Heteroskedasticity

1. Consequences
2. Heteroskedasticity-Robust Inference
3. Testing
4. Weighted Least Squares

Consequences

OLS estimators $\hat{\beta}_j$ still unbiased for β_j

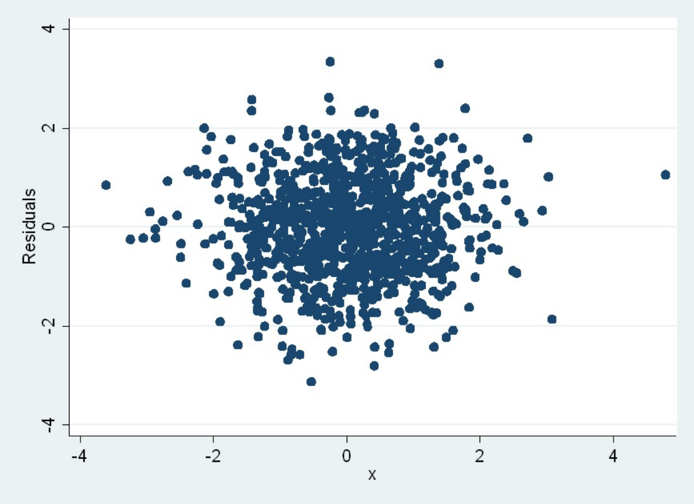
- Model

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- (MLR.5) Homoskedasticity - violated
  - $Var(u|x_1, ..., x_k)$ $\neq \sigma^2$ but depends on x_j
- Under MLR.1-MLR.4
- Usual standard errors and test statistics - no longer valid

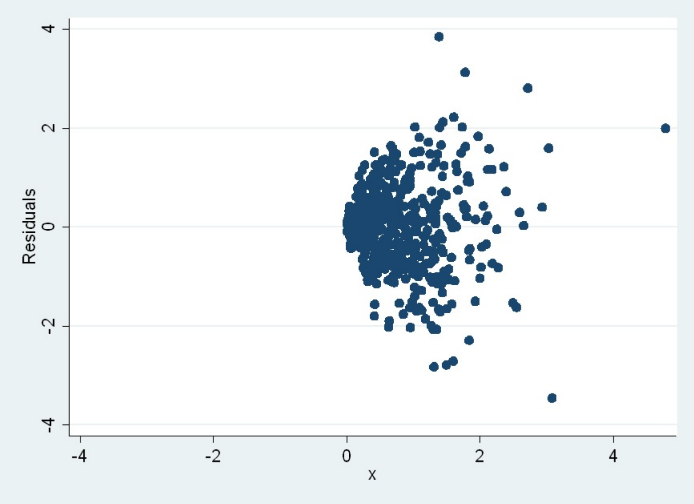
Consequences (cont.)

Homoskedasticity example



Consequences (cont.)

Heteroskedasticity example



Heteroskedasticity-Robust Inference

- Heteroskedasticity of unknown form
  - Relationship between $Var(u|x_1, ..., x_k)$ and x_j - unknown
- Heteroskedasticity-robust standard errors
  - Works only in case of large samples

Heteroskedasticity-Robust Inference (cont.)

std. errors of slope

- Simple regression
  - Under homoskedasticity

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{SST_x}}$$

where $SST_x = \sum_{i=1}^{n}(x_i - \overline{x})^2$

  - Heteroskedasticity-robust

$$se(\hat{\beta}_1) = \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \widehat{u}_i^2}}{SST_x}$$

Heteroskedasticity-Robust Inference (cont.)

- Multiple regression
  - Under homoskedasticity

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1 - R_j^2)}}$$

where $SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$ and R_j^2: R^2 from the regression of x_j on all other explanatory variables

  - Heteroskedasticity-robust

$$se(\hat{\beta}_j) = \frac{\sqrt{\sum_{i=1}^{n} \widehat{r}_{ij}^2 \widehat{u}_i^2}}{SST_j(1 - R_j^2)}$$

where $\widehat{r}_{ij}$: residual from the regression of x_j on all other explanatory variables for observation i

Heteroskedasticity-Robust Inference (cont.)

Haiku by Keisuke Hirano

T-stat looks too good.
Use robust standard errors-
significance gone.

https://keihirano.github.io/haiku.html

Testing

- Model (assuming MLR.1-MLR.4)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- Null hypothesis of homoskedasticity

$$H_0 : Var(u|x_1, \ldots, x_k) = \sigma^2$$

- Equivalently

$$H_0 : E(u^2|x_1, \ldots, x_k) = \sigma^2$$

Testing (cont.)

- Suppose

$$u^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \nu$$

- Null of homosk. $\Rightarrow$

$$H_0 : \delta_1 = \ldots = \delta_k = 0$$

- Can estimate

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \ldots + \delta_k x_k + \nu$$

- Joint test in H_0
  - Works only in case of large samples
  - Can also include squares and interactions or fitted values

$x_1^2, x_2^2 \ldots$ $x_1 x_2$ $x_1 x_k$ $\hat{y}$

Weighted Least Squares

- Model

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i$$

- Suppose

$$Var(u_i | x_{i1}, \ldots, x_{ik}) = \sigma^2 h_i$$

- Can transform the equation into

$$\frac{y_i}{\sqrt{h_i}} = \beta_0 \frac{1}{\sqrt{h_i}} + \beta_1 \frac{x_{i1}}{\sqrt{h_i}} + ... + \beta_k \frac{x_{ik}}{\sqrt{h_i}} + \frac{u_i}{\sqrt{h_i}}$$

$Var(u_i | x_{i1}, \ldots, x_{ik}) = \sigma^2 h_i$ and thus $Var(\frac{u_i}{\sqrt{h_i}} | x_{i1}, \ldots, x_{ik}) = \frac{\sigma^2 h_i}{h_i} = \sigma^2$

- Generalized least squares (GLS) or weighted least squares (WLS)
- Feasible or estimated GLS (FGLS or EGLS) $\because$ h must be estimated