Econometrics

Lecture 7

Nathaniel Higgins

ERS and JHU

This week

- Fatima's TA session
- My office hours are Wed from 8a to 9a
- I can be available Wed evening starting as early as 6p
  - Where to meet? Office? Coffee shop? Lab?
- Extra office hours (I will be very flexible on Friday)

Study help

- Don't forget: old homeworks are available — see what you got wrong the first time

Cheat sheet

- Everybody can have a cheat sheet
- Because nobody practices econometrics without a computer, the internet, and a few good books
- Cheat sheet can be one side of a sheet of paper (no giant pieces of paper — that's just cheesy)
- Don't write the pdf or cdf of a distribution on your sheet
- If the actual formula for a distribution is needed, I'll provide it to you
- You might not actually need the sheet. Don't get mad at me if you don't use it (that is once again cheesy)

What to study

- Homeworks
  - Pay special attention to problems which require reasoning
  - Questions of basic interpretation aren't very rewarding, from a cost-benefit standpoint
  - Examples: compound comparative statics (bedroom and sqrft example, C3.2; using coefficients to derive useful measures of change, C3.8(iv))
- Lectures
  - Don't worry about deriving any formulas
  - Interpretation is important
  - Final formulas (of a derivation) are worth reviewing
- Book
  - Definitions
  - Final formulas (not derivations, but formulas of the OLS coefficients, e.g.)

How to study
My recommendation

- Go through the book and write down on your CS definitions that you have heard before, but that you have trouble remembering
- Go through homeworks (without the answers) and see how quickly you can recall the general thrust of the answer
  - Study the homework problems that are difficult for you
  - Emphasize homework problems that involve compound reasoning
- Go through lecture notes
  - Write down formulas that represent a concept (e.g. Variance of an estimator, conditional expectation function, R-squared in terms of the variation in the data) — make sure the interpretation makes sense
  - Play with comparative statics (huh?)
  - Do computer exercises (not getting hung up on Stata coding) — the point of these exercises is to build intuition

How the test will look

- It will probably take you less than 2 hours
- It will contain: (1) some easy questions, (2) some thinkers, and (3) some questions you are totally unprepared for (use your intuition and what you understand so far to answer as best you can)
- The second two types of questions generate separation in the grades
- Examples of each kind:
  1. Definitions, such as R-squared, std. dev.; interpretations of linear, log-linear, linear-log, and log-log coefficients; T/F or mult. choice
  2. In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
  3. What is the R-squared of a regression of y on nothing but a constant (i.e. no independent x-variables)?

How the test will look
Example type 1

- R-squared is the proportion of of variation in the dependent variable that is explained by the independent variables, i.e. it is the explained sum of squares over the total sum of squares

How the test will look
Example type 1

- R-squared is the proportion of of variation in the dependent variable that is explained by the independent variables, i.e. it is the explained sum of squares over the total sum of squares
  - The first clause is necessary; the second clause helps because it makes clear you know what you are talking about if you bungle the first clause a little bit
  - If you write down the formula but no interpretation, you get 10 dork points (note: dork points do not translate to real points)

How the test will look

Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.

How the test will look

Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward

How the test will look

Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?

How the test will look

Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$?

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases?

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated).

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated). If when x_1 decreases y also decreases, then x_1 and y are positively related (they move in the same direction). What does this mean for the estimate of β_1?

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated). If when x_1 decreases y also decreases, then x_1 and y are positively related (they move in the same direction). What does this mean for the estimate of β_1?
- Case 1: β_1 is positive:

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated). If when x_1 decreases y also decreases, then x_1 and y are positively related (they move in the same direction). What does this mean for the estimate of β_1?
- Case 1: β_1 is positive: The omission of x_2 makes $\widehat{\beta_1}$ *more* positive (upward bias)

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated). If when x_1 decreases y also decreases, then x_1 and y are positively related (they move in the same direction). What does this mean for the estimate of β_1?
- Case 1: β_1 is positive: The omission of x_2 makes $\widehat{\beta_1}$ *more* positive (upward bias)
- Case 2: β_1 is negative:

How the test will look
Example type 2

- In a model with two independent variables (e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$) what is the direction of bias in the coefficient β_1 when x_2 is omitted from the equation if: β_2 is negative and x_1 and x_2 are negatively correlated.
- The direction of the bias is upward
- How do we know?
- When x_2 increases, what happens to $\widehat{y}$? It decreases (bc β_2 is negative). What else happens when x_2 increases? x_1 decreases (bc x_1 and x_2 are negatively correlated). If when x_1 decreases y also decreases, then x_1 and y are positively related (they move in the same direction). What does this mean for the estimate of β_1?
- Case 1: β_1 is positive: The omission of x_2 makes $\widehat{\beta_1}$ *more* positive (upward bias)
- Case 2: β_1 is negative: The omission of x_2 makes $\widehat{\beta_1}$ *less* negative (upward bias)

- In both cases, it is correct to say that the bias is upward. Reasoning it out, however, makes more clear what you mean.
- You mean that the bias is in the *positive direction*
- This is different from saying that it makes the coefficient larger in absolute value
- In fact, whether the bias makes x_1 appear to have a *larger effect* (in the sense of the significance of its value in predicting y) depends on whether β_1 is positive or negative

- This is why including reasoning is important. If you included the reasoning above, you would earn full points (because it is clear you understand what is going on). If you gave the answer without the reasoning, you might be right (if you were referring to an increase in value) or wrong (if you were referring to an increase in absolute value, i.e. an increase in power of prediction).

