Accordingly, b is more efficient than $\hat{\beta}$.

7. If $u \sim N(0, \sigma^2\Omega)$, then $b \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$.

   Consider testing the hypothesis $H_0 : R\beta = r$.

   $$R : \ G \times k, \qquad \text{rank}(R) = G \leq k.$$

   $$Rb \sim N(R\beta, \sigma^2 R(X'\Omega^{-1}X)^{-1}R').$$

   Therefore, the following quadratic form is distributed as:

   $$\frac{(Rb - r)'(R(X'\Omega^{-1}X)^{-1}R')^{-1}(Rb - r)}{\sigma^2} \sim \chi^2(G)$$

8. Because $(y^\star - X^\star b)'(y^\star - X^\star b)/\sigma^2 \sim \chi^2(n-k)$, we obtain:

   $$\frac{(y - Xb)'\Omega^{-1}(y - Xb)}{\sigma^2} \sim \chi^2(n-k)$$

9. Furthermore, from the fact that b is independent of $y - Xb$, the following F distribution can be derived:

$$\frac{(Rb-r)'(R(X'\Omega^{-1}X)^{-1}R')^{-1}(Rb-r)/G}{(y-Xb)'\Omega^{-1}(y-Xb)/(n-k)} \sim F(G, n-k)$$

10. Let b be the unrestricted GLSE and $\tilde{b}$ be the restricted GLSE.

    Their residuals are given by e and $\tilde{u}$, respectively.

$$e = y - Xb, \qquad \tilde{u} = y - X\tilde{b}$$

    Then, the F test statistic is written as follows:

$$\frac{(\tilde{u}'\Omega^{-1}\tilde{u} - e'\Omega^{-1}e)/G}{e'\Omega^{-1}e/(n-k)} \sim F(G, n-k)$$

8.1 Example: Mixed Estimation (Theil and Goldberger Model)

A generalization of the restricted OLS $\Longrightarrow$ Stochastic linear restriction:

$$r = R\beta + v, \qquad \mathrm{E}(v) = 0 \ \text{and} \ \mathrm{V}(v) = \sigma^2\Psi$$

$$y = X\beta + u, \qquad \mathrm{E}(u) = 0 \ \text{and} \ \mathrm{V}(u) = \sigma^2 I_n$$

Using a matrix form,

$$\begin{pmatrix} y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix}\beta + \begin{pmatrix} u \\ v \end{pmatrix}, \qquad \mathrm{E}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ \text{and} \ \mathrm{V}\begin{pmatrix} u \\ v \end{pmatrix} = \sigma^2\begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}$$

For estimation, we do not need normality assumption.

Applying GLS, we obtain:

$$\begin{aligned} b &= \left((\,X' \quad R'\,)\begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1}\begin{pmatrix} X \\ R \end{pmatrix}\right)^{-1}\left((\,X' \quad R'\,)\begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1}\begin{pmatrix} y \\ r \end{pmatrix}\right) \\ &= \left(X'X + R'\Psi^{-1}R\right)^{-1}\left(X'y + R'\Psi^{-1}r\right). \end{aligned}$$

Mean and Variance of b: b is rewritten as follows:

$$b = \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} y \\ r \end{pmatrix} \right)$$

$$= \beta + \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1} \begin{pmatrix} u \\ v \end{pmatrix}$$

Therefore, the mean and variance are given by:

$$\mathrm{E}(b) = \beta \qquad \Longrightarrow \qquad b \text{ is unbiased.}$$

$$\mathrm{V}(b) = \sigma^2 \left(\begin{pmatrix} X' & R' \end{pmatrix} \begin{pmatrix} I_n & 0 \\ 0 & \Psi \end{pmatrix}^{-1} \begin{pmatrix} X \\ R \end{pmatrix} \right)^{-1}$$

$$= \sigma^2 \left(X'X + R'\Psi^{-1}R \right)^{-1}$$

9 Maximum Likelihood Estimation (MLE, 最尤法)

⟶ **Review**

1. The distribution function of $\{X_i\}_{i=1}^n$ is $f(x;\theta)$, where $x = (x_1, x_2, \cdots, x_n)$.

   θ is a vector or matrix of unknown parameters, e.g., $\theta = (\mu, \Sigma)$, where $\mu = \mathrm{E}(X_i)$ and $\Sigma = \mathrm{V}(X_i)$.

   Note that X is a vector of random variables and x is a vector of their realizations (i.e., observed data).

   Likelihood function $L(\cdot)$ is defined as $L(\theta; x) = f(x;\theta)$.

   Note that $f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$ when X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed.

The maximum likelihood estimate (MLE) of θ is the θ such that:

$$\max_{\theta} \; L(\theta; x). \qquad \Longleftrightarrow \qquad \max_{\theta} \; \log L(\theta; x).$$

Thus, MLE satisfies the following two conditions:

(a) $\dfrac{\partial \log L(\theta; x)}{\partial \theta} = 0. \quad \Longrightarrow \quad$ Solution of θ: $\tilde{\theta} = \tilde{\theta}(x)$

(b) $\dfrac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

2. $x = (x_1, x_2, \cdots, x_n)$ are used as the observations (i.e., observed data).

   $X = (X_1, X_2, \cdots, X_n)$ denote the random variables associated with the joint distribution $f(x; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$.