4 Generalized Least Squares Method (GLS, 一般化最小自乗法)

1. Regression model: $y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega)$

2. **Heteroscedasticity (不等分散，不均一分散)**

$$\sigma^2\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix}$$

First-Order Autocorrelation (一階の自己相関，系列相関)

In the case of time series data, the subscript is conventionally given by t, not i .

$u_t = \rho u_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{ iid } N(0, \sigma_\epsilon^2)$

$$\sigma^2\Omega = \frac{\sigma_\epsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix}$$

$$\mathrm{V}(u_t) = \sigma^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$$

3. The Generalized Least Squares (GLS，一般化最小二乗法) estimator of β, denoted by b, solves the following minimization problem:

$$\min_b \ (y - Xb)'\Omega^{-1}(y - Xb)$$

The GLSE of β is:

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

4. In general, when Ω is symmetric, Ω is decomposed as follows.

$$\Omega = A'\Lambda A$$

Λ is a diagonal matrix, where the diagonal elements of Λ are given by the eigen values.

A is a matrix consisting of eigen vectors.

When Ω is a positive definite matrix, all the diagonal elements of Λ are positive.

5. There exists P such that $\Omega = PP'$ (i.e., take $P = A'\Lambda^{1/2}$). $\Longrightarrow$ $P^{-1}\Omega P'^{-1} = I_n$

Multiply P^{-1} on both sides of $y = X\beta + u$.

We have:

$$y^\star = X^\star\beta + u^\star,$$

where $\quad y^\star = P^{-1}y, \quad X^\star = P^{-1}X, \quad$ and $\quad u^\star = P^{-1}u.$

The variance of $u^\star$ is:

$$\mathrm{V}(u^\star) = \mathrm{V}(P^{-1}u) = P^{-1}\mathrm{V}(u)P'^{-1} = \sigma^2 P^{-1}\Omega P'^{-1} = \sigma^2 I_n.$$

because $\Omega = PP'$, i.e., $P^{-1}\Omega P'^{-1} = I_n$.

Accordingly, the regression model is rewritten as:

$$y^\star = X^\star\beta + u^\star, \qquad u^\star \sim (0, \sigma^2 I_n)$$

Apply OLS to the above model.

Let b be as estimator of β from the above model.

That is, the minimization problem is given by:

$$\min_b \; (y^\star - X^\star b)'(y^\star - X^\star b),$$

which is equivalent to:

$$\min_{b} \; (y - Xb)'\Omega^{-1}(y - Xb).$$

Solving the minimization problem above, we have the following estimator:

$$\begin{aligned} b &= (X^{\star\prime}X^{\star})^{-1}X^{\star\prime}y^{\star} \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \end{aligned}$$

which is called GLS (Generalized Least Squares) estimator.

b is rewritten as follows:

$$b = \beta + (X^{\star\prime}X^{\star})^{-1}X^{\star\prime}u^{\star} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$$

The mean and variance of b are given by:

$$\begin{aligned} \mathrm{E}(b) &= \beta, \\ \mathrm{V}(b) &= \sigma^2(X^{\star\prime}X^{\star})^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}. \end{aligned}$$

6. Suppose that the regression model is given by:

$$y = X\beta + u, \qquad u \sim N(0, \sigma^2\Omega).$$

In this case, when we use OLS, what happens?

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$$

$$\mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Compare GLS and OLS.

(a) Expectation:

$$\mathrm{E}(\hat{\beta}) = \beta, \qquad \text{and} \qquad \mathrm{E}(b) = \beta$$

Thus, both $\hat{\beta}$ and b are unbiased estimator.

(b) Variance:

$$\mathrm{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$
$$\mathrm{V}(b) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Which is more efficient, OLS or GLS?.

$$
\begin{aligned}
\mathrm{V}(\hat{\beta}) - \mathrm{V}(b) &= \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} - \sigma^2 (X'\Omega^{-1}X)^{-1} \\
&= \sigma^2 \Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \Big) \Omega \\
&\qquad \times \Big((X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1} \Big)' \\
&= \sigma^2 A\Omega A'
\end{aligned}
$$

Note that A is $k \times n$ and Ω is $n \times n$.

Ω is the variance-covariance matrix of u, which is a positive definite matrix.

Therefore, except for $\Omega = I_n$, $A\Omega A'$ is also a positive definite matrix.

(From $\Omega = PP'$ and $A\Omega A' = AP(AP)'$, we have $xAP(xAP)' = \sum_{i=1}^{k} z_i^2 > 0$ for $x \neq 0$, where x is $1 \times k$, $z = xAP$ is $1 \times k$ and $z = (z_1, z_2, \cdots, z_k)$.)

This implies that $\mathrm{V}(\hat{\beta}_i) - \mathrm{V}(b_i) > 0$ for the ith element of β.

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

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