Maximum Likelihood (ML)

- an estimation method
- ML = Mode Regression while
  LS = Mean Regression
- assume the probability distribution of the involved random variables up to (all or some of) their parameters, e.g., normal

ML Principle

The observed data set is the most likely. It must be at the mode of their joint probability distribution.

Choose the estimator that maximizes the likelihood (mode) of the observed data set.

ML Estimator for μ (1)

<u>Assumption</u>

$$X \sim \mathrm{N}(\mu, \sigma^2)$$

pdf of X

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right)$$

μ and σ^2 are unknown parameters

ML Estimator for μ (2)

Given the randomly sampled data $[x_1, x_2, \ldots, x_n]$, its joint pdf (Likelihood function) can be written as

$$L(\mu, \sigma^2; \mathbf{X}) = \prod_{i=1}^{n} f(x_i)$$

ML Estimator for μ (3)

In general, maximization of Log of the likelihood function is much easier

$$\ln L(\mu, \sigma^2; \mathbf{X}) = \sum_{i=1}^{n} \ln(f(x_i))$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Same solution as maximization of the likelihood function. Why?

ML Estimator for μ (4)

$$\max_{\mu, \sigma^2} \quad -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

<u>Solution</u>

$$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML})^2$$

ML Estimator for μ (5)

$\hat{\mu}_{ML}$ is an unbiased (fortunately) estimator of μ

$\hat{\sigma}^2_{ML}$ is a biased but consistent estimator of σ^2

ML Estimator for μ (6)

First-order Conditions $\underbrace{\nabla \ln L}_{2\text{x}1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

where

$$\nabla \ln L = \begin{bmatrix} \dfrac{\partial \ln L}{\partial \mu} \\ \\ \dfrac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \\ \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 \end{bmatrix}$$

ML Estimator for μ (7)

Asymptotic variance-covariance matrix is

$$\left[-\nabla^2 \ln L\right]^{-1} = \begin{bmatrix} -\dfrac{\partial^2 \ln L}{\partial \mu^2} & -\dfrac{\partial^2 \ln L}{\partial \sigma^2 \partial \mu} \\ -\dfrac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} & -\dfrac{\partial^2 \ln L}{\partial \sigma^4} \end{bmatrix}^{-1}$$

evaluated at the solution.

ML Estimator for μ (8)

Asymptotic variance-covariance matrix

$$= \begin{bmatrix} \dfrac{n}{\hat{\sigma}^2} & 0 \\ 0 & \dfrac{n}{2}\left(\dfrac{1}{\hat{\sigma}^2}\right)^2 \end{bmatrix}^{-1} = \begin{bmatrix} \dfrac{\hat{\sigma}^2}{n} & 0 \\ 0 & \dfrac{2}{n}\left(\hat{\sigma}^2\right)^2 \end{bmatrix}$$

ML Estimator for μ (9)

Asymptotic Distribution

$$\sqrt{n}(\hat{\mu}_{ML} - \mu) \overset{A}{\sim} \mathrm{N}(0, \hat{\sigma}^2_{ML})$$

$$\sqrt{n}(\hat{\sigma}^2_{ML} - \sigma^2) \overset{A}{\sim} \mathrm{N}\left(0, 2\left(\hat{\sigma}^2_{ML}\right)^2\right)$$

ML Estimator for μ (10)

Approx. Variances (for stat. Inference)

$$V(\hat{\mu}_{ML}) = \frac{\hat{\sigma}^2_{ML}}{n}$$

$$V(\hat{\sigma}^2_{ML}) = \frac{2}{n}\left(\hat{\sigma}^2_{ML}\right)^2$$

ML Estimator for μ (11)

$(1-\alpha)100\%$ approx. Confidence Interval

for $\mu = \hat{\mu}_{ML} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{\sigma}^2_{ML}/n}$

$(1-\alpha)100\%$ approx. Confidence Interval

for $\sigma^2 = \hat{\sigma}^2_{ML} \pm z_{\frac{\alpha}{2}} \sqrt{2\left(\hat{\sigma}^2_{ML}/n\right)^2}$

