- Hypothesis test about multiple linear combinations of the regression coefficients

① Wald approach.

$\mathbb{C}$: $r \times k$ constant matrix, $r \leqq k$, full rank

Estimation of $\underset{r\times k}{\mathbb{C}}\underset{k\times 1}{\beta}$ by $\mathbb{C}\hat{\beta}$

$\mathbb{C}\hat{\beta} \sim N(\mathbb{C}\beta,\ \underbrace{\sigma^2\mathbb{C}(X'X)^{-1}\mathbb{C}'}_{(=\Gamma\Gamma' \text{ by Cholesky's decomposition})})$ Thus, $\overset{\Gamma^{-1}}{\overset{\parallel}{[\mathbb{C}(X'X)^{-1}]^{-\frac{1}{2}}}}(\mathbb{C}\hat{\beta}-\mathbb{C}\beta) \sim N(0, I)$

Therefore,

$$\underbrace{(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)'(\mathbb{C}\sigma^2(X'X)^{-1}\mathbb{C}')^{-1}(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)}_{:=A} \sim \chi^2_{(r)}$$

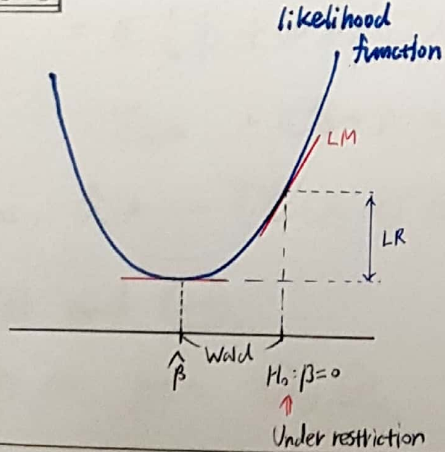
$$\frac{A/r \quad \sim \chi^2(r)/r}{\frac{\hat{\sigma}^2}{\sigma^2} \quad \sim \frac{\chi^2(N-k)}{N-k}} = \boxed{(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)'(\mathbb{C}\hat{\sigma}^2(X'X)^{-1}\mathbb{C}')^{-1}(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)/r \sim F(r, N-k) \Updownarrow \frac{\chi^2(r)/r}{\chi^2(N-k)/N-k}}$$

$H_0 : \mathbb{C}\beta = \underset{r\times 1}{A}$: hypothesis test using F-distribution (Wald approach)

$\Rightarrow \exists$ 2 alternative ways for hypothesis tests

; Likelihood Ratio (LR) and Lagrangean multiplier approaches (LM)

Note



Wald : how far $\hat{\beta}$ to $H_0 : \beta = 0$
LR : minimum vs. under restriction
LM : slope of $\hat{\beta}$ vs slope of under restriction.

$\Rightarrow$ No difference for linear models
(Wald, LR and LM are numerically indifferent)

3 LR approach

SSR_R : the Restricted Sum of Squared Residuals $= \hat{u}_R'\hat{u}_R$, $\hat{u}_R = Y - X\hat{\beta}_R$

SSR_{UR} : the UnRestricted 〃 $= \hat{u}'\hat{u}$, $\hat{u} = Y - X\hat{\beta}$

$\hat{\beta}_R$ minimizes $\sum_{i=1}^{N}(y_i - x_i'b)^2$ s.t. $\underbrace{Cb = A}_{\text{restriction}}$.

For example, $y_i = x_i'\beta + u_i = x_{1i}'\beta_1 + x_{2i}'\beta_2 + u_i$

unrestricted : $\beta = 0$ $\quad \hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}$

Restricted : $\beta_R = 0$ $\quad \hat{\beta}_R = \begin{pmatrix} \hat{\beta}_{1R} \\ \hat{\beta}_{2R} \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{\beta}_{2R} \end{pmatrix}$ → Given $\beta_1 = 0$, minimize β_2

$(Y - Xb)'(Y - Xb)$ s.t $Cb = A$

$$\mathcal{L} = \frac{1}{2}(Y - Xb)'(Y - Xb) - \lambda'(Cb - A) \quad \text{where } \underset{r\times 1}{\lambda} \in \mathbb{R}^r$$

