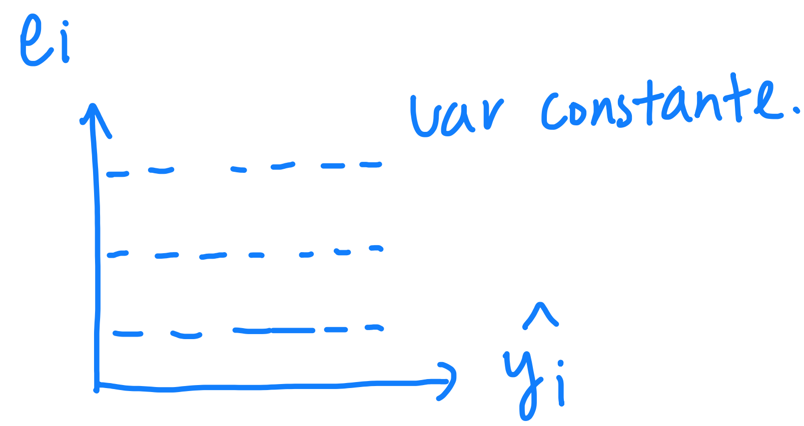
Chapiter 6 Heteroscedasticity (Unequal variance)

One of the great values of Gauss-Markov theorem is that it provides conditions which, if they hold, assure us that least squares is a good procedure. These conditions can be checked and if we find that one or more of them are seriously violated, we can take action that will cause at least approximate compliance.

This chapiter is devoted to the second G-M condition, which states that $var(\varepsilon_i) = var(y_i) = \sigma^2$. Violation of this condition is often called heteroscedasticity, which compliance is referred to as homoscedasticity.

If there is heteroscedasticity:

$$E(\hat{\beta}) = \beta, \quad cov(\hat{\beta}) = (X'X)^{-1}X' \underbrace{cov(y)}_{\neq \sigma^2 I.} X(X'X)^{-1} \neq \sigma^2 (X'X)^{-1}.$$



Detecting Heteroscedasticity.

1. We can determine if heteroscedasticity is likely to be present from an understanding of the underlying situation and also determine what corrective measures might be taken.

ex: if $y \sim$ Poisson, $var(\varepsilon_i) = var(y_i) = \sigma_i^2 = E(y_i)$

if $y_i = \frac{X_i}{n_i}$, $X \sim$ binomial, $var(y_i) = \sigma_i^2 = \frac{p_i(1-p_i)}{n_i} = \frac{[E(y_i)][1-E(y_i)]}{n_i}$

2. Another way of checking to see if heteroscedasticity is present is through plots.

If $\sigma_i^2 = var(\varepsilon_i)$ varies with $E(y_i)$, a plot of the residuals against the $\hat{y}_i$'s might show the residuals e_i to be more spread out. for some values of $\hat{y}_i$ than for others.

standard residuals. $e_i^{(s)} = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$

Rstudent: $e_i^{*} = \dfrac{e_i}{S_{(i)}\sqrt{1-h_{ii}}}$

are used to detect the presence of heteroscedasticity.

Variance Stabilizing Transformation.

When heteroscedasticity occurs we can take one of two types of actions to make the σ_i's approximate equal.

1. One consists of transforming y_i when $var(y_i)$ depends on $E(y_i)$
2. The other involves weighting the regression. (Weighted least squares)

Transformation of dependent variable.

For any function $f(y)$ of y with continuous first derivative $f'(y)$ and finite second derivative $f''(y)$, we know from elementary calculus that

$$f(y_i) - f(\mu_i) = (y_i - \mu_i) f'(\mu_i) + \frac{1}{2} (y_i - \mu_i)^2 f''(\theta)$$

where θ lies between y_i and μ_i. $\mu_i = E(y_i)$

Thus, when $(y_i - \mu_i)^2$ is small, we have

$$f(y_i) - f(\mu_i) \approx f'(\mu_i)(y_i - \mu_i) \quad (*)$$

Squaring and taking expectations of both side of $(*)$, we get approximately,

$$var(f(y_i)) \approx (f'(\mu_i))^2 \sigma_i^2(\mu_i), \quad \sigma_i^2(\mu_i) = var(y_i)$$