ML Estimator for λ (1)

<u>Assumption</u>

$$X \sim \text{Poisson}(\lambda)$$

pmf of X

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

ML Estimator for λ (2)

Given the randomly sampled data $[x_1, x_2, \ldots, x_n]$, its log Likelihood function can be written as

$$\ln L(\lambda; \mathbf{X}) = \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)$$

$$= -n\lambda + \ln(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

ML Estimator for λ (3)

$$\max_{\lambda} \quad -n\lambda + \ln(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

<u>FOC</u> $-n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0$

<u>Solution</u> $\hat{\lambda}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$

unbiased => $E(\hat{\lambda}_{ML}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \lambda$

ML Estimator for λ (4)

Asymptotic Variance of $\hat{\lambda}_{ML}$

$$\left[-\frac{d^2 \ln L}{d\lambda^2}\right]^{-1} = \left[\frac{1}{\hat{\lambda}^2}\sum_{i=1}^{n} x_i\right]^{-1}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} x_i = \frac{1}{n}\hat{\lambda}_{ML}$$

ML Estimator for λ (5)

Asymptotic Distribution of $\hat{\lambda}_{ML}$

$$\sqrt{n}(\hat{\lambda}_{ML} - \lambda) \overset{A}{\sim} N\left(0, \hat{\lambda}_{ML}\right)$$

(1-α)100% approx. Confidence Interval

for $\lambda = \quad \hat{\lambda}_{ML} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n}\hat{\lambda}_{ML}}$

ML Estimator for CLNRM (1)

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \sim MVN(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

pdf for each ε_i

$$f(\varepsilon_i) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{1}{2} \left(\frac{Y_i - X_i \boldsymbol{\beta}}{\sigma} \right)^2 \right)$$

ML Estimator for CLNRM (2)

Given the observation $(\mathbf{Y}, \mathbf{X})$, the log Likelihood function can be written as

$$\ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(Y_i - X_i \boldsymbol{\beta} \right)^2$$

ML Estimator for CLNRM (3)

$$\max_{\boldsymbol{\beta},\sigma^2} -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]^{\mathrm{T}}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]$$

Solution

$$\hat{\boldsymbol{\beta}}_{ML} = [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \implies \hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{OLS}$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n}[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}]^{\mathrm{T}}[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}] = \frac{1}{n}SSR$$

ML Estimator for CLNRM (4)

FOC

$$\nabla \ln L = \begin{bmatrix} \dfrac{\partial \ln L}{\partial \boldsymbol{\beta}} \\ \dfrac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sigma^2}\mathbf{X}^{\mathrm{T}}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}] \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]^{\mathrm{T}}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}] \end{bmatrix}$$

$$= \mathbf{0}$$

ML Estimator for CLNRM (5)

Asymptotic Variance of $\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}^2_{ML}$

$$\left[-\nabla^2 \ln L\right]^{-1} = \left[\begin{array}{c|c} \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X} & \frac{1}{\sigma^4}\mathbf{X}^T[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}] \\ \hline \frac{1}{\sigma^4}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]^T\mathbf{X} & -\frac{n}{2\sigma^4}+\frac{1}{\sigma^6}[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}]^T[\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}] \end{array}\right]^{-1}$$

evaluated at the solution.

ML Estimator for CLNRM (6)

Asymptotic Variance of $\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}^2_{ML}$

$$= \left[\begin{array}{c|c} \frac{1}{\hat{\sigma}^2}\mathbf{X}^T\mathbf{X} & 0 \\ \hline 0 & \frac{n}{2\hat{\sigma}^4} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \hat{\sigma}^2[\mathbf{X}^T\mathbf{X}]^{-1} & 0 \\ \hline 0 & \frac{2\hat{\sigma}^4}{n} \end{array}\right]$$

ML Estimator for CLNRM (7)

