Econometrics
Lecture 6

Nathaniel Higgins

JHU

Outline

- The properties of the OLS model
- Variance of the OLS estimator (setting-up our work on statistical inference using the OLS model)
- Talk about t-tests
- Testing compound hypotheses
- Then we will look to relax some of the assumptions that we have relied on to get where we are (MLR.1 – MLR.6)
- If time: begin discussion of functional form: esp. dummy variables and interaction terms

- OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
- Endogeneity is present any time an x-variable is related to the unobservable (the u term) in a regression equation

- OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
- Endogeneity is present any time an x-variable is related to the unobservable (the u term) in a regression equation
- We don't like endogeneity because it biases our coefficient estimates (think of what *bias* means in this context)

Review of last time
Endogeneity

- OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
- Endogeneity is present any time an x-variable is related to the unobservable (the u term) in a regression equation
- We don't like endogeneity because it biases our coefficient estimates (think of what *bias* means in this context)
- Omitted variables generate bias because the omitted variable is subsumed into u

- OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
- Endogeneity is present any time an x-variable is related to the unobservable (the u term) in a regression equation
- We don't like endogeneity because it biases our coefficient estimates (think of what *bias* means in this context)
- Omitted variables generate bias because the omitted variable is subsumed into u
- This is bad. When u is related to the x-variable(s), we accidentally attribute to x some movements that are really movements in u (and these are ultimately really movements in the omitted variables)

- OVB is just one type of endogeneity. It happens to be a type that we can understand pretty easily.
- Endogeneity is present any time an x-variable is related to the unobservable (the *u* term) in a regression equation
- We don't like endogeneity because it biases our coefficient estimates (think of what *bias* means in this context)
- Omitted variables generate bias because the omitted variable is subsumed into *u*
- This is bad. When *u* is related to the x-variable(s), we accidentally attribute to x some movements that are really movements in *u* (and these are ultimately really movements in the omitted variables)
- We don't know yet what to do about endogeneity, but we know to look out for it

Endogeneity: OVB

- Here's a mathematical-model-based version of the material on the previous slide

Endogeneity: OVB

- Here's a mathematical-model-based version of the material on the previous slide
- Suppose that the *true* DGP of y is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Endogeneity: OVB

- Here's a mathematical-model-based version of the material on the previous slide
- Suppose that the *true* DGP of y is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- But we lack data on x_2 (this happens *a lot*)

Endogeneity: OVB

- Here's a mathematical-model-based version of the material on the previous slide
- Suppose that the *true* DGP of y is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- But we lack data on x_2 (this happens *a lot*)
- As a consequence, we run the following regression instead of the one we really want to run (above)

$$y = \alpha_0 + \alpha_1 x_1 + v$$

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (1)$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \quad (2)$$

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \tag{2}$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y?

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \tag{2}$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y? It goes up

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (1)$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \quad (2)$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y? It goes up
- when x_2 increases, what is likely to be true of x_1?

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (1)$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \quad (2)$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y? It goes up
- when x_2 increases, what is likely to be true of x_1? It increases, too

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \tag{2}$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y? It goes up
- when x_2 increases, what is likely to be true of x_1? It increases, too
- If we are running regression (2) instead of (1), what do we see?

Endogeneity: OVB

- True:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$

- What we do:

$$y = \alpha_0 + \alpha_1 x_1 + v \tag{2}$$

- Suppose two things: (i) that both β_1 and β_2 are positive and (ii) that x_1 and x_2 are positively correlated
- When x_2 increases, what happens to y? It goes up
- when x_2 increases, what is likely to be true of x_1? It increases, too
- If we are running regression (2) instead of (1), what do we see?
- We see: x_1 going up and y going up. This causes us to believe that x_1 is *responsible* for the increase in y

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$
- When x_2 increases, what happens to y?

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$
- When x_2 increases, what happens to y? It goes down

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$
- When x_2 increases, what happens to y? It goes down
- When x_2 increases, what is likely to be true of x_1?

