ECON 1123 Section 3

Slides at github.com/cjleggett/1123-section

Outline

- Name Circle
- Pset Feedback
- Lecture Recap / Questions
- Examples + Practice

Name Circle

Name Circle

- Name
- What’s a weird food combo you like?

Problem Set Feedback

Be careful with Wording

- Don’t say “causes” or “affects” unless you’re talking about a causal relationship!
- Make sure to answer all parts of each question.
- Interpreting coefficients:
    - A **[value] [units]** increase in **[variable]** is associated with a **[value] [units]** increase/decrease in **[variable]** on average, holding fixed **[variables]**
    - “A **10 percentage point** increase in the **fraction of people who are obese** is associated with a **.326 year** decrease in **life expectancy** on average, holding fixed the **fraction of people who currently smoke and exercise**”

Steps in OVB Problem

1. Choose an omitted variable

2. Sign β_2: correlation between omitted variable and outcome. Provide economic or real-world intuition for this sign.

3. Sign γ_1: correlation between omitted variable and variable of interest. Provide economic or real-world intuition for this sign.

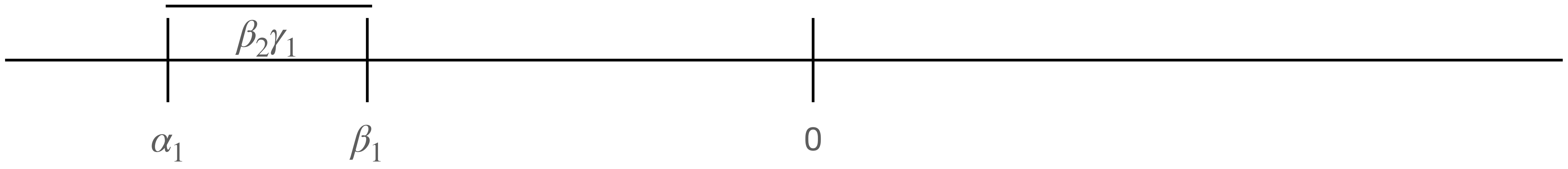
4. Based on the above two, say the bias is positive or negative.

5. Conclude whether we have overstated or understated the causal effect. (Is α_1 too big or too small?)

6. (Optional) Draw a number line.

Steps in OVB Problem

1. Omitted Variable is stress
2. I think people with more stress live shorter lives because stress can make it easier to get sick. ($\beta_2 < 0$)
3. I think people with more stress smoke more because some people use cigarettes as a coping mechanism ($\gamma_1 > 0$)
4. This means the bias ($\beta_2\gamma_1$) is negative.
5. This means we have overstated the causal effect of smoking on life expectancy, as we believe β_1 will be less negative than α_1



Lecture Recap

Nonlinear Regression Functions

- Often our data will not look like a nice straight line
- In this case we’ll want to be able to use functions of variables (mostly logs and polynomials)

Logarithms

Linear-Linear Regression	No logarithms
Log-Linear Regression	Taking log of dependent variable (left)
Linear-Log Regression	Taking log of independent variable (right)
Log-Log Regression	Taking log of both variables

Log Interpretations

- For small values of x, $log(1+x) = x$
  - (why? not super important for this class)
- We can use this for interpretations!

x

Linear-Linear Regression	A 1 unit change in x is associated with an alpha_1 unit change in y
Log-Linear Regression	A 1 unit change in x is associated with a 100alpha_1% change in y
Linear-Log Regression	A 1% change in x is associated with a 0.01alpha_1 unit change in y
Log-Log Regression	A 1% change in x is associated with an alpha_1% change in y

log(0)

- Log of zero is negative infinity
- This makes math very tricky!
- Two main strategies for dealing with this:
  - Drop 0 values
  - Replace 0s with some standard value

Polynomials

- We're still using linear regression, but including extra variables that are polynomial terms in the right side
- Unlike logs, we'll rarely ever use this on the left side of the equation
- Why is this helpful?
  - Effect of X on Y could depend on the value of X (think income vs. happiness)
  - We can flexibly control for other variables

Interpreting Polynomials

- We no longer get the simple interpretations we got before
- But we can still predict effects at a certain point!
- We can also calculate the standard error of this prediction, but it takes a bit more work
- We'll practice this later today

Joint Hypotheses

- We often want to test joint hypotheses (H0: x = y = z = 0)
- To do this, we use an F test (Wald test)
- Why might we want to do this?
  - We want to know if x has a significant relationship with y, but we have both the x and x^2 term
  - We have several indicator variables for one overall factor (eg. region)

Joint Hypotheses

- $H_0 : \beta_1 = 0$ and $\beta_2 = 0$
- $H_1 : \beta_1 \neq 0$ or $\beta_2 \neq 0$
- Why not just use the t-statistic for each?
- What's the probability of rejecting
- Size vs. Power