F.O.C. w.r.t b : $-X'(Y - Xb) - C\lambda = 0$

F.O.C. w.r.t λ : $-(Cb - A) = 0$

$$X'Xb = X'Y + C'\lambda \qquad \underline{\hat{\beta}_R} = (X'X)^{-1}X'Y + (X'X)^{-1}\underset{k\times r}{C'}\underset{r\times 1}{\lambda}$$

$$= \underline{\hat{\beta} + (X'X)^{-1}C'\hat{\lambda}} \quad \cdots \text{(*)}$$

Let $D := C\hat{\beta}_R - A$

$= C[\hat{\beta} + (X'X)^{-1}C'\lambda] - A = 0$ under the null.

Then, $-C\hat{\beta} + A = C(X'X)^{-1}C'\lambda$

Thus, $\hat{\lambda} = -[C(X'X)^{-1}C']^{-1}[C\hat{\beta} - A] \quad \cdots \text{(**)}$

By (*) and (**),

we can get $\hat{u}_R'\hat{u}_R$

Now, $$\hat{u}_R = \mathbb{Y} - \mathbb{X}\hat{\beta}_R$$

$$= \mathbb{Y} - \mathbb{X}\left[\hat{\beta} + (\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}\right]$$

$$= \mathbb{Y} - \mathbb{X}\hat{\beta} - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}$$

$$= \hat{u} - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}$$

Thus, $$\hat{u}_R'\hat{u}_R = \left[\hat{u} - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}\right]'\left[\hat{u} - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}\right]$$

$$= \hat{u}'\hat{u} - \underbrace{\hat{u}'\mathbb{X}}_{=0}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda} - \hat{\lambda}'\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\underbrace{\mathbb{X}'\hat{u}}_{=0}$$

$$+ \hat{\lambda}'\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\cancel{\mathbb{X}'\mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}}\mathbb{C}'\hat{\lambda}$$

$$= \hat{u}'\hat{u} + \hat{\lambda}'\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}$$

$$LR = \frac{(SSR_R - SSR_{UR})/\overset{\text{rank of }\mathbb{C}}{r}}{SSR_{UR}/(N-K)} = \frac{(\hat{u}_R'\hat{u}_R - \hat{u}'\hat{u})/r}{\hat{u}'\hat{u}/(N-K)}$$

$$= \frac{\hat{\lambda}'\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\hat{\lambda}/r}{\hat{u}'\hat{u}/(N-K)}$$

$\downarrow$ $\hat{\sigma}^2 = \hat{u}'\hat{u}/(N-K)$ and consider (**) : $\hat{\lambda} = -\left[\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\right]^{-1}\left[\mathbb{C}\hat{\beta} - A\right]$

$$= \frac{\left[\mathbb{C}\hat{\beta} - A\right]'\left[\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\right]^{-1}\cancel{\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\left[\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\right]^{-1}}\left[\mathbb{C}\hat{\beta} - A\right]/r}{\hat{u}'\hat{u}/(N-K)}$$

Thus,

$$\boxed{LR = \frac{\left[\mathbb{C}\hat{\beta} - A\right]'\left[\mathbb{C}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{C}'\right]^{-1}\left[\mathbb{C}\hat{\beta} - A\right]}{\hat{\sigma}^2}}$$ → the same as Wald statistics.

※ LR is convenient for calculation.

Wald can be used for the null as well as the alternative hypothesis

$$\tilde{\sigma}^2 = \frac{1}{N-K+r} \sum_{i=1}^{N} (y_i - x_i' \hat{\beta}_R)^2$$

$\uparrow$

variance for the restricted case.

$\Rightarrow$ We do not use this since $\hat{\sigma}^2 < \tilde{\sigma}^2$ using $\frac{1}{N-K}$

$$\mathbb{C}\hat{\beta} \sim N(\mathbb{C}\beta,\ \mathbb{C}\sigma^2(X'X)^{-1}\mathbb{C}')$$

Density of $\mathbb{C}\hat{\beta}$ = constant # $\cdot \exp\{-\frac{1}{2}(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)'[\mathbb{C}\sigma^2(X'X)^{-1}\mathbb{C}']^{-1}(\mathbb{C}\hat{\beta}-\mathbb{C}\beta)\}$

$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\{-\frac{1}{2}A\}$. where we defined A before.