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$
- When x_2 increases, what happens to y? It goes down
- When x_2 increases, what is likely to be true of x_1? It increases, too

Endogeneity: OVB

- We can go through this exact same exercise starting with different assumptions about β_1 and β_2 and the correlation of x_1 and x_2 (try it!)
- Let's always assume that β_1 is positive (easy to switch this one assumption later)
- We already know what happens when $corr(x_1, x_2) > 0$ and $\beta_2 > 0$
- Suppose instead that $corr(x_1, x_2) > 0$ and $\beta_2 < 0$
- When x_2 increases, what happens to y? It goes down
- When x_2 increases, what is likely to be true of x_1? It increases, too
- We see: x_1 going up and y going down. This causes us to believe that x_1 is *responsible* for the decrease in y Downward bias in our estimate of β_1

Endogeneity: OVB

- Figure out the other cases on your own when you review these notes!
- (Consult Table 3.2 of Wooldridge to check your answers; p. 90 of ed. 5)
- Always follow the same basic procedure
- Ask yourself: as x_2 increases, what happens to (i) y and (ii) x_1
- What is the net effect of what we see?
- What will this cause us to conclude (what is the direction of the bias)?

- OVB causes bias in our coefficient estimates
- This bias comes as a result of attributing to x_1 (the included variable) movements in y that are really due to movements in x_2 (the omitted variable)
- This is one manifestation of *endogeneity*, which is a problem that causes bias in our coefficient estimates whenever u is correlated with included independent variable(s)
- We haven't explored any other flavors of engoeneity yet, but we will
- In general, we want $cov(u, x) = 0$
- We don't know yet what to do about endogeneity, but we know to look out for it

Multivariate regression

- This concludes — for now — or discussion of the unique issues associated with multivariate regression analysis
- We have talked exclusively about *point estimates* — estimates of the value of β_k
- We now move on to discussing the *variability* of these point estimates
- The following material applies to both bivariate and multivariate regression models, but we tackle the material solely in the context of multivariate models

Multivariate regression

- This concludes — for now — or discussion of the unique issues associated with multivariate regression analysis
- We have talked exclusively about *point estimates* — estimates of the value of β_k
- We now move on to discussing the *variability* of these point estimates
- The following material applies to both bivariate and multivariate regression models, but we tackle the material solely in the context of multivariate models
- To set the stage for this discussion, we need to examine some assumptions we usually make when we estimate a basic multivariate regression model

Where we're going

- We haven't talked much about distributional assumptions
- Most of we've done so far hasn't depended on distribution of unobservables (exception is the ML estimator)
- Today, we'll investigate the assumptions of the classical linear model: the Gauss-Markov assumptions
- We've talked casually about the first four — now we're going to be explicit

Where we're going

- We haven't talked much about distributional assumptions
- Most of we've done so far hasn't depended on distribution of unobservables (exception is the ML estimator)
- Today, we'll investigate the assumptions of the classical linear model: the Gauss-Markov assumptions
- We've talked casually about the first four — now we're going to be explicit
- The first four assumptions give us:
  - OLS is unbiased

Where we're going

- We haven't talked much about distributional assumptions
- Most of we've done so far hasn't depended on distribution of unobservables (exception is the ML estimator)
- Today, we'll investigate the assumptions of the classical linear model: the Gauss-Markov assumptions
- We've talked casually about the first four — now we're going to be explicit
- The first four assumptions give us:
  - OLS is unbiased
- We'll add the fifth and sixth assumptions, which give us:
  - An explicit formula for the variation of our estimator ($\widehat{\beta}$)
  - This formula, along with an assumption about the distribution of u, to give us test statistics (enables statistical inference)

1. MLR.1 (p. 83)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

2. MLR.2 (p. 84)
   Our data (x,y) represent a random sample from the relevant population
3. MLR.3 (p. 84)

$$x_i \neq z * x_j$$

4. MLR.4 (p. 86)

$$E(u|x) = 0$$

Gauss-Markov assumptions
5-6

5. MLR.5 (p. 93)

$$Var(u|x) = \sigma^2$$

6. MLR.6 (p. 118)
   - u is independent of x's and is distributed normally with mean of zero and variance of σ^2

Gauss-Markov assumptions

- These are the “base” assumptions
- Much of applied econometrics deals with figuring out how to proceed when these assumptions fail
- Of course, in reality they *always* fail a little

Gauss-Markov assumptions

- These are the “base” assumptions
- Much of applied econometrics deals with figuring out how to proceed when these assumptions fail
- Of course, in reality they *always* fail a little
- In practice, we worry about them a lot when there is a compelling reason to think that one of the assumptions fails badly
- Then we look for solutions

Gauss-Markov assumptions

- These are the “base” assumptions
- Much of applied econometrics deals with figuring out how to proceed when these assumptions fail
- Of course, in reality they *always* fail a little
- In practice, we worry about them a lot when there is a compelling reason to think that one of the assumptions fails badly
- Then we look for solutions
- But as a first cut at almost every problem, we do the analysis assuming that the Gauss-Markov assumptions hold
    - (Homoskedasticity might be a slight caveat to that example — more on that later)

Variance of the OLS estimators

- With MLR.1-MLR.5 we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- Wait, the beta-hats have variance?

