

## 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 Estimator for $\mu$ (1)

### Assumption

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

pdf of X

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

$\mu$  and  $\sigma^2$  are unknown parameters

## 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 $\mu$ (2)

Given the randomly sampled data  $[x_1, x_2, \dots, x_n]$ , 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 $\mu$ (3)

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

$$\begin{aligned}\ln L(\mu, \sigma^2; \mathbf{X}) &= \sum_{i=1}^n \ln(f(x_i)) \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\end{aligned}$$

Same solution as maximization of the likelihood function. Why?

## ML Estimator for $\mu$ (5)

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

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

## ML Estimator for $\mu$ (4)

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

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

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

## ML Estimator for $\mu$ (6)

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

where

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

## ML Estimator for $\mu$ (7)

Asymptotic variance-covariance matrix is

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

evaluated at the solution.

## ML Estimator for $\mu$ (9)

Asymptotic Distribution

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

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

## ML Estimator for $\mu$ (8)

Asymptotic variance-covariance matrix

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

## ML Estimator for $\mu$ (10)

Approx. Variances (for stat. Inference)

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

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

## ML Estimator for $\mu$ (11)

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

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

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

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

## ML Estimator for $\lambda$ (2)

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

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

## ML Estimator for $\lambda$ (1)

Assumption

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

pmf of X

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

## ML Estimator for $\lambda$ (3)

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

$$\text{FOC} \quad -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0$$

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

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

## ML Estimator for $\hat{\lambda}$ (4)

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

$$\begin{aligned} \left[ -\frac{d^2 \ln L}{d\lambda^2} \right]^{-1} &= \left[ \frac{1}{\hat{\lambda}^2} \sum_{i=1}^n x_i \right]^{-1} \\ &= \frac{1}{n^2} \sum_{i=1}^n x_i = \frac{1}{n} \hat{\lambda}_{ML} \end{aligned}$$

## ML Estimator for CLNRM (1)

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

pdf for each  $\boldsymbol{\varepsilon}_i$

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

## ML Estimator for $\hat{\lambda}$ (5)

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

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

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

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

## ML Estimator for CLNRM (2)

Given the observation  $(\mathbf{Y}, \mathbf{X})$ , the log Likelihood

function can be written as

$$\ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}, \mathbf{X}) = -\frac{n}{2} \ln(2\pi\sigma^2)$$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - X_i \boldsymbol{\beta})^2$$

### ML Estimator for CLNRM (3)

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

Solution

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

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

### ML Estimator for CLNRM (5)

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

$$[-\nabla^2 \ln L]^{-1} = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X} & \frac{1}{\sigma^4} \mathbf{X}^T [\mathbf{Y} - \mathbf{X}\beta] \\ \frac{1}{\sigma^4} [\mathbf{Y} - \mathbf{X}\beta]^T \mathbf{X} & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} [\mathbf{Y} - \mathbf{X}\beta]^T [\mathbf{Y} - \mathbf{X}\beta] \end{bmatrix}^{-1}$$

evaluated at the solution.

### ML Estimator for CLNRM (4)

FOC

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

$$= \mathbf{0}$$

### ML Estimator for CLNRM (6)

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

$$= \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \mathbf{X}^T \mathbf{X} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\sigma}^2 [\mathbf{X}^T \mathbf{X}]^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

## ML Estimator for CLNRM (7)

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

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

## 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.

## ML Estimator for CLNRM (8)

Log Likelihood Value

$$\begin{aligned} \ln L(\hat{\boldsymbol{\beta}}_{ML}, \hat{\sigma}_{ML}^2, \mathbf{Y}, \mathbf{X}) &= -\frac{n}{2} \ln \left( 2\pi \frac{SSR}{n} \right) - \frac{n}{2} \\ &= -\frac{n}{2} \left\{ 1 + \ln(2\pi) + \ln \left( \frac{SSR}{n} \right) \right\} \end{aligned}$$

## Restricted ML (1)

Elimination Approach (see RLS)

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

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

Solution: same as RLS except for  $\sigma^2$

## Restricted ML (2)

Lagrange Method

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

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

## Restricted ML (4)

Define  $\theta = \sigma^2 \lambda$

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

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

$$\hat{\beta}_R = [\mathbf{X}^T \mathbf{X}]^{-1} [\mathbf{X}^T \mathbf{Y} - \mathbf{R}^T \hat{\theta}]$$

$$= [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{Y} - [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{R}^T \hat{\theta}$$

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

## Restricted ML (3)

$$\text{FOC} \quad \nabla \ln L = \begin{bmatrix} \frac{\partial \ln L}{\partial \beta} - \mathbf{R}^T \lambda \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}$$

