Weighted Least Squares

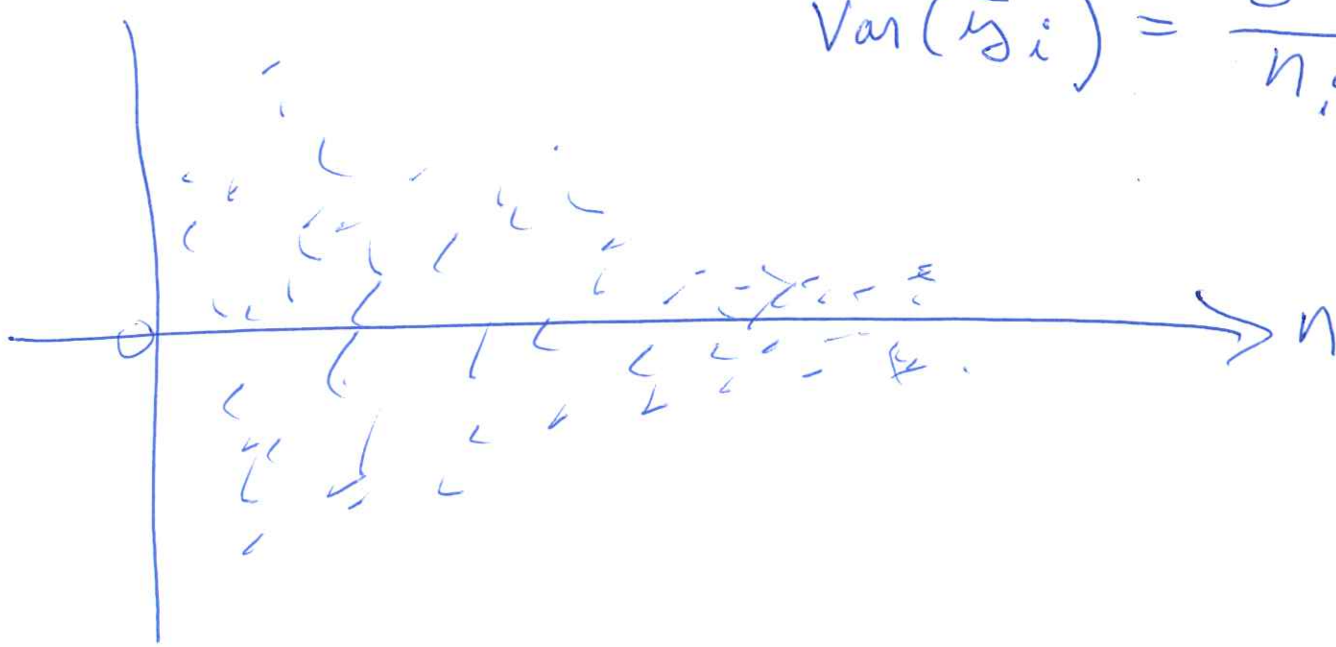
An antidote to unequal variance (of a certain kind.)

<u>Ex</u> Aggregated data. Teaching evaluations. Have

- $\bar{y}_1, \ldots, \bar{y}_m$
- $n_1, \ldots, n_n$
- Lots of IVs

Residual plot

$$\mathrm{Var}(\bar{y}_i) = \frac{\sigma^2}{n_i}$$



Have $y = X\beta + \varepsilon$, with $\underset{\Omega}{\Sigma}$

$$\mathrm{Cov}(\varepsilon) = \begin{pmatrix} \sigma^2/n_1 & & & 0 \\ & \sigma^2/n_2 & & \\ & & \ddots & \\ 0 & & & \sigma^2/n_m \end{pmatrix} = \sigma^2 \begin{pmatrix} 1/n_1 & & & 0 \\ & 1/n_2 & & \\ & & \ddots & \\ 0 & & & 1/n_n \end{pmatrix}$$

Generalize: $y = X\beta + \varepsilon$, $\mathrm{Cov}(\varepsilon) = \sigma^2 \Omega$

omega

Ω $m \times m$ known symmetric positive definite matrix.

GENERALIZED LEAST SQUARES

Transform the data

$$\underbrace{\Omega^{-1/2} y}_{y^*} = \underbrace{\Omega^{-1/2} X}_{X^*} \beta + \underbrace{\Omega^{-1/2} \varepsilon}_{\varepsilon^*}$$

same β

$$\mathrm{Cov}(\Omega^{-1/2}\varepsilon) = \Omega^{-1/2} \sigma^2 \Omega \, \Omega^{-1/2'}$$
$$= \sigma^2 I_m$$

3

An amazing Scalar example with no independent variables

$y_{ij} \overset{iid}{\sim} ?(\mu, \sigma^2)$ Have

$\bar{y}_1, \ldots, \bar{y}_m \quad \bar{y}_j \;\dot{\sim}\; N\left(\mu, \frac{\sigma^2}{n_j}\right)$

Want to estimate μ

$$\hat{\mu}_1 = \frac{1}{m}\sum_{j=1}^{m} \bar{y}_j \quad E(\hat{\mu}_1) = \mu$$

$$Var(\hat{\mu}_1) = \frac{\sigma^2}{m^2}\sum_{j=1}^{m}\frac{1}{n_j}$$

$$L = \frac{1}{m}\sum_{j=1}^{m} c_j \bar{y}_j$$

unbiased iff $\sum_{j=1}^{m} c_j = 1$

Want the BEST

(B.L.U.E.)

Try Weighted Least squares (4)

$$\bar{y}_i = 1\!\!\!/\,\mu + \varepsilon_i, \quad \varepsilon_i$$

β_0

$$E(\varepsilon_i) = 0$$
$$Var(\varepsilon_i) = \frac{\sigma^2}{n_i}$$

$$\Longrightarrow \sqrt{n_i}\,\bar{y}_i = \sqrt{n_i}\,1\!\!\!/\,\mu + \sqrt{n_i}\,\varepsilon_i$$

$$y_i = x_i \beta + \varepsilon_i$$
$$Var(\varepsilon_i) = \sigma^2$$

$$\begin{pmatrix} \sqrt{n_1} & & & 0 \\ & \sqrt{n_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{n_m} \end{pmatrix} \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_m \end{pmatrix} = \begin{pmatrix} \sqrt{n_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{n_m} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \mu$$

$$+ \begin{pmatrix} \sqrt{n_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{n_m} \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

5

$$\sqrt{n_i}\,\bar{y}_i = \sqrt{n_i}\,\overset{\mu}{\overset{\shortparallel}{\beta}} + \sqrt{n_i}\,\varepsilon_i$$

$$y_i^* = x_i^*\,\beta + \varepsilon_i^*$$

$$b_{GLS} = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}}$$

$$= \frac{\sum_{i=1}^{m} n_i \bar{y}_i}{\sum_{i=1}^{m} n_i}$$