- MLE of β, σ^2

$y_i | x_i \sim \underline{N(x_i'\beta, \sigma^2)} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right\} \cdots \circledast$

by iid sample.

$(y_1, y_2, \dots, y_N) | x = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left\{-\sum_{i=1}^{N} \frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right\}$

($\rightarrow$ This is joint p.d.f of $y_1, \dots, y_N$. By iid, we can multiply $\circledast$ N times)

MLE is obtained by maximizing likelihood over β & σ^2.

Log likelihood $(\beta, \sigma^2) = \text{constant \#} - \frac{N}{2}\log\sigma^2 - \boxed{\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)^2}$

※ max. log likelihood is the same as $\underline{\min \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)^2}$ for linear cases.

Therefore, $\underline{\hat{\beta}_{MLE} = \hat{\beta}_{OLS} \text{ so, unbiased}}$.

$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})^2 = 0 \quad \hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\beta)^2$

Therefore, $\underline{\hat{\sigma}^2_{MLE} \text{ is biased}}$. ($\because$ divided by $\frac{1}{N}$, Not $\frac{1}{N-K}$)

- Information matrix

$\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - x_i'\beta)x_i$

) First order derivative.

$\frac{\partial L}{\partial \sigma^2} = -\frac{N}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\cdot\sum_{i=1}^{N}(y_i - x_i'\beta)^2$

$\frac{\partial^2 L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}\sum_{i=1}^{N}x_i x_i' \qquad \frac{\partial^2 L}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{N}(y_i - x_i'\beta)x_i$

) Second order derivative.

$\frac{\partial^2 L}{(\partial\sigma^2)^2} = \frac{N}{2}\cdot\frac{1}{\sigma^4} - \frac{1}{\sigma^6}\cdot\sum_{i=1}^{N}(y_i - x_i'\beta)^2$

Take expectation!

$$E\left\{\frac{\partial^2 L}{\partial\beta\partial\beta'}\,\middle|\,X\right\} = -\frac{1}{\sigma^2}\sum_{i=1}^{N} x_i x_i'$$

$$E\left\{\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\,\middle|\,X\right\} = 0$$

$$E\left\{\frac{\partial^2 L}{(\partial\sigma^2)^2}\,\middle|\,X\right\} = \frac{N}{2}\cdot\frac{1}{\sigma^4} - N\cdot\frac{1}{\sigma^4} = \frac{-N}{2\sigma^4}$$

Information matrix $\mathcal{J} = \begin{pmatrix} -E\left\{\frac{\partial^2 L}{\partial\beta\partial\beta'}\,\middle|\,X\right\} & -E\left\{\frac{\partial^2 L}{\partial\sigma^2\partial\beta}\,\middle|\,X\right\} \\ -E\left\{\frac{\partial^2 L}{\partial\beta\partial\sigma^2}\,\middle|\,X\right\} & -E\left\{\frac{\partial^2 L}{(\partial\sigma^2)^2}\,\middle|\,X\right\} \end{pmatrix}$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{N} x_i x_i' & 0 \\ 0 & \frac{N}{2\sigma^4} \end{pmatrix}$$

Thus.

$$\mathcal{J}^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{pmatrix}$$

※ Under the following assumptions : $\left(\begin{array}{l} E(y_i|x_i) = x_i'\beta \\ V(y_i|x_i) = \sigma^2 \text{ (No heteroskedasticity)} \\ y_i = x_i'\beta + u_i \\ \quad iid \end{array} \right) \rightarrow$ OLS is BLUE

※ Normality so that $y_i|x_i \sim N(x_i'\beta, \sigma^2)$

Then, OLS = $\hat{\beta}_{MLE}$ & OLS is the Best unbiased estimator.

(Under Normality, without linear assumption,

OLS & MLE is the Best estimator.)