Generalized least-squares (GLS)

We consider a more general multiple linear regression model:

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}$$

($\underline{y}$: of size $n\times 1$
X : of size $n\times p$
$\underline{\beta}$: of size $p\times 1$
$\underline{\varepsilon}$: of size $n\times 1$
n is the number of data points)

Here, we assume

$$E(\underline{\varepsilon}) = \underline{0}, \quad \boxed{Cov(\underline{\varepsilon}) = \sigma^2 V}$$

where

σ^2 is unknown, V is known

matrix V, of size $n\times n$, represents the structure of the variances and covariances among random errors.

Note: V is non-singular and positive definite.

GLS model parameter estimator:

- Since V is non-singular and positive definite, we let

$$V = K^T K = KK$$

, where K is symmetric, square-root of V
K is also positive definite

Due to the above equality, we let

$$\underline{z} = K^{-1}\underline{y}, \quad B = K^{-1}X, \quad \underline{g} = K^{-1}\underline{\varepsilon}$$

therefore, we have $\boxed{\underline{z} = B\cdot\underline{\beta} + \underline{g}}$, where

multiplication of K^{-1} to $\underline{\varepsilon}$ "whitens" the original error terms!

implies

$$E(\underline{g}) = K^{-1}E(\underline{\varepsilon}) = \underline{0}$$

$$Cov(\underline{g}) = E\{(\underline{g}-E(\underline{g}))(\underline{g}-E(\underline{g}))^T\} = E\{(K^{-1}\underline{\varepsilon})(K^{-1}\underline{\varepsilon})^T\}$$

$$= K^{-1}\cdot E(\underline{\varepsilon}\,\underline{\varepsilon}^T)K^{-1} = \sigma^2\cdot K^{-1}VK^{-1} = \sigma^2 I$$

Now, after the transformation, the elements of $\underline{z}$ have zero-mean, constant variance σ^2, and are uncorrelated.

The model parameter estimator $\hat{\underline{\beta}}$ is obtained as:

$$\hat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}}\ S(\underline{\beta})$$

$$= \underset{\underline{\beta}}{\operatorname{argmin}}\ (\underline{z} - B\underline{\beta})^T (\underline{z} - B\underline{\beta})$$

where $(\underline{z} - B\underline{\beta})^T(\underline{z} - B\underline{\beta})$ is equivalent to $(\underline{y} - X\underline{\beta})^T V^{-1} (\underline{y} - X\underline{\beta})$.

This is because $(\underline{z} - B \cdot \underline{\beta})^T (\underline{z} - B\underline{\beta}) = (K^{-1}\underline{y} - K^{-1}X\underline{\beta})^T (K^{-1}\underline{y} - K^{-1}X\underline{\beta})$

$$= (\underline{y} - X\underline{\beta})^T K^{-1} K^{-1} (\underline{y} - X\underline{\beta})$$

$$= (\underline{y} - X\underline{\beta})^T V^{-1} (\underline{y} - X\underline{\beta}).$$

Now, we solve $\hat{\underline{\beta}}$ from the following minimization problem:

$$\hat{\underline{\beta}} = \underset{\underline{\beta}}{\operatorname{argmin}}\ (\underline{y} - X\underline{\beta})^T V^{-1} (\underline{y} - X\underline{\beta})$$

Similarly, we take the derivative of the cost function w.r.t. $\underline{\beta}$ and set it equal to zero, i.e.,

$$\frac{\partial (\underline{y} - X\underline{\beta})^T V^{-1} (\underline{y} - X\underline{\beta})}{\partial \underline{\beta}} = 0 \Rightarrow \boxed{\hat{\underline{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{y}}$$

The properties of $\hat{\beta}$ are as follows:

- $E(\hat{\beta}) = (X^T V^{-1} X)^{-1} X^T V^{-1} E(\underline{y}) = (X^T V^{-1} X)^{-1} X^T V^{-1} X \underline{\beta} = \underline{\beta}$

This is due to $E(\underline{\varepsilon}) = \underline{0}$

- $\mathrm{Cov}(\hat{\beta}) = E\{(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))^T\}$

$$= E\{(X^T V^{-1} X)^{-1} X^T V^{-1} \underline{\varepsilon} \cdot \underline{\varepsilon}^T V^{-1} X (X^T V^{-1} X)^{-1}\}$$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} E(\underline{\varepsilon}\,\underline{\varepsilon}^T) V^{-1} X (X^T V^{-1} X)^{-1}$$

$$= \sigma^2 \cdot (X^T V^{-1} X)^{-1}$$

This is because $E(\underline{\varepsilon}\,\underline{\varepsilon}^T) = \sigma^2 V$.