Asymptotic Distribution of $\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}^2_{ML}$

$$\sqrt{n}\begin{bmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \\ \hat{\sigma}^2 - \sigma^2 \end{bmatrix} \overset{A}{\sim} \text{MVN}\left(\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2 \left[\dfrac{\mathbf{X}^T\mathbf{X}}{n} \right]^{-1} & \mathbf{0} \\ \mathbf{0} & 2\hat{\sigma}^4 \end{bmatrix} \right)$$

ML Estimator for CLNRM (8)

Log Likelihood Value

$$\ln L(\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}^2_{ML}, \mathbf{Y}, \mathbf{X}) = -\frac{n}{2}\ln\left(2\pi \frac{SSR}{n} \right) - \frac{n}{2}$$

$$= -\frac{n}{2}\left\{ 1 + \ln(2\pi) + \ln\left(\frac{SSR}{n} \right) \right\}$$

ML Estimator for CLNRM (9)

Since the log likelihood value is simply a function of *SSR* from OLS, it is generally also reported in the OLS result report. No ML estimation is actually done.

Restricted ML (1)

Elimination Approach (see RLS)

$$\mathbf{P} = \mathbf{Z\delta} + \boldsymbol{\varepsilon}$$

$$\max_{\boldsymbol{\delta},\sigma^2} \; -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\mathbf{P}-\mathbf{Z\delta}]^{\mathrm{T}}[\mathbf{P}-\mathbf{Z\delta}]$$

Solution: same as RLS except for σ^2

Restricted ML (2)

Lagrange Method

$$\max_{\boldsymbol{\beta},\sigma^2} \; -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}]^{\mathrm{T}}[\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}]$$

subject to $\underbrace{\mathbf{R}}_{M\text{x}K}\ \underbrace{\boldsymbol{\beta}}_{K\text{x}1} = \underbrace{\mathbf{r}}_{M\text{x}1}$

Restricted ML (3)

<u>FOC</u>

$$\nabla \ln L = \begin{bmatrix} \dfrac{\partial \ln L}{\partial \boldsymbol{\beta}} - \mathbf{R}^{\mathbf{T}}\boldsymbol{\lambda} \\ \\ \dfrac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \\ 0 \end{bmatrix}$$

$\boldsymbol{\lambda}$ is the Mx1 Lagrange Multiplier vector

Restricted ML (4)

Define $\boldsymbol{\theta} = \sigma^2 \boldsymbol{\lambda}$

From FOC, $\mathbf{X}^{\mathrm{T}}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_R] - \mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}} = \mathbf{0}$

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} - \mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}}_R - \mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}} = \mathbf{0}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_R &= [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}[\mathbf{X}^{\mathrm{T}}\mathbf{Y} - \mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}}] \\ &= [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} - [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}} \\ &= \hat{\boldsymbol{\beta}}_U - [\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}}\end{aligned}$$

Restricted ML (5)

Substitute into **R**β=**r**

$$[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}] - \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\hat{\boldsymbol{\theta}} = \mathbf{0}$$

$$\hat{\boldsymbol{\theta}} = \mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$

where $\mathbf{S} = \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}$

<u>Solution</u>: same $\hat{\boldsymbol{\beta}}_R$ as that in RLS but

$$\hat{\sigma}_R^2 = \frac{SSR_R}{n}$$

Restricted ML (6)

Asymptotic Var-cov matrix of $\hat{\boldsymbol{\beta}}_R, \hat{\sigma}_R^2$

$$\left[-\nabla^2 \ln L\right]^{-1} = \hat{\sigma}_R^2 \left[\begin{array}{c|c} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \frac{1}{\hat{\sigma}_R^2}\mathbf{X}^{\mathrm{T}}[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_R] \\ \hline \frac{1}{\hat{\sigma}_R^2}[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_R]^{\mathrm{T}}\mathbf{X} & \frac{n}{2\hat{\sigma}_R^2} \end{array}\right]^{-1}$$

Restricted ML (7)

Asymptotic Var-cov matrix of $\hat{\boldsymbol{\beta}}_R, \hat{\sigma}_R^2$

$$\left[-\nabla^2 \ln L\right]^{-1} = \hat{\sigma}_R^2 \left[\begin{array}{c|c} \mathbf{X}^{\mathrm{T}}\mathbf{X} & \frac{1}{\hat{\sigma}_R^2}\mathbf{R}^{\mathrm{T}}\hat{\boldsymbol{\theta}} \\ \hline \frac{1}{\hat{\sigma}_R^2}\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{R} & \frac{n}{2\hat{\sigma}_R^2} \end{array}\right]^{-1}$$

Restricted ML (8)

Inverse of the Second-order matrix ??

