Lecture 16

Quantitative Political Science

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/11/02

Slides Updated: 2023-12-23

Agenda

1. Variance of OLS estimators
2. Heteroskedasticity

Recap

- We went over 4 assumptions to characterize the bias of our OLS estimators

1. Relationship between x and y is **linear in its parameters**
2. x and y are drawn from a random sample, making them **i.i.d.**
3. $VAR(X) \neq 0$
4. $E(u|x) = 0$

- With these, we demonstrated that $\hat{\beta}_0$ and $\hat{\beta}_1$ are **unbiased** for β_0 and β_1

Sampling Distributions

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are statistics, just like $\bar{Y}$
- When we evaluate statistics, we care about both their **bias** (last class) and their **variance**
  - How far can we expect them to be from their true value (i.e., the population parameter) on average?
- In the univariate case, we were interested in the **sampling distribution** of $\bar{Y}$
- Here, we are also interested in the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

Variance

- We already know the means of $\hat{\beta}_0$ and $\hat{\beta}_1$: they are β_0 and β_1 (from last class)
- To compute the **variances** of $\hat{\beta}_0$ and $\hat{\beta}_1$, we need a **fifth assumption**

Assumption 5: $VAR(u|x) = \sigma^2$

- The error has the **same variance** regardless of the value of x
- This is known as **homoskedasticity**
- If this fails, we have **heteroskedasticity**

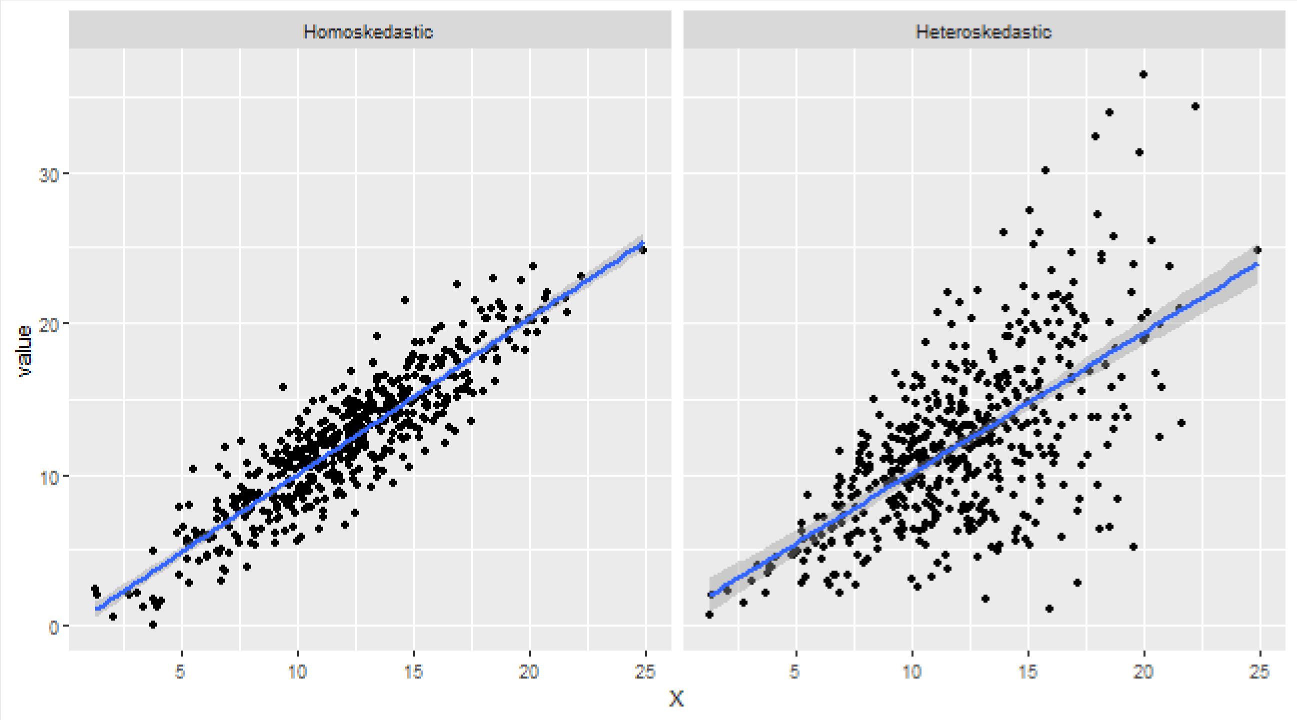
Variance

```r
require(tidyverse)
set.seed(123)
X <- rnorm(500,mean = 12,sd = 4)
Y <- rnorm(500,mean = X,sd = X/3)
Y2 <- rnorm(500,mean = X,sd = 2)

p <- data.frame(X = X,Heteroskedastic = Y,Homoskedastic = Y2) %>%
  gather(outcome,value,-X) %>%
  mutate(outcome = factor(outcome,levels = c('Homoskedastic','Heteroskedastic'))) %>%
  ggplot(aes(x = X,y = value)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  facet_wrap(~outcome)
```

