Classwork 7 (analytical)

EC421

2022-10-23

1 Signing the Bias

Suppose the true data generating processes in the examples below are of the form:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 o_i + u_i$$

Where x_i and o_i are exogenous such that $E[u_i|x, o] = 0$.

However, o is a variable that cannot be observed and so instead, we run $y_i = \beta_0 + \beta_1 x_i + w_i$, where $w_i = \beta_2 o_i + u_i$.

Then recall from the workbook chapter on Consistency that:

$$plim(\hat{\beta}_1) = \beta_1 + \frac{\beta_2 \text{Cov}(x, o)}{\text{Var}(x)}$$

1.1 Causal effect of Migration on Earnings

Suppose we took survey data that included people's earnings and whether they have recently moved to a new city or not, and we estimated this model:

$$\text{Earnings}_i = \beta_0 + \beta_1 \text{Migration}_i + w_i$$

Consider the unobservable variable absorbed in w_i which is $o_i = ambition_i$.

a) How does Ambition correlate with Earnings? That is, would you expect β_2 to be positive, negative, or 0?

b) How does Ambition correlate with Migration? That is, would you expect $Cov(migration, ambition)$ to be positive, negative, or 0?

c) We have to omit *ambition* from the model because it's not observable. Considering your answers to parts a) and b), we should expect that $\hat{\beta}_1$ will be biased (circle one): up / down / not at all.

1.2 Causal effect of friends of the opposite sex on a high school student's GPA

Suppose we took survey data that included high school students' GPAs and the number of friends of the opposite sex they have, and we estimated this model:

$$\text{GPA}_i = \beta_0 + \beta_1 \text{Opposite Sex Friends}_i + w_i$$

Consider the unobservable variable absorbed in w_i which is $o_i = \text{strict parents}_i$.

a) How does Strict Parents correlate with GPA? That is, would you expect β_2 to be positive, negative, or 0?

b) How does Strict Parents correlate with Opposite Sex Friends? That is, would you expect $Cov(\text{Opposite Sex Friends}, \text{Strict Parents})$ to be positive, negative, or 0?

c) We have to omit Strict Parents from the model because it's not observable. Considering your answers to parts a) and b), we should expect that $\hat{\beta}_1$ will be biased (circle one): up / down / not at all.

2.1 Covariance

Let A and B be random variables. Use the definition of covariance: $\text{Cov}(A, B) = E[(A - E[A])(B - E[B])]$ to show that $\text{Cov}(A, B) = E[AB] - E[A]E[B]$.

2.2 Explain why, in the workbook, we've called it a "freebie" to assume $E[u] = 0$.

Hint: take a look at this data generating process and the fitted model. The true model is: $y = 5 + 3x + u$, but we estimate β_0 to be much higher than the true value which is 5. Why is this?

```
library(tidyverse)

tibble(
  x = sample(-5:5, size = 100, replace = T),
  y = 5 + 3 * x + rnorm(n = 100, mean = 5)
) %>%
  lm(y ~ x, data = .) %>%
  broom::tidy()
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic   p.value
##   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
## 1 (Intercept)     9.97    0.0955     104.  2.62e-102
## 2 x               2.98    0.0305      97.7 1.74e- 99
```

2.3 Show exogeneity ($E[u|X] = 0$) implies that Cov$(X, u) = 0$.

You can use the facts that you proved in the previous two questions: $\text{Cov}(A, B) = E[AB] - E[A]E[B]$ and $E[u] = 0$.

The interpretation of this: If we have exogeneity, OLS estimators are unbiased AND consistent.

3 Consistency in the Presence of Measurement Error

Suppose the true data generating process for y is this: $y_i = \beta_0 + \beta_1 z_i + u_i$

We'd like to measure z, but we can't. Instead, we measure x, which is equal to $z + w$ (z with error).

w is zero on average ($E(w) = 0$) and has constant variance ($Var(w) = \sigma^2$). w is also independent of z and u, so $Cov(w, z) = Cov(w, u) = 0$. You can also assume that $Cov(z, u) = 0$.

When we try to estimate $y = \beta_0 + \beta_1 z + u$ using x, we're estimating $y = \beta_0 + \beta_1 x - \beta_1 w + u$, and the error term is $-\beta_1 w + u$.

So we have: $plim(\hat{\beta}_1) = \beta_1 + \frac{Cov(x, -\beta_1 w + u)}{Var(x)}$

3.1 (1 point Extra Credit) Show that $plim(\hat{\beta}_1) = \beta_1 + \frac{-\beta_1 Var(w)}{Var(x)}$**.**

3.2 (1 point Extra Credit) Finally, show that $plim(\hat{\beta}_1) = \beta_1(\frac{Var(z)}{Var(z)+Var(w)})$

3.3 Based on the formula in 3.2, does measurement error make estimates inconsistent?

That is, do we have $plim(\hat{\beta}_1) = \beta_1$ in the presence of measurement error?

3.4 Again based on the formula in 3.2, does measurement error bias estimates up or down?

Recall that variances are always nonnegative.