ECON-320-Lab-6

Sonan Memon

Introduction

- There are many possible causes of failure in accurate causal inference with econometric models; one of them is omitted variable bias.
- Correlation is not causation.

Packages

```r
library(tinytex)

library(tidyverse)

library(dslabs)
library(dplyr)

library(ggplot2)

library(tibble)
library(modelsummary)

library(broom)

library(haven)
```

Example with Simulation

```r
n <- 1000

set.seed(1)

# Generate data in a tibble
data_sim = tibble(
e1 = rnorm(n, sd = 3),
e2 = rnorm(n, sd = 2),
e3 = rnorm(n, sd = 1),
x = runif(n, min = 0, max = 10),
y = runif(n, min = 10, max = 20),
z = 20 - 0.3*y + 3*x + e1,
a = 6 + 2*x - 1.5*y + e2,
b = 10 - 0.5*y + 4*z + e3)

lm1 = lm(data = data_sim, a ~ x)
lm2 = lm(data = data_sim, a ~ y)
lm3 = lm(data = data_sim, a ~ x + y)
```

Example with Simulation

- Is there omitted variable bias in any of the following models?

$$a = \beta_0 + \beta_1 x$$

$$a = \beta_0 + \beta_1 y$$

$$a = \beta_0 + \beta_1 x + \beta_2 y$$

Regression Table: Models for a

Table 1: Results From Simulated Data

	Model 1	Model 2	Model 3
(Intercept)	-16.565***	17.031***	5.842***
	(0.308)	(1.026)	(0.370)
x	2.037***		1.994***
	(0.052)		(0.023)
y		-1.548***	-1.490***
		(0.068)	(0.023)

$+$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Example with Simulation

- Is there omitted variable bias in any of the following models?

$$b = \beta_0 + \beta_1 x$$

$$b = \beta_0 + \beta_1 y$$

$$b = \beta_0 + \beta_1 x + \beta_2 y$$

Example with Simulation

```
lm4 = lm(data = data_sim, b ~ y)
lm5 = lm(data = data_sim, b ~ z)
lm6 = lm(data = data_sim, b ~ y + z)
```

Regression Table: Models for b

Table 2: Results From Simulated Data

	Model 1	Model 2	Model 3
(Intercept)	152.698***	2.093***	9.971***
	(6.230)	(0.193)	(0.213)
y	-1.751***		-0.498***
	(0.411)		(0.011)
z		4.015***	4.001***
		(0.006)	(0.004)

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Bad Controls

- Irrelevant and bad controls can soak up all the variance and block causal pathways of interest.
- Example below adds an irrelevant control which is one example of a bad control.
- Look at adjusted R^2 values.

```
data_sim <- data_sim %>%
  mutate(c = rnorm(n, mean = 10, sd = 5))

lm9 = lm(data = data_sim, a ~ c + y)
```

F Tests

- Consider the model: $b = \beta_0 + \beta_1 y + \beta_2 z$
- If we want to test null hypotheses such as $H_0 : \beta_1 = \beta_2$ or $H_0 : \beta_1 = 2 * \beta_2$ or $H_0 : \beta_1 > \beta_2$ etc, we can use F tests.
- Restricted versus unrestricted models, Under $H_0 : \beta_1 = \beta_2$, model is: $b = \beta_0 + \beta_1(y + z)$.

F Tests

```
data_sim <- data_sim %>%
mutate(plus = y + z)

restricted = lm(data = data_sim, b ~ plus)
unrestricted = lm(data = data_sim, b ~ y + z)
```

- $F_{q,n-k-1} = \frac{\frac{RSS_r - RSS_u}{q}}{\frac{RSS_u}{n-k-1}}$
- For $H_o : \beta_1 = \beta_2,\ q = 1,\ k = 2,\ n = 1000$

Manual Computation of F-Test

```r
res_r_sq = (unname(resid(restricted)))^2
res_u_sq = (unname(resid(unrestricted)))^2

rss_r = sum(res_r_sq)
rss_u = sum(res_u_sq)

q = 1
k = 2

F_stat <- ((rss_r - rss_u)/q)/(rss_u/(n - k - 1))
F_stat
```

```
[1] 150181.6
```

F-statistic and Critical Value

- Once you have a F-statistic, you need to compare it against a critical value, which comes from a F-Table.
- $F_{1,1000-3} = Critical\,Value \approx 3.841$ and $F_{stat} = 150181.6 > 3.841$; we reject at $\alpha = 0.05$ and beyond.

Repeated Simulations