How the test will look

Example type 3

- What is the R-squared of a regression of y on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)

How the test will look

Example type 3

- What is the R-squared of a regression of *y* on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)
- Put down what you can — starting with a definition is never a bad idea
- What does R-squared *represent*? Break it down to its component parts

How the test will look
Example type 3

- What is the R-squared of a regression of *y* on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)
- Put down what you can — starting with a definition is never a bad idea
- What does R-squared *represent*? Break it down to its component parts
- R-squared represents the explained variation in a dependent variable *y*
- What is variation?

How the test will look
Example type 3

- What is the R-squared of a regression of *y* on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)
- Put down what you can — starting with a definition is never a bad idea
- What does R-squared *represent*? Break it down to its component parts
- R-squared represents the explained variation in a dependent variable *y*
- What is variation? It is movement around the mean value

How the test will look

Example type 3

- What is the R-squared of a regression of *y* on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)
- Put down what you can — starting with a definition is never a bad idea
- What does R-squared *represent*? Break it down to its component parts
- R-squared represents the explained variation in a dependent variable *y*
- What is variation? It is movement around the mean value
- What does a regression without an x-variable look like?

How the test will look
Example type 3

- What is the R-squared of a regression of *y* on nothing but a constant (i.e. no independent x-variables)?
- Talk to your neighbors
- If you don't know how to start, start small (define R-squared)
- Put down what you can — starting with a definition is never a bad idea
- What does R-squared *represent*? Break it down to its component parts
- R-squared represents the explained variation in a dependent variable *y*
- What is variation? It is movement around the mean value
- What does a regression without an x-variable look like? It is a regression of *y* on a constant
- What is the interpretation of the constant?

How the test will look
Example type 3

- The constant is the mean of y

How the test will look
Example type 3

- The constant is the mean of y
- If your predicted value of y is always equal to the mean, how much variation *around the mean* can the model explain?

How the test will look

Example type 3

- The constant is the mean of y
- If your predicted value of y is always equal to the mean, how much variation *around the mean* can the model explain?
- zilch

How the test will look

Example type 3

- The constant is the mean of y
- If your predicted value of y is always equal to the mean, how much variation *around the mean* can the model explain?
- zilch
- The answer is 0.

How the test will look
Example type 3

- The constant is the mean of y
- If your predicted value of y is always equal to the mean, how much variation *around the mean* can the model explain?
- zilch
- The answer is 0.
- A good answer to this question would include the answer (i.e. the number "0") plus the reason why (because a model that predicts the same value of y all the time cannot possibly explain any variation, and R-squared is a measure of explained variation).

How the test will look

- Multiple choice? possible
- T/F? possible
- Definitions? possible
- At least one of the previous three? definitely
  - These questions are graded without any partial credit: you were either right or wrong
- Short answer? definitely
  - You will earn credit for reasoning and interim steps (i.e. you will earn points for parts of your work other than the eventual statement of the answer)
- Essay? no way

How the test will look

- Context vs. no-context
  - You can expect several problems where a model is described as having generic components, such as y, x_1, x_2, and u, e.g.
  - You can also expect some problems where instead of giving variables generic names, variables will take on names of actual data, e.g. `bwght`, `cornYield`, etc.
  - When a variable takes the identity of actual data, the intuition surrounding the real-world situation is in play

How the test will look

- Context vs. no-context
  - You can expect several problems where a model is described as having generic components, such as y, x_1, x_2, and u, e.g.
  - You can also expect some problems where instead of giving variables generic names, variables will take on names of actual data, e.g. `bwght`, `cornYield`, etc.
  - When a variable takes the identity of actual data, the intuition surrounding the real-world situation is in play
  - You can expect some questions which are open to interpretation — argue your point!

Tricky business

- I may include irrelevant information in the questions
- This has the effect of testing that you understand the material, rather than having simply memorized a procedure

Definitions

- Type I error
- Type II error
- Randomness
- Distribution function
- Omitted variable bias
- Multicollinearity
- Explained sum of squares
- Homoskedasticity / heteroskedasticity
- BLUE (I won't ask you what it stands for, but I want you to know what "Best Linear Unbiased Estimator" means, i.e. what does it mean for an estimator to be unbiased, what does it mean to be best)

Definitions
Out

- Underspecifying the model
- Micronumerosity
- Variance Inflation Factor