Thus, in order to find a suitable transformation f of y_i which would make $var(f(y_i))$ approximately a constant, we need to solve the equation:

$$f'(\mu_i) = \frac{c}{\sigma_i(\mu_i)}$$

where c is any constant. Such a transformation f is called a variance stabilizing transformation.

ex: let $y_i \sim$ Poisson, then $\sigma_i^2(\mu_i) = \mu_i$

$$\Rightarrow f'(\mu_i) = \frac{c}{\sqrt{\mu_i}}$$

$$\frac{df}{d\mu} = \frac{c}{\sqrt{\mu}} \Rightarrow df = c \cdot \mu^{-\frac{1}{2}} \cdot d\mu.$$

$$\Rightarrow f = c \int \mu^{-\frac{1}{2}} d\mu = 2c\mu^{\frac{1}{2}} + C_1 = 2c\sqrt{\mu} + C_1.$$

In practice, we can use $f(y) = \sqrt{y}$ with $y > 0$ to stabilize relatively the variance.

ex: $y_i = \frac{X_i}{n_i} \sim$ Binomial $\qquad \mu_i = E(y_i)$

$$\sigma_i^2(\mu_i) = \frac{\mu_i(1-\mu_i)}{n_i}$$

$$f'(\mu_i) = \frac{c}{\sigma_i(\mu_i)} = \frac{c}{\sqrt{\frac{\mu_i(1-\mu_i)}{n_i}}} = \frac{c\sqrt{n_i}}{\sqrt{\mu_i(1-\mu_i)}}$$

$$\Rightarrow f(\mu) = c\sqrt{n_i} \int \frac{d\mu}{\sqrt{\mu(1-\mu)}} \Rightarrow f(\mu) = 2c\sqrt{n_i}\arcsin(\sqrt{\mu}) + C_1.$$

In practice, we use $f(y) = \sqrt{n} \arcsin(\sqrt{y})$.

ex: If $\sigma_i = \mu_i$, then.

$$f'(\mu_i) = \frac{c}{\sigma_i} = \frac{c}{\mu_i}$$

$$\Rightarrow f(\mu) = c \int \frac{1}{\mu} \, d\mu = c \ln(|\mu|) + c_1.$$

In practice, we use $f(y) = \ln y$ with $y > 0$.

⊛ Box-cox transformation:

$$f(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \in \mathbb{R},\ \lambda \neq 0 \\ \ln(y) & \text{if } \lambda \to 0 \end{cases}$$

Weighting

Suppose $var(\varepsilon_i) = \sigma_i^2 = c_i^2 \sigma^2$ where c_i^2 are known constants. Then constancy of variance can also be achieved by dividing both sides of each of the equations of the regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

by c_i, i.e.

$$\frac{y_i}{c_i} = \frac{\beta_0}{c_i} + \frac{\beta_1 x_{i1}}{c_i} + \cdots + \frac{\beta_k x_{ik}}{c_i} + \frac{\varepsilon_i}{c_i} \quad (1)$$

is clearly homoscedastic. Each $w_i = \dfrac{1}{c_i^2}$ is called a weight, the nomenclature coming from the fact that now we are minimizing a weighted sum of squares,

$$\sum w_i (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 \quad (2)$$

The estimate of β obtained from the model (1), by minimizing (2) is called a weighted least squares (WLS) estimate of β. When $C_i = 1$, i.e. when least squares is not "weighted", we call it ordinary least squares (OLS).

Theorem. $y = X\beta + \varepsilon$. $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 \Omega$

$\Omega = diag(C_1^2, C_2^2, \dots, C_n^2)$, C_i^2 known constant.

1. weighted least squares estimate of β, $\hat{\beta}_{WLS} =$

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

2. $Cov(\hat{\beta}_{WLS}) = \sigma^2 (X' \Omega^{-1} X)^{-1}$.