Variance of the OLS estimators

- With MLR.1-MLR.5 we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- Wait, the beta-hats have variance?
- Before charging ahead, this is a good time to get philosophical, to remind ourselves what we're doing

Variance of the OLS estimators

- With MLR.1-MLR.5 we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- Wait, the beta-hats have variance?
- Before charging ahead, this is a good time to get philosophical, to remind ourselves what we're doing
- When we estimate a model based on a random sample of data, we *expect* to get an unbiased estimate of the true beta
- What we get is only an estimate, of course — we don't expect it to be exactly equal to the true value
- We expect that if we were to run the same model, but on an alternative sample of data, we would get a different estimate. Let's go through some code to illustrate this...

Variance of the OLS estimators

- Moral of the story: Our estimate of $\widehat{\beta}$ is itself a random variable, dependent on the randomness in the data that we collect.
- Although we expect our estimate to be unbiased (correct) *on average*, the $\widehat{\beta}$ has variance, so sometimes we will get estimates that are more right (or more wrong) than others
- All else equal, it would be nice to have estimators that produce $\widehat{\beta}$s that have low variance, so we can be confident that our estimates cluster around the true values of the unknown parameters
- Draw a distribution of $\widehat{\beta}$ for yourself to illustrate bias and variance

Variance of the OLS estimators

- With MLR.1-MLR.5 we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- The formula can be found on p. 94 (I don't care to have you derive it)
- That said, the formula has some intuitive content

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

Variance of the OLS estimators

- With MLR.1-MLR.5 we can derive a formula for the variance of the beta-hats ($\widehat{\beta}$)
- The formula can be found on p. 94 (I don't care to have you derive it)
- That said, the formula has some intuitive content

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- The variance of our estimator increases when the numerator increases and decreases when the denominator increases
- The numerator is a constant that represents the true variance of the unobservables in our equation. This makes sense. As the unexplainable portion of y varies more, our estimator is less capable of picking out variation in y that is caused by x (think of a moving target)

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- The denominator is made out of two related things
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - How much a whole column of x-data varies
     - More variation in a given x-variable makes estimation easier
     - If we want to estimate β_1, then we want x_1 to be varying like crazy — if we want to estimate the effect of education on earnings, then a sample made out of our class would be a good/bad sample?

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- The denominator is made out of two related things
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - How much a whole column of x-data varies
     - More variation in a given x-variable makes estimation easier
     - If we want to estimate β_1, then we want x_1 to be varying like crazy — if we want to estimate the effect of education on earnings, then a sample made out of our class would be a good/bad sample? Bad. Everyone has a pretty similar educational background. We need to vary x to see how y reacts. (With the same instruments, it's easier to uncover the truth when the data jump around a lot)

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- The denominator is made out of two related things
  1. $SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$
     - How much a whole column of x-data varies
     - More variation in a given x-variable makes estimation easier
     - If we want to estimate β_1, then we want x_1 to be varying like crazy — if we want to estimate the effect of education on earnings, then a sample made out of our class would be a good/bad sample? Bad. Everyone has a pretty similar educational background. We need to vary x to see how y reacts. (With the same instruments, it's easier to uncover the truth when the data jump around a lot)
  2. $(1 - R_j^2)$
     - see next slide

Variance of the OLS estimators

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

- Denominator is made of
  1. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
     - see previous slide
  2. $(1 - R_j^2)$
     - $R_j^2 =$ The proportion of variation in x_j explained by all the other x-variables
     - Saying that the x-variables are highly correlated is like saying that the x-variables don't do a lot of varying independently (think using the Ballantine)
     - This is bad, because we don't have much variation in the x-variables to use to estimate each independent relationship

Variance of the OLS estimators

Takeaways 1

- All else equal, the variance of $\widehat{\beta}_i$ is smaller when:
  1. the variance of the unobservables (u) is smaller (i.e. σ is smaller)
  2. the variance of x_i is larger
  3. the *independent* variability of x_i is larger