Identical Var-Covar matrix as in RLS??

Testing CLNRM using ML

Assume large sample

- Likelihood Ratio(LR) Test
- Wald Test
- Lagrange Multiplier(LM) Test

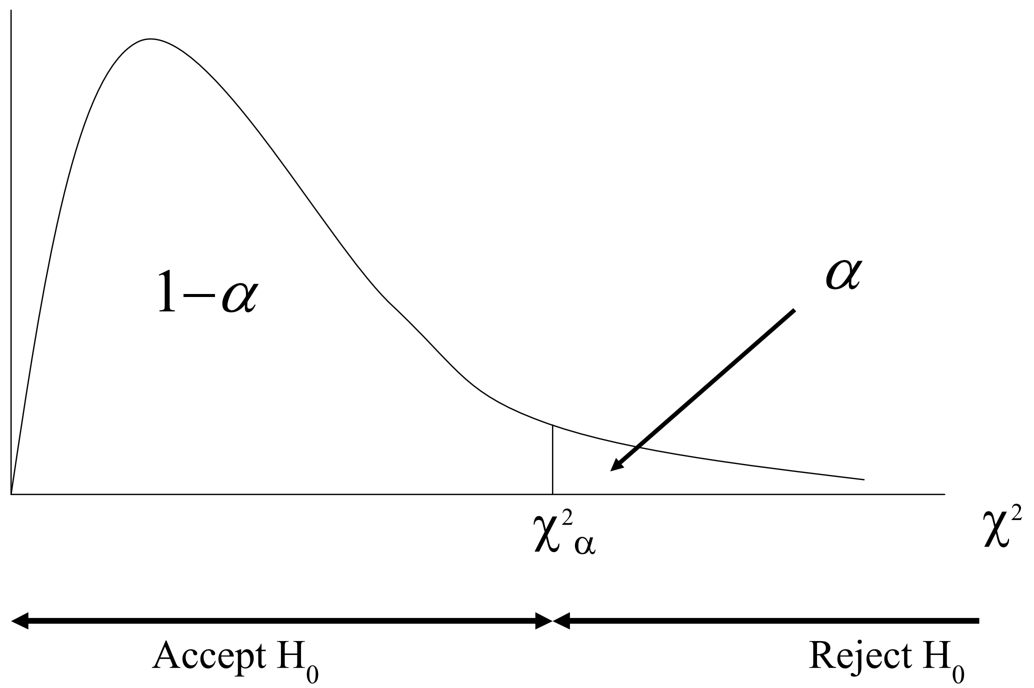
LR Test (1)

- An alternative to Generalized F-test (RLS) for large sample
- Compare likelihood of the restricted model (L_R) with that of the unrestricted model (L_U)
- $\ln L_R$ is always less than or equal to $\ln L_U$
- Small gap ==> H_0 is true

$$LR = -2\ln\left(\frac{L_R}{L_U}\right) = -2(\ln L_R - \ln L_U) \sim \chi^2(M)$$

- Perform the right-tailed Chi-square test

LR Test (2)



LR Test (3)

Technical Details

$$\ln L_U = -\frac{n}{2}\left\{1 + \ln(2\pi) + \ln\left(\frac{SSR_U}{n}\right)\right\}$$

$$\ln L_R = -\frac{n}{2}\left\{1 + \ln(2\pi) + \ln\left(\frac{SSR_R}{n}\right)\right\}$$

$$LR = -2(\ln L_R - \ln L_U) = n\ln\left(\frac{SSR_R}{SSR_U}\right)$$

LR Test (4)

Why does LR have a Chi-square distribution?