Variance

p



Homoskedastic
Heteroskedastic
value
30
20
10
0
5
10
15
20
25
X

Variance

- Note the difference between Assumptions 4 and 5!
  - $E(u|x) = 0$
  - $VAR(u|x) = \sigma^2$
- We don't need assumption 5 for unbiasedness, but we do for variance!
- What is $VAR(y|x)$?

Variance

$$\begin{aligned} VAR(y|x) &= VAR(\beta_0 + \beta_1 x + u|x) \\ &= VAR(\beta_0|x) + VAR(\beta_1 x|x) + VAR(u|x) \\ &= 0 + 0 + \sigma^2 \\ &= \sigma^2 \end{aligned}$$

- What is σ^2?
- It is a measure of the **extent to which unexplained factors are affecting** y
  - These factors are not related to x (from assumption 4)
  - These factors are constant regardless of x (from assumption 5)
  - When σ^2 is big, it means that other factors explain a lot of variation in y beyond just x
  - When σ^2 is small, it means that x explains a lot of variation in y
- Note that σ^2 is a **parameter**, something that exists in the population

Variance of estimators

- $VAR(\hat{\beta}_0) = \frac{\sigma^2 \frac{\sum x_i^2}{n}}{SST_x}$ and $VAR(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$
- I'll leave it to you to prove $VAR(\hat{\beta}_0)$, but let's dig into $VAR(\hat{\beta}_1)$

$$
\begin{aligned}
\hat{\beta}_1 &= \beta_1 + \frac{\sum(x_i - \bar{x})u_i}{SST_x} \\
VAR(\hat{\beta}_1|x) &= VAR\left[\beta_1 + \frac{\sum(x_i - \bar{x})u_i}{SST_x} \;\middle|\; x\right] \\
&= VAR(\beta_1|x) + \frac{1}{SST_x^2}\sum(x_i - \bar{x})^2 VAR(u_i \mid x) \\
&= 0 + \frac{SST_x}{SST_x^2} VAR(u_i|x) \\
&= \frac{\sigma^2}{SST_x}
\end{aligned}
$$

Sampling Variance of $\hat{\beta}_1$

- So $VAR(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$
- Want this to be as small as possible (bias-variance tradeoff)
- As σ^2 gets smaller, so does $VAR(\hat{\beta}_1)$
- As SST_x gets bigger, $VAR(\hat{\beta}_1)$ gets smaller
- Unpack SST_x for more insights!