- For more precise estimates of the β's we care about, we can do a few things
  1. Make the sample size as big as possible (increase n whenever possible). This is listed as #1 for a reason! There is no downside to getting more data.
  2. Collect x-variables that explain as much of y as possible (practically, this reduces the role of u)
  3. Be aware of the co-variation of x-variables (multicollinearity)

$$Var(\widehat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

1. $\sigma^2 = Var(u|x)$
   - Estimate this value using residuals from our regression model: $\widehat{\sigma^2} = \frac{(\widehat{u_1^2} + \widehat{u_2^2} + \ldots + \widehat{u_n^2})}{(n-k-1)}$
2. $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$
3. $R_j^2 = R^2$ from a regression of x_j on all the other x-variables (x_{-j})

- We estimate the variance of the OLS estimator

$$Var(\widehat{\beta}_j) = \frac{\widehat{\sigma^2}}{SST_j(1 - R_j^2)}$$

- ...then we take the square root of it

$$se(\widehat{\beta}_j) = \sqrt{Var(\widehat{\beta}_j)}$$

- ...and THIS is the value that we see in our regression output

- OK, so we know how to come up with a formula that estimates the variance of the OLS estimator

$$Var(\widehat{\beta}_j) = \frac{\widehat{\sigma^2}}{SST_j(1 - R_j^2)}$$

- ...and we know how to use this to calculate a thing called the *standard error* of the j^{th} coefficient estimate
- Now what?

- So what? Who cares?

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - “Oh *that value*? We see that value all the time.

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - "Oh *that value*? We see that value all the time. My guess was that β_1 was zero, but I'm not at all surprised by $\widehat{\beta_1} = 1$."

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - "Oh *that value*? We see that value all the time. My guess was that β_1 was zero, but I'm not at all surprised by $\widehat{\beta_1} = 1$."
- Knowing how likely a value is to have occurred by chance means that we can say intelligent things about how likely it is that our coefficient estimates were obtained because of something *other than chance*
  - "Not a likely value just by chance, eh? Must be something deterministic going on here."

- So what? Who cares?
- Knowing how variable a statistic is means that we can determine how likely a value is to have occurred by chance
  - "Oh *that value*? We see that value all the time. My guess was that β_1 was zero, but I'm not at all surprised by $\widehat{\beta_1} = 1$."
- Knowing how likely a value is to have occurred by chance means that we can say intelligent things about how likely it is that our coefficient estimates were obtained because of something *other than chance*
  - "Not a likely value just by chance, eh? Must be something deterministic going on here."
- Take our example dset from earlier (dset1)

- When we ran a regression of y on x_1 and x_2, we obtained an average point estimate of $\beta_1 = 2.01$.
- Is this estimate "big?"

- When we ran a regression of y on x_1 and x_2, we obtained an average point estimate of $\beta_1 = 2.01$.
- Is this estimate “big?” I don’t know either. “Big” is relative. Only context of the problem makes an estimate “big.” (think of a returns-to-education example if you want to think about “big” in context)

- When we ran a regression of y on x_1 and x_2, we obtained an average point estimate of $\beta_1 = 2.01$.
- Is this estimate "big?" I don't know either. "Big" is relative. Only context of the problem makes an estimate "big." (think of a returns-to-education example if you want to think about "big" in context)
- Inference does NOT help us to answer questions like the above

- When we ran a regression of y on x_1 and x_2, we obtained an average point estimate of $\beta_1 = 2.01$.
- Is this estimate "big?" I don't know either. "Big" is relative. Only context of the problem makes an estimate "big." (think of a returns-to-education example if you want to think about "big" in context)
- Inference does NOT help us to answer questions like the above
- Is the coefficient estimate of 2.01 an anomaly? Maybe the fact that we obtained 2.01 rather than 1 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.01). How likely is it that the true coefficient is actually 0? or1?

- When we ran a regression of y on x_1 and x_2, we obtained an average point estimate of $\beta_1 = 2.01$.
- Is this estimate "big?" I don't know either. "Big" is relative. Only context of the problem makes an estimate "big." (think of a returns-to-education example if you want to think about "big" in context)
- Inference does NOT help us to answer questions like the above
- Is the coefficient estimate of 2.01 an anomaly? Maybe the fact that we obtained 2.01 rather than 1 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.01). How likely is it that the true coefficient is actually 0? or1?
- This is the type of question we can start to answer now

- Let me ask you a question: if the point estimate for β_1 was 2.009991 and the standard error (a measure of variation) was 0.00000000000001, how likely do you think it is that the true value of β_1 is 0? 1?