★ When we additionally assume $\underline{\varepsilon} \sim N(\underline{0}, \sigma^2 V)$, then we have as well

$$\hat{\beta} \sim N(\underline{\beta}, \sigma^2 (X^T V^{-1} X)^{-1})$$

The parameter estimator of σ^2:

- First, define SS_{Res} to be:

$$SS_{Res} = (\underline{z} - \hat{\underline{z}})^T (\underline{z} - \hat{\underline{z}}) = (K^{-1}\underline{y} - K^{-1} X \hat{\beta})^T (K^{-1}\underline{y} - K^{-1} X \hat{\beta})$$

$$= (K^{-1}\underline{y} - K^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{y})^T (K^{-1}\underline{y} - K^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{y})$$

$$= \underline{y}^T \underbrace{(K^{-1} - K^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1})^T (K^{-1} - K^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1})}_{A} \underline{y}$$

It can be easily verified that

$$A = V^{-1} - V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1}$$ (to be shown on WB)

We know that :

- If A is a $k \times k$ matrix of constants, and $\underline{y}$ is a $k \times 1$ random vector with mean $\underline{\mu}$ and non-singular covariance matrix Σ, then.

$$E(\underline{y}^T A \underline{y}) = \text{trace}(A\Sigma) + \underline{\mu}^T A \underline{\mu}$$

Normal distribution is NOT assumed.

We apply the above result to

$$E(SS_{Res}) = tr\left(\sigma^2 V \cdot (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})\right)$$

$$+ (X\underline{\beta})^T (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}) X\underline{\beta}$$

$$= \sigma^2 \cdot \text{trace}(I - X(X^TV^{-1}X)^{-1}X^TV^{-1})$$

$$+ \underline{\beta}^T (X^TV^{-1}X - X^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}X)\underline{\beta}$$

$$= \sigma^2 \cdot \left(tr(I_n) - tr(X(X^TV^{-1}X)^{-1}X^TV^{-1})\right)$$

$$= \sigma^2 (n-p)$$

And we let $MS_{Res} = \dfrac{SS_{Res}}{n-p}$ to be an estimator of σ^2. It can easily shown with the aid of the above result :

$$E(MS_{Res}) = \sigma^2. \longrightarrow \text{unbiased estimator of } \sigma^2$$

The explicit expression of MS_{Res} is :

$$\boxed{MS_{Res} = \frac{\underline{y}^T (V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1})\underline{y}}{n-p}}$$

Generalized Gauss-Markov Theorem:

$$\hat{\underline{\beta}} = (X^T V^{-1} X)^{-1} X^T V^{-1} \underline{y}$$ is the BLUE estimator

when $E(\underline{\varepsilon}) = \underline{0}$, and $cov(\underline{\varepsilon}) = \sigma^2 V$.

Yet another Version:

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} \underline{y}$$ is the BLUE estimator

when $E(\underline{\varepsilon}) = 0$ and $Cov(\underline{\varepsilon}) = \Sigma$

proof omitted! $\rightarrow$ see textbook Appendix C.11 [optional]

Two special cases:

① $V = I$, this leads to "ordinary" LS.

② $V = diag\left(\frac{1}{w_1}, \frac{1}{w_2}, \cdots \frac{1}{w_n}\right)$, with $w_i > 0 \ \forall i = 1, 2, \dots n$ $\Rightarrow$ uncorrelated error but with non-constant variance!

The second special case is called "weighted LS" (WLS). in the textbook.

For this case, we could simply let $W = V^{-1} = diag(w_1, w_2, \cdots w_n)$, then

$$\hat{\underline{\beta}} = (X^T W X)^{-1} X^T W \underline{y}$$

and the transformation involves:

$$B = K^{-1} X = \begin{bmatrix} \sqrt{w_1} & & \\ & \sqrt{w_2} & \\ & & \ddots \\ & & & \sqrt{w_n} \end{bmatrix} X, \quad \underline{z} = K^{-1} \underline{y} = \begin{bmatrix} y_1 \sqrt{w_1} \\ y_2 \sqrt{w_2} \\ \vdots \\ y_n \sqrt{w_n} \end{bmatrix}$$