Sampling Variance of $\hat{\beta}_1$

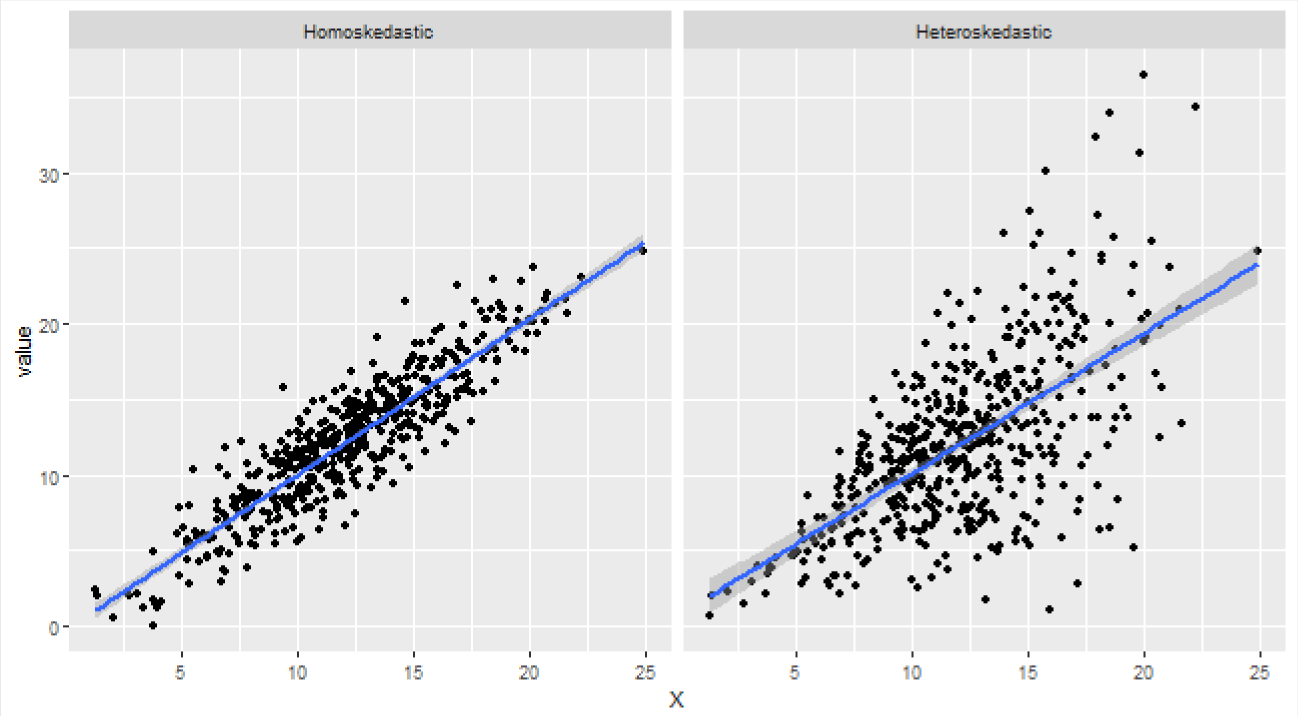
$$\begin{aligned}
SST_x &= \sum (x_i - \bar{x})^2 \\
\frac{SST_x}{n} &= \frac{\sum (x_i - \bar{x})^2}{n} \\
&= var(x) \text{ sample variance of } x \\
SST_x &= n * var \\
VAR(\hat{\beta}_1) &= \frac{\sigma^2}{n * var(x)}
\end{aligned}$$

- What can we actually manipulate here as a researcher?
  - σ^2 is a parameter: it declines when x explains y well. But we don't have a lot of control over this.
  - $var(x)$ is the empirical variance of x in our sample. It approximates the population variance, but we don't have a ton of control over this either.
  - n: we choose this!

Heteroskedasticity

- The preceding results rely on the assumption of $VAR(u|x) = \sigma^2$
- What if this doesn't hold?

p



Heteroskedasticity

- We would then say that the variance of the errors conditional on x is specific to that unit
  - $VAR(u_i|x_i) = \sigma_i^2$
- Recall that $\hat{\beta}_1 = \frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}$ and that

$$\begin{aligned} VAR(\hat{\beta}_1) &= VAR\left[\frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}\right] \\ &= \frac{1}{SST_x^2}\sum(x_i-\bar{x})^2 VAR(u_i|x_i) \\ &= \frac{\sum(x_i-\bar{x})^2\sigma_i^2}{SST_x^2} \end{aligned}$$

Heteroskedasticity

- What to do?
- In 1980, one of the most cited economics papers was written by Halbert White
- In it, he proposed calculating heteroskedastic robust standard errors as $\widehat{VAR}(\hat{\beta}_1) = \frac{\sum(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$
    - $\hat{u}_i^2$ is just the squared residual associated with each observation i
- These standard errors have LOTS of different names:
    - "White standard errors"
    - "Huber-White standard errors"
    - "Robust standard errors"
    - "Heteroskedasticity-robust standard errors"

A Preview of What's to Come

- The easiest thing to do is just calculate robust standard errors with $V\widehat{AR(\hat{\beta}}_1) = \frac{\sum(x_i - \bar{x})^2 \hat{u}_i^2}{SST_x^2}$
- However, if our fifth assumption fails (remember: $VAR(u|x) = \sigma^2$), it means that the OLS estimator is no longer the best'' estimator
- What do we mean by best''? The lowest-variance!
  - As an aside, **B**est **L**inear **U**nbiased **E**stimator or BLUE is a commonly used acronym for the OLS estimators. We will come back to this next week in more detail, but for now, note that we can prove **U**nbiasedness with the first four assumptions, and assumption 5 gives us **B**est