- Let me ask you a question: if the point estimate for β_1 was 2.009991 and the standard error (a measure of variation) was 0.0000000000001, how likely do you think it is that the true value of β_1 is 0? 1?
- Not very
- That is a loose notion, though. The whole idea of test statistics is to put a finer set of bounds on when the standard error is “too big” for us to consider our estimate of β_j “significantly different” from 0 or 1 (or some other value we choose)

Inference (the distribution step)
MLR.6

- We have always assumed that the u are random
- But we haven't assumed much else
- In order to determine the s.e. of the coefficient estimates, we assumed homoskedasticity (fixed σ^2)
- In order to take the next step, we need to make an assumption about exactly how u is distributed

Inference (the distribution step)
MLR.6

- We have always assumed that the u are random
- But we haven't assumed much else
- In order to determine the s.e. of the coefficient estimates, we assumed homoskedasticity (fixed σ^2)
- In order to take the next step, we need to make an assumption about exactly how u is distributed
- Next step: determine how $\widehat{\beta}_j$ is distributed

- We have always assumed that the u are random
- But we haven't assumed much else
- In order to determine the s.e. of the coefficient estimates, we assumed homoskedasticity (fixed σ^2)
- In order to take the next step, we need to make an assumption about exactly how u is distributed
- Next step: determine how $\widehat{\beta}_j$ is distributed
- In more advanced work, you might look at your estimates of u (the $\widehat{u}$ that you obtain when you estimate a model) and use them to make a guess of how u is really distributed
- In most cases you will assume that u is distributed in a certain way — the most common assumption is that u is distributed normally

- We add to our set of working assumptions the *normality assumption*
- Our accumulated assumptions up to this point were MLR.1 – MLR.5
- Add to this a normality assumption and we get MLR.6

$$u \sim N(0, \sigma^2)$$

- We add to our set of working assumptions the *normality assumption*
- Our accumulated assumptions up to this point were MLR.1 – MLR.5
- Add to this a normality assumption and we get MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period.

- We add to our set of working assumptions the *normality assumption*
- Our accumulated assumptions up to this point were MLR.1 – MLR.5
- Add to this a normality assumption and we get MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period. But we need MLR.6 to use our basic tools of statistical inference

- We add to our set of working assumptions the *normality assumption*
- Our accumulated assumptions up to this point were MLR.1 – MLR.5
- Add to this a normality assumption and we get MLR.6

$$u \sim N(0, \sigma^2)$$

- We can operate an OLS model without MLR.6. Period. But we need MLR.6 to use our basic tools of statistical inference
- Our total set of assumptions so far (MLR.1 – MLR.6) are the assumptions of the Classic Linear Regression Model (or CLM, in Wooldridge's parlance)

- The yield of MLR.1 - MLR.6 ...

- The yield of MLR.1 - MLR.6 ...
- Under CLM assumptions

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

  - The fact that the mean of $\widehat{\beta}_j$ is β_j is “unbiasedness”
  - The formula for $Var(\widehat{\beta}_j)$ was given earlier in the lecture; way too long to print here

- The yield of MLR.1 - MLR.6 . . .
- Under CLM assumptions

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

  - The fact that the mean of $\widehat{\beta}_j$ is β_j is “unbiasedness”
  - The formula for $Var(\widehat{\beta}_j)$ was given earlier in the lecture; way too long to print here
- Knowing all this allows us to conclude that

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

Inference

- Pause

Inference

- Pause
- We know

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

Inference

- Pause
- We know

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

- If we know the distribution of $\widehat{\beta}_j$, we can do what we said we wanted to do

Inference

- Pause
- We know

$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

- If we know the distribution of $\widehat{\beta}_j$, we can do what we said we wanted to do
- Recall: "Is the coefficient estimate of 2.01 an anomaly? Maybe the fact that we obtained 2.01 rather than 0 or 1 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.01). How likely is it that the true coefficient is actually 1?"

