Heteroscedasticity

Additional note

Difference between OLS and GLS

OLS we minimize

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

GLS we minimize

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2, w_i = 1/\sigma_i^2$$

- In GLS we minimized a weighted sum of residual squares with $w_i = \frac{1}{\sigma^2}$ acting as the weights, but in OLS we minimize an unweighted or equally weighted residual sum of squares (RSS)

- In GLS the weight assigned to each observation is inversely proportional to its σ_i, that is, observations coming from a population with larger σ_i will get relatively smaller weight and those from a population with smaller σ_i will get proportionately larger weight in minimizing the RSS

To see the difference between OLS and GLS clearly, consider the hypothetical scattergram

- In the (unweighted) OLS each $\hat{u}_i^2$ associated with points A, B, and C will receive the same weight in minimizing the RSS.
- Obviously, in this case the $\hat{u}_i^2$ associated with point C will dominate the RSS.
- But in GLS the extreme observation C will get relatively smaller weight than the other two observations.
- In estimating the PRF more reliably we would like to give more weight to observations that are closely clustered around their (population) mean than to those that are widely scattered about.

Hypothetical scattergram