3. residuals

$$e_i^{(\Omega)} = \sqrt{w_i}\,(y_i - \hat{y}_{i,WLS})$$

4. unbiased estimate of σ^2:

$$s^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (e_i^{(\Omega)})^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} w_i (y_i - \hat{y}_{i,WLS})^2.$$

with $w_i = \frac{1}{c_i^2}$. $\hat{y}_{i,WLS} = x_i' \cdot \hat{\beta}_{WLS}$.

5. $var(a'\hat{\beta}_{WLS}) \le var(a'\hat{\beta})$

proof:

(1) $y = X\beta + \varepsilon$. $E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 \Omega$

$\Omega = diag(c_1^2, \cdots, c_n^2)$

$y^{(\Omega)} = X^{(\Omega)}\beta + \varepsilon^{(\Omega)}$

où $y^{(\Omega)} = Ay$, $X^{(\Omega)} = AX$, $\varepsilon^{(\Omega)} = A\varepsilon$.

$A = diag\left(\frac{1}{c_1}, \frac{1}{c_2}, \cdots \frac{1}{c_n}\right)$

$= diag(\sqrt{w_1}, \sqrt{w_2}, \cdots \sqrt{w_n})$

$= \Omega^{-\frac{1}{2}}$

$E(\varepsilon^{(\Omega)}) = 0$. $cov(\varepsilon^{(\Omega)}) = \sigma^2 I_n$. $\beta = (X'X)^{-1}X'y$.

$\hat{\beta}_{WLS} = (X'^{(\Omega)} X^{(\Omega)})^{-1} X'^{(\Omega)} y^{(\Omega)}$

$= (X'A'AX)^{-1}(X'A')Ay$

$$A'A = \mathrm{diag}\left(\frac{1}{c_1^2}, \cdots \frac{1}{c_n^2}\right) = \Omega^{-1}$$

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

(2) $$\mathrm{Cov}(\hat\beta_{WLS}) = \sigma^2 (X'^{(\Omega)} X^{(\Omega)})^{-1}$$

$$= \sigma^2 (X'A'AX)^{-1}$$

$$= \sigma^2 (X'\Omega X)^{-1}.$$

(3) $$e^{(\Omega)} = y^{(\Omega)} - \hat y^{(\Omega)} = Ay - AX\hat\beta_{WLS}$$

$$= A(y - X\hat\beta_{WLS}) \qquad A = \mathrm{diag}(\sqrt{w_1}, \cdots \sqrt{w_n})$$

$$\Rightarrow e_i^{(\Omega)} = \sqrt{w_i}\,(y_i - \hat y^{(\Omega)}_{WLS})$$

$$E(e'^{(\Omega)} e^{(\Omega)}) = E(y'^{(\Omega)} M'^{(\Omega)} M^{(\Omega)} y^{(\Omega)})$$

où $M^{(\Omega)} = I_n - X^{(\Omega)} (X'^{(\Omega)} X^{(\Omega)})^{-1} X'^{(\Omega)}$

$$e^{(\Omega)} = M^{(\Omega)} . y^{(\Omega)}$$

$$\because M^{(\Omega)} = M'^{(\Omega)} M^{(\Omega)}$$

$$\Rightarrow E(e'^{(\Omega)} e^{(\Omega)}) = E(y'^{(\Omega)} M^{(\Omega)} y^{(\Omega)})$$

$$= tr[M^{(\Omega)} E(y^{(\Omega)} y'^{(\Omega)})] = tr(M^{(\Omega)} \sigma^2 I_n)$$

$$= \sigma^2 tr(M^{(\Omega)}) = \sigma^2 (n-k-1)$$

$$\Rightarrow E(s^2) = \frac{e'^{(\Omega)} e^{(\Omega)}}{n-k-1} = \sigma^2$$

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

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

$$= \sigma^2 (X'X)^{-1} X' \Omega X (X'X)^{-1}.$$

Gauss-Markov, $\Rightarrow var(c'\hat{\beta}_{WLS}) \leq var(c'\hat{\beta})$.