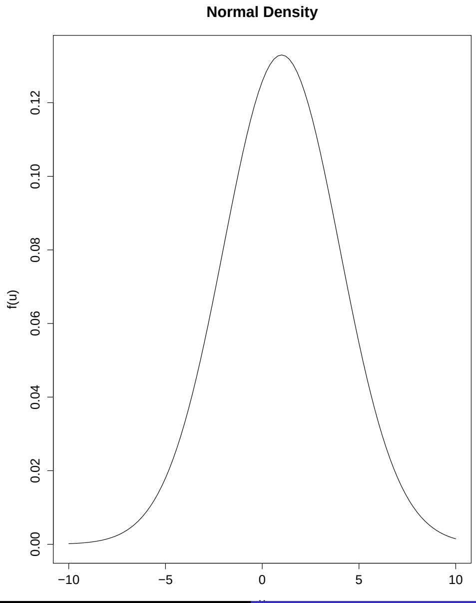
Inference

- Pause
- We know

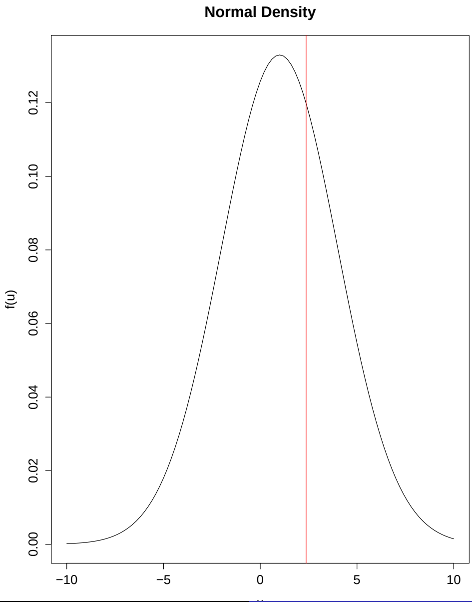
$$\widehat{\beta}_j \sim N(\beta_j, Var(\widehat{\beta}_j))$$

- If we know the distribution of $\widehat{\beta}_j$, we can do what we said we wanted to do
- Recall: "Is the coefficient estimate of 2.01 an anomaly? Maybe the fact that we obtained 2.01 rather than 0 or 1 is just due to random fluctuations (rather than the fact that the true value of the coefficient is around 2.01). How likely is it that the true coefficient is actually 1?"
- We could look at a normal distribution with mean 1 and variance $V(\widehat{\beta}_j)$, and we could determine whether the value of $\widehat{\beta}_j$ that we observed was in the "meaty" part of the distribution, or whether it was "in the tails"

Inference



Inference

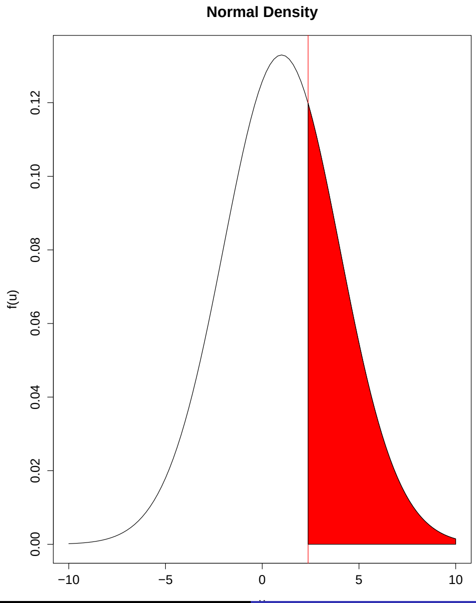


Inference

- Specifically, we could look at a normal distribution with a mean of 1, and a variance of 0.35 (approximate $V(\widehat{\beta_1}) = 0.35$) and we could say "how likely are we to have pulled a value at least as large as 2.01 from this distribution?"