Decision	H_0 is True	H_1 is True
Reject H_0	*Type I error*	Correct decision
Accept H_0	Correct decision	*Type II error*

F Test (Wald Test)

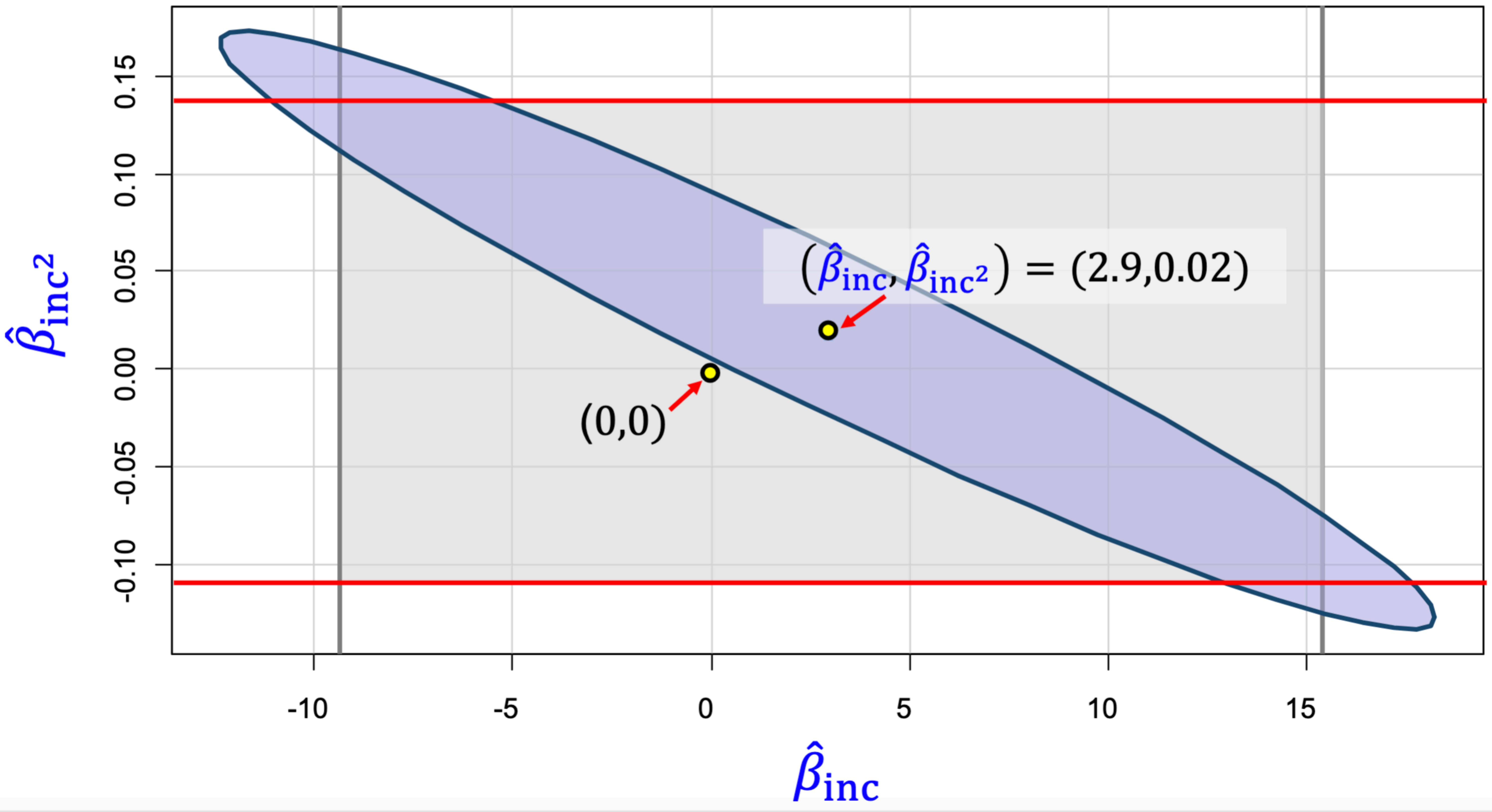
- $H_0 : \beta_1 = 0$ and $\beta_2 = 0$
- $H_1 : \beta_1 \neq 0$ or $\beta_2 \neq 0$
- q degrees of freedom when constricting q variables in null hypothesis
- F statistic has $F_{q,\infty} = \chi^2_q / q$ distribution.
    - Why? Interesting linear algebra stuff beyond scope of this clas

F Test (Wald Test) in practice

1. Run linear regression to get model
2. Run a Wald test (built into stata, R, and Python) to get F statistic
3. Calculate p-value by subtracting CDF of $\chi^2(df)$ at $df \times F$ from 1
4. Calculate critical value by using the inverse CDF of the F-distribution at df degrees of freedom for .95

Let’s draw what this really means because it’s confusing!

Visualization



Exercises!