Tutorial: Hetroscedasticity

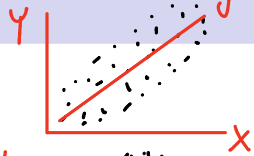
Feedback Form: tiny.cc/hammadfeedback

Hammad Shaikh

November 10, 2019

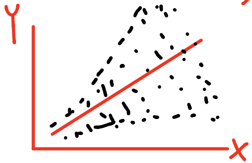
Heteroskedasticity Definition

Homoscedasticity:



- $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon$

Hetroscedastic



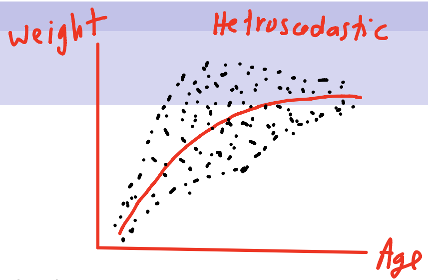
- Homoscedasticity: ϵ has constant variance in all covariates

  ↳ $V(\varepsilon | x) = \sigma^2$ (constant)

- Hetroscedasticity: $V(\epsilon)$ depends on covariates

  ↳ $V(\varepsilon | x) = f(x) \neq$ constant but depends on X

Heteroskedasticity Example



- Suppose we regress weight (lbs) on age
  - $Weight = \beta_0 + \beta_1 Age + \epsilon$

- Is the error term, ϵ, homoskedastic?

↳ Variance in weight changes dramatically as people age $\Rightarrow V(\epsilon | age) \neq$ constant

$\Rightarrow$ Hetroscedastic

Consequences of Heteroskedasticity

- Suppose $V(\epsilon_i|x_i) = \sigma_i^2$ (hetroscedastic)
  - OLS estimator given other assumptions hold is still unbiased

    $\hookrightarrow E(\hat{\beta}|X) = \beta$
  - OLS estimator is inefficient

    $\hookrightarrow V(\hat{\beta}|X)$ is no longer smallest when compared to other unbiased estimator
  - t-statistics and F-statistics not as reliable

    $\hookrightarrow t_{stat} = \dfrac{\hat{\beta}}{SE(\hat{\beta})}$

    $H_0: \beta = 0$

    $H_1: \beta \neq 0$

Under Hetrosced. $\Rightarrow$ $SE(\hat{\beta})$ will be wrong $\Rightarrow$ t_{stat} wrong

Correcting of Heteroskedasticity

Doesn't solve effiency loss ($\uparrow Var(\hat{\beta}|X)$)

- Robust standard errors
  - Helps correct the standard errors

    ↳ Alternative formular for SE under Hetro.

    ↳ Reliable Hypo. test

- Generalized Least Squares (GLS)
  - Efficient estimator if GLS assumptions satisfied

    ↳ Solves problem but requires strong assumption

Robust Standard Errors

$SE(\hat{\beta}_1|x) = \sqrt{V(\hat{\beta}_1|x)}$

Homoscedastic $\Rightarrow$ $Var(\hat{\beta}_1|x) = \dfrac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$

- Simple linear regression: $y = \beta_0 + \beta_1 x_i + \epsilon_i$

if $\sigma_i^2 = \sigma^2$

- Suppose $V(\epsilon_i|x_i) = \sigma_i^2$ (hetroscedastic)
  - Possible to derive $V(\hat{\beta}) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2\sigma_i^2}{\left[\sum_{i=1}^{n}((x_i-\bar{x})^2\right)^2}$

$V(\varepsilon_i|x_i) = E(u_i^2|x_i) - [E(u_i|x_i)]^2 = E(u_i^2|x_i) = \sigma_i^2$

- Obtain robust standard error involves estimate σ_i^2 using $\hat{u}_i^2$

Hetro $\Rightarrow$ $V_{robust}(\hat{\beta}|x) = \dfrac{\sum(x_i-\bar{x})^2 \cdot \hat{u}_i^2}{\left[\sum(x_i-\bar{x})^2\right]^2} \Rightarrow SE_{robust}(\hat{\beta}) = \sqrt{V_{robust}(\hat{\beta}|x)}$

Generalized Least Squares (GLS)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \Rightarrow \frac{Y_i}{\sqrt{g(x_i)}} = \frac{\beta_0}{\sqrt{g(x_i)}} + \beta_1 \frac{X_i}{\sqrt{g(x_i)}} + \tilde{\varepsilon}_i$$

$$\Downarrow$$

- Suppose $V(\epsilon_i|x_i) = \sigma_i^2$ (hetroscedastic)

$$V(\tilde{\varepsilon}_i|X_i) = \sigma^2 \Leftarrow \tilde{Y}_i = \tilde{\beta}_0 + \beta_1 \tilde{X}_i + \tilde{\varepsilon}_i$$

- Suppose we further assume $V(\epsilon_i|x_i) = \sigma^2 * g(x_i)$

$$\hookrightarrow \tilde{\varepsilon}_i = \frac{\varepsilon_i}{\sqrt{g(x_i)}} \Rightarrow V(\tilde{\varepsilon}_i|X_i) = \frac{1}{g(x_i)} \cdot V(\varepsilon_i) = \sigma^2$$

- Can we apply some transformation to ϵ_i and recover homoscedasticity?

$$\hookrightarrow \tilde{\varepsilon}_i = h(\varepsilon_i) \text{ s.t } V(\tilde{\varepsilon}_i|X_i) = \sigma^2$$

GLS Example (Dec 2017, Exam)

$$\tilde{Y} = \beta_0 \cdot \tilde{X}_0 + \beta_1 \tilde{X}_1 + \beta_2 \tilde{X}_2 + \tilde{U}$$

- $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$
- Assume $V(U|X_1, X_2) = \sigma^2 \times e^{X_2}$

$$= \sigma^2$$

$$\tilde{U} = \frac{U}{\sqrt{e^{X_2}}} \Rightarrow V(\tilde{U}|X_1, X_2) = \frac{1}{e^{X_2}} \cdot V(U|X_1, X_2)$$

- Transform model so it has homoscedastic error

$$\frac{Y}{\sqrt{e^{X_2}}} = \frac{\beta_0}{\sqrt{e^{X_2}}} + \beta_1 \frac{X_1}{\sqrt{e^{X_2}}} + \beta_2 \frac{X_2}{\sqrt{e^{X_2}}} + \tilde{U}$$