Topic 6 Limitations of Linear Regression Part II

MG205 Econometrics: Theory and Applications

José Ignacio González Rojas
j.i.gonzalez-rojas@lse.ac.uk.
London School of Economics and Political Science

9 December, 2024

Last week

- OVB
- Sample selection
- Projection vs. regression
- Non-random sampling
- Unbiasedness.
- Reverse causality

Today

- Take attendance
- What is in there for me to do the formative?
- Recap of the or TG P II.
- AT Wil Class Exercises
- Holidays.

L4: Outlying observations.

- Leverage values
- Influential observations / Outliers.
- Correction methods: Leave-one out estimator.
- Outlier ≠ typo ≠ different units ≠ measurement error.

Alternative methods:

- Estimate log model.
- Drop outliers.
- Winsorise.
- Model median, 99th percentile, etc.

LG: Heteroskedasticity.

- Makes inference harder. $Var(\beta_j)$ $IE[\hat{\sigma}^2 - \sigma^2]$

identically independently distributed

non-heteroskedas ticity

non-random sampling

$Var(e_i | X) = \sigma_i^2$

LS: Measurement error

Dependent $y = y^* + e_0 \quad e_0 \sim N(0, \delta_{e_0}^2)$

Independent $x = x^* + e_1 \quad e_1 \sim N(0, \delta_{e_1}^2)$

$y = \beta_0 + \beta_1 x + \underbrace{(e - \beta_1 e_1)}_{v}$

$cov(x, e_1) = E[e_1^2] = var(e_1)$

"OVB": special case. Attenuation bias: drives $plim(\hat{\beta}_1) = \beta_1 \cdot \frac{1}{1 + \frac{var(e_1)}{var(x^*)}} < \beta_1$ magnitude towards 0.

Q9. $N = 3,000$

admin data — subjective judgement.

$\widehat{wages} = 3.15 - 0.71$ below

(0.66) (0.33)

Measurement error: independent variable

=> attenuation bias.

-0.71 is closer to 0 than it should w/o bias, it could be -1.75. think in terms of magnitude

$t = |-2.15| > 1.96$ even with bias.

Q10. $\log(\text{wages}) = \underset{(1.3)}{3.2} + \underset{(0.2)}{0.4}\ \text{educ}$ $N = 2,000$.

a. $H_0: \beta = 0$ $H_a: \beta > 0$ => one-tailed test

$t = 2 > 1.645$ w/o abs value.

b. wages dependent variable. would reduce overall variance.

c. education. independent variable. attenuation bias.

Still rejected in both cases.

Q11. Same answer as lecture slides.

Feel free to ask for clarification!

QD. Why measurement error in the independent variable is not "a severe problem"?

Attenuation bias: $\text{plim}(\hat{\beta}_1) = \beta_1 \frac{1}{1 + \frac{\text{var}(e_i)}{\text{var}(x^*)}} < \beta_1$

Magnitude is always lower. Estimate goes to 0.

Lower bound.

Does not affect the sign. Persuasive of the effect.
Bad when you need the point estimate.

Q B. profits = $\beta_0 + \beta_1 \log(sales) + e$

β_1: ↑ 1% of sales will increase profits by $\frac{\hat{\beta}_1}{100}$ millions of dollars.

Heteroskedasticity does not affect point estimate.
Only affect inference.

Q14. Same argument as slides.

Q and A

Summary

- Measurement error.
- Lower bound
- Attenuation bias
- Non-random sampling
- Heteroskedasticity
- Outlying observations.

See you on January!