$\lambda$  is the  $M \times 1$  Lagrange Multiplier vector

## Restricted ML (5)

Substitute into  $\mathbf{R}\beta = \mathbf{r}$

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

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

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

Solution: same  $\hat{\beta}_R$  as that in RLS but

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

## Restricted ML (6)

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

$$[-\nabla^2 \ln L]^{-1} = \hat{\sigma}_R^2 \begin{bmatrix} \mathbf{X}'\mathbf{X} & \frac{1}{\hat{\sigma}_R^2} \mathbf{X}'[\mathbf{Y} - \mathbf{X}\hat{\beta}_R] \\ \frac{1}{\hat{\sigma}_R^2} [\mathbf{Y} - \mathbf{X}\hat{\beta}_R]'\mathbf{X} & \frac{n}{2\hat{\sigma}_R^2} \end{bmatrix}^{-1}$$

## Restricted ML (8)

Inverse of the Second-order matrix ??

Identical Var-Covar matrix as in RLS??

## Restricted ML (7)

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

$$[-\nabla^2 \ln L]^{-1} = \hat{\sigma}_R^2 \begin{bmatrix} \mathbf{X}'\mathbf{X} & \frac{1}{\hat{\sigma}_R^2} \mathbf{R}'\hat{\theta} \\ \frac{1}{\hat{\sigma}_R^2} \hat{\theta}'\mathbf{R} & \frac{n}{2\hat{\sigma}_R^2} \end{bmatrix}^{-1}$$

## 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  $\implies H_0$  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 (3)

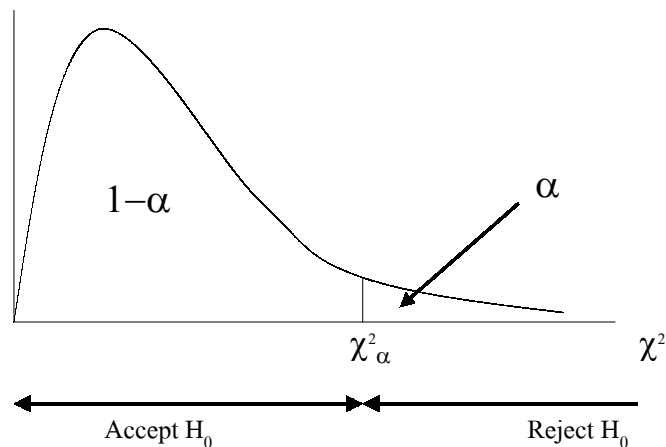
Technical Details

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

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

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

## LR Test (2)



## LR Test (4)

Why does LR have a Chi-square distribution?

## Wald Test (1)

For large sample,

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

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

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

## Wald Test (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.

## Wald Test (2)

Wald statistic

$$\begin{aligned} W &= [\mathbf{R}\hat{\beta} - \mathbf{r}]^T [\mathbf{R}V(\hat{\beta})\mathbf{R}^T]^{-1} [\mathbf{R}\hat{\beta} - \mathbf{r}] \\ &= [\mathbf{R}\hat{\beta} - \mathbf{r}]^T [\hat{\sigma}_{ML}^2 \mathbf{R}[\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{R}^T]^{-1} [\mathbf{R}\hat{\beta} - \mathbf{r}] \\ &= [\mathbf{R}\hat{\beta} - \mathbf{r}]^T [\mathbf{R}[\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{R}^T]^{-1} [\mathbf{R}\hat{\beta} - \mathbf{r}] \frac{1}{\hat{\sigma}_{ML}^2} \\ &\sim \chi^2(M) \end{aligned}$$

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

## LM Test (3)

Solution  $\hat{\lambda}, \hat{\beta}_R, \hat{\sigma}_R^2, L_R$

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

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

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

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

## LM Test (7)

Technical Details

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

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

and its asymptotic variance

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

## LM Test (6)

Technical Details

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

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

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

where

$$\begin{aligned} V(\hat{\theta}) &= \sigma^2 \mathbf{S}^{-1} \mathbf{R} [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{R}^T \mathbf{S}^{-1} \\ &= \sigma^2 \mathbf{S}^{-1} \end{aligned}$$

## LM Test (8)

Technical Details

$$\begin{aligned} LM &= \frac{1}{\hat{\sigma}_R^2} [\mathbf{R}\hat{\beta}_U - \mathbf{r}]^T \mathbf{S}^{-1} \mathbf{S} \mathbf{S}^{-1} [\mathbf{R}\hat{\beta}_U - \mathbf{r}] \\ &= \frac{1}{\hat{\sigma}_R^2} [\mathbf{R}\hat{\beta}_U - \mathbf{r}]^T \mathbf{S}^{-1} [\mathbf{R}\hat{\beta}_U - \mathbf{r}] \end{aligned}$$

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

$$LM \leq LR \leq W$$

They give different values of  $\chi^2$  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

LM test

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

Feel free to prove them.

## Overall F-test with large sample

LR test

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

Wald test

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