- Specifically, we could look at a normal distribution with a mean of 1, and a variance of 0.35 (approximate $V(\widehat{\beta_1}) = 0.35$) and we could say "how likely are we to have pulled a value at least as large as 2.01 from this distribution?"
- That is, we could evaluate $\Pr(2.01 \leq \widehat{\beta}_j)$

Inference



Inference

- So if we can do all that just by knowing

$$\widehat{\beta}_j \sim N(\beta_j, V(\widehat{\beta}_j))$$

. . . then why do we need:

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

Inference

- So if we can do all that just by knowing

$$\widehat{\beta}_j \sim N(\beta_j, V(\widehat{\beta}_j))$$

...then why do we need:

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- Turns out, it's sort of a pain to look up the probability that a value greater than *something* (like 2.01, e.g.) occurs in a normal distribution with mean 1 and variance 0.35 — every time we want to evaluate a new question, we'd have to look at a new distribution (with new mean and new variance)

Inference

- So if we can do all that just by knowing

$$\widehat{\beta}_j \sim N(\beta_j, V(\widehat{\beta}_j))$$

  ... then why do we need:

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- Turns out, it's sort of a pain to look up the probability that a value greater than *something* (like 2.01, e.g.) occurs in a normal distribution with mean 1 and variance 0.35 — every time we want to evaluate a new question, we'd have to look at a new distribution (with new mean and new variance)
- So instead, we *standardize* the random variable

Inference

- Standardizing the random variable means that we only have to look at one distribution, no matter the value of $\widehat{\beta}_j$ we are evaluating, no matter its variance

Inference

- Standardizing the random variable means that we only have to look at one distribution, no matter the value of $\widehat{\beta}_j$ we are evaluating, no matter its variance

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- The estimated value, minus the true value, divided by the standard error of the estimated value, is always distributed as a t with $n - k - 1$ degrees of freedom

Inference

- Standardizing the random variable means that we only have to look at one distribution, no matter the value of $\widehat{\beta}_j$ we are evaluating, no matter its variance

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- The estimated value, minus the true value, divided by the standard error of the estimated value, is always distributed as a t with $n - k - 1$ degrees of freedom
- We can use this distribution to evaluate the *same* question: "how likely are we to have pulled a value at least as large as 2.01 from this distribution?"

Inference

- When we start asking questions like we've just been asking — evaluating the likelihood of observing the values we observe — we are one baby step away from hypothesis testing

- Language of hypothesis testing . . .
- We have a “null hypothesis” — think of this as our default position

- Language of hypothesis testing . . .
- We have a "null hypothesis" — think of this as our default position
- We can use our knowledge of the distribution of $\widehat{\beta}_j$ (and its standardized cousin) to *reject* the null hypotheses

- Language of hypothesis testing ...
- We have a "null hypothesis" — think of this as our default position
- We can use our knowledge of the distribution of $\widehat{\beta}_j$ (and its standardized cousin) to *reject* the null hypotheses
- All we can do with evidence (data + models) is disprove things — we can never *prove* something affirmatively

- Language of hypothesis testing ...
- We have a "null hypothesis" — think of this as our default position
- We can use our knowledge of the distribution of $\widehat{\beta}_j$ (and its standardized cousin) to *reject* the null hypotheses
- All we can do with evidence (data + models) is disprove things — we can never *prove* something affirmatively
- Here is how we write a null hypothesis (carrying over from our graphical example)

$$H_0 : \beta_j = 1$$

- When we write a null hypothesis (a “default” position), we need an *alternative* hypothesis

- When we write a null hypothesis (a “default” position), we need an *alternative* hypothesis
- Alternative hypothesis: if we reject the null (reject H_0), what will we be rejecting it in favor of

- When we write a null hypothesis (a “default” position), we need an *alternative* hypothesis
- Alternative hypothesis: if we reject the null (reject H_0), what will we be rejecting it in favor of

$$H_0 : \beta_j = 1$$
$$H_A : \beta_j \neq 1$$

- Recall again our regression results from earlier
- (type `summary(m1)`)
- What is $t_{\widehat{\beta_1}}$ for the null that $\widehat{\beta_1} = 1$?

- Recall again our regression results from earlier
- (type `summary(m1)`)
- What is $t_{\widehat{\beta_1}}$ for the null that $\widehat{\beta_1} = 1$?

$$\frac{2.2839 - 1}{0.5641} = 2.276015$$

- Recall again our regression results from earlier
- (type `summary(m1)`)
- What is $t_{\widehat{\beta_1}}$ for the null that $\widehat{\beta_1} = 1$?

$$\frac{2.2839 - 1}{0.5641} = 2.276015$$

- To verify: `(2.2839-1)/0.5641` or use `coeff(summary(m1))` tricks

- Now, take the t-statistic and compare it to the critical value on page 833 of Wooldridge
- If the value of the t-statistic is greater (in absolute value) than the critical value, reject the null