Wald Test (1)

For large sample,

asymptotic Variance of $\hat{\boldsymbol{\beta}}_U = \hat{\sigma}^2_{ML}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}$

Same concept as the single-run Generalized F-test.

If $\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r} = \mathbf{0}$, then, H_0 cannot be rejected. Only when the difference from zero is significant enough, then, H_0 will be rejected.

Wald Test (2)

Wald statistic

$$W = [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}}[\mathbf{R}\mathrm{V}(\hat{\boldsymbol{\beta}})\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]$$

$$= [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}}[\hat{\sigma}^2_{ML}\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]$$

$$= [\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]^{\mathrm{T}}[\mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathrm{T}}]^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}]\frac{1}{\hat{\sigma}^2_{ML}}$$

$$\sim \chi^2(M)$$

where $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_U$

Wald Test (3)

Perform the same right-tailed Chi-square test. Same criterion as in LR test.

It is the Chi-square statistic reported in EViews Wald test output.

LM Test (3)

<u>Solution</u> $\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\beta}}_R, \hat{\sigma}_R^2, L_R$

Undoubtedly accept H_0 if $\hat{\boldsymbol{\lambda}} = \mathbf{0}$

when the constraints did not affect solution (same as that of unrestricted model). Only the difference from zero becomes significant before H_0 will be rejected

LM Test (4)

LM statistic

$$LM = \hat{\boldsymbol{\lambda}}^{\mathrm{T}}[\mathrm{V}(\hat{\boldsymbol{\lambda}})]^{-1}\hat{\boldsymbol{\lambda}} \sim \chi^2(M)$$

Perform a right-tailed Chi-square test with the same criterion as LR and Wald tests.

LM Test (6)

Technical Details

From $\hat{\boldsymbol{\theta}} = \mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$

where $\mathbf{S} = \mathbf{R}[\mathbf{X}^{\mathrm{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}$

$$\mathrm{V}(\hat{\boldsymbol{\theta}}) = \mathbf{S}^{-1}\mathbf{R}\mathrm{V}(\hat{\boldsymbol{\beta}}_U)\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}$$

where

$$\mathrm{V}(\hat{\boldsymbol{\theta}}) = \sigma^2\mathbf{S}^{-1}\mathbf{R}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-1}\mathbf{R}^{\mathbf{T}}\mathbf{S}^{-1}$$
$$= \sigma^2\mathbf{S}^{-1}$$

LM Test (7)

Technical Details

Since $\hat{\sigma}_R^2$ is a consistent estimator, the consistent estimator of λ

$$\hat{\boldsymbol{\lambda}} = \frac{1}{\hat{\sigma}_R^2}\mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$

and its asymptotic variance

$$\mathrm{V}(\hat{\boldsymbol{\lambda}}) = \frac{1}{\hat{\sigma}_R^2}\mathbf{S}^{-1}$$

LM Test (8)

Technical Details

$$\mathrm{LM} = \frac{1}{\hat{\sigma}_R^2}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]^{\mathbf{T}}\mathbf{S}^{-1}\mathbf{S}\mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$
$$= \frac{1}{\hat{\sigma}_R^2}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]^{\mathbf{T}}\mathbf{S}^{-1}[\mathbf{R}\hat{\boldsymbol{\beta}}_U - \mathbf{r}]$$

which is similar to Wald statistic except that $\hat{\sigma}_R^2$ is used instead of $\hat{\sigma}_U^2$

Order of Magnitude

For a Multiple Linear Regression

$$LM \le LR \le W$$

They give different values of χ^2 cal for the same hypothesis testing.

Note that the Wald test might reject H_0 while LM accepts it.

Overall F-test with large sample

LR test

$$\chi^2_{cal} = -n \ln\left(\frac{SSR}{SST}\right) = -n \ln(1 - R^2)$$

Wald test

$$\chi^2_{cal} = \frac{n(SST - SSR)}{SSR} = \frac{nR^2}{1 - R^2}$$

Overall F-test with large sample

LM test

$$\chi^2_{cal} = \frac{n(SST - SSR)}{SST} = nR^2$$

Feel free to prove them.