- So, it turns out that this expression is a really useful way to summarize what we know about the distribution of $\widehat{\beta}_j$

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- We can use it to test hypotheses about β_j
- The most common null hypothesis is $\beta_j = 0$

- So, it turns out that this expression is a really useful way to summarize what we know about the distribution of $\widehat{\beta}_j$

$$\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)} \sim t_{n-k-1}$$

- We can use it to test hypotheses about β_j
- The most common null hypothesis is $\beta_j = 0$
- This is akin to testing that the j^{th} x-variable has *no effect* on *y*
- Proving otherwise (i.e. proving that the x-variable does have some effect on *y*) is often a powerful conclusion

- Does education have any effect on earnings?
  - Rejecting the null bolsters arguments for spending money on education

- Does education have any effect on earnings?
  - Rejecting the null bolsters arguments for spending money on education
- Does access to micro-credit increase the (long-term) income of people in developing countries?
  - Rejecting the null argues in favor of micro-credit

- The most common null hypothesis is that the true value of β_j is zero

$$H_0 : \beta_j = 0$$

- The most common alternative hypothesis is

$$H_A : \beta_j \neq 0$$

Testing null hypotheses in R

- Using `lecture-06-dset-01.RData`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$

Testing null hypotheses in R

- Using `lecture-06-dset-01.RData`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy.

Testing null hypotheses in R

- Using `lecture-06-dset-01.RData`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 833 of Wooldridge

Testing null hypotheses in R

- Using `lecture-06-dset-01.RData`
- Test the null hypothesis that $\beta_1 = 0$ against the alternative hypothesis that $\beta_1 \neq 0$
- Easy. Take the t-statistic reported in the row labeled `x1` and compare it to the critical value on page 833 of Wooldridge
- If the t-statistic is greater than the critical value, then we reject the null (we reject H_0)

Degrees of freedom

- To look up the proper critical value you need to know one thing: the degrees of freedom of the test (the number of observations we are basing the test statistic on)

Degrees of freedom

- To look up the proper critical value you need to know one thing: the degrees of freedom of the test (the number of observations we are basing the test statistic on)
- We are using one degree of freedom to estimate the intercept in the model, one to estimate β_1, and one to estimate β_2. We started out with 100 observations. How many are left to use for inference? 100 - 3 = 97.

Degrees of freedom
What you talkin' 'bout?

- You start with 100 observations
- Then you estimate the value of β_0

Degrees of freedom

What you talkin' 'bout?

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean

Degrees of freedom

What you talkin' 'bout?

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean
- I give you the first observation: 7
- I give you my estimate of the mean: 8
- What was the other observation in my sample?

- You start with 100 observations
- Then you estimate the value of β_0
- If you gave me the first 99 observations in your dset, plus your estimate of β_0, I could tell you what the 100th observation in your dset was
- You used-up an observation when you calculated $\widehat{\beta_0}$
- What? Think of it this way:
- Suppose I had a sample composed of two observations (crappy dset, I know, but this is a toy example)
- I use this observation to estimate the population mean
- I give you the first observation: 7
- I give you my estimate of the mean: 8
- What was the other observation in my sample?
- Seeeee?

Looking up the t-stat

- We know the degrees of freedom
- We decide on the significance level of the test (0.05 is very common, but there is absolutely, positively nothing special about it)
- We look up the critical value in our t-table
- If our t-stat is greater in absolute value than the critical t, then we reject the null
- For our test, the critical t is somewhere in the neighborhood of 2
- So we can reject our null like nobody's business

Looking up the t-stat

- We know the degrees of freedom
- We decide on the significance level of the test (0.05 is very common, but there is absolutely, positively nothing special about it)
- We look up the critical value in our t-table
- If our t-stat is greater in absolute value than the critical t, then we reject the null
- For our test, the critical t is somewhere in the neighborhood of 2
- So we can reject our null like nobody's business
- Can we tell anything from the fact that our t-stat was so damn big? Hold that thought

Next time

- Finish talking about t-tests
- Next time we will move to testing compound hypotheses
- Properties of OLS
- Then we will look to relax some of the assumptions that we have relied on to get where we are (MLR.1 – MLR.6)
- Then all we have left before spring break is a detailed discussion of functional form: esp. dummy variables and interaction terms

Homework

- Homework #4 posted on net (15 problems)
- Merge your gdp data (the data you turned in this week) with temperature data that I send you via email