Applied Econometrics

Lecture 2

Nathaniel Higgins

ERS and JHU

13 September 2010

Artefactual world

- Let’s simulate a world where we know the ”true” parameters
-

Artefactual world

Stata code

```
* Hospital example:
* Let's create an example (from scratch) with
* 30 people from the streets of DC
* In Stata, this means that we need to start
* with a blank dset, and declare that we will
* create 30 observations
clear
set obs 30
* Let's create an id number for all of the
* individuals in our dset so that we can refer
* to them easily
gen id = _n
```

Artefactual world

Stata code

```
* Instead of inventing every single piece of
* data from scratch, we'll use a random number
* generator to help make some of our data.  To
* make sure that we can replicate our results
* later (this is important!), set the seed
* first.
set seed 12345
* Now let's create some data.  Let's make all
* 30 people either sick or healthy.  Let's
* make it a 50/50 shot that someone is sick.
* We'll create a variable called "sick" that
* takes on value 1 if the person is sick, and
* 0 otherwise.
```

Artefactual world

Stata code

```
* First, initialize the variable by setting
* ALL observations to 0
gen sick = 0
* Then, with probablity 0.5, set the
* variable's value to 1
gen temp = uniform()
replace sick = 1 if temp > 0.5
* Now, let's create a baseline health
* measurement for everybody (we'll make sure
* to account for the fact that sick people
* feel worse in just a moment).
* Let's make the health measurement vary from
* 1 to 5, with equal probability.
gen health = 1+int((5-1+1)*uniform())
```

Artefactual world

Stata code

```
* Now, let's make sure that sick people
* feel worse than healthy people.  Let's make
* them feel 2-worse than they would otherwise
* feel.
replace health = health - 2 if sick == 1
* But we don't want people to have health
* values below 1, so we'll censor the data to
* 1.
replace health = 1 if health < 1
```

Artefactual world

Stata code

```
* Now, let's do an experiment.  Let'send
* half the population to the hospital -- a
* RANDOM half, not just the half that happens
* to be sick
gen hospital = 0
drop temp
gen temp = uniform()
replace hospital = 1 if temp > 0.5
* If sick people go to the hospital, they get
* 1 better.
replace health = health + 1 if ///
(sick == 1 & hospital == 1)
```

Artefactual world

Stata code

```
* OK, we've run the experiment.  Now let's
* analyze the results.
* First, let's analyze them using a simple
* ttest
ttest health, by(hospital)
```

- Magic!
- We get the result we wanted:

- Magic!
- We get the result we wanted:
  - The experiment revealed to us (a very close approximation of) the treatment effect (remember it was 1)

- Magic!
- We get the result we wanted:
  - The experiment revealed to us (a very close approximation of) the treatment effect (remember it was 1)
- Now, just to *preview* the next few lectures, let's see what would have happened if we just used *observational* data to examine the same question ("do hospitals make people better?")

Stata code

```
* Re-run all commands (see why it's good
* to set the seed?)  right up until the part
* where we sent people to the hospital.
set seed 12345
drop sick
gen sick = 0
drop temp
gen temp = uniform()
replace sick = 1 if temp > 0.5
drop health
gen health = 1+int((5-1+1)*uniform())
replace health = health - 2 if sick == 1
replace health = 1 if health < 1
```

Stata code

```
* This time, let whoever is sick go to the
* hospital
drop hospital
gen hospital = 0
replace hospital = 1 if sick == 1
* Hospitals really do make sick people a
* little better
replace health = health + 1 if (hospital == 1)
* Finally, let's test again and see what we
* get
ttest health, by(hospital)
```

- So what happened?
- Why didn't we get the "right" (true) treatment effect when we didn't run an experiment?

- So what happened?
- Why didn’t we get the ”right” (true) treatment effect when we didn’t run an experiment?
- Preview of next time:

- So what happened?
- Why didn't we get the "right" (true) treatment effect when we didn't run an experiment?
- Preview of next time:
  - The people who went to the hospital were sicker than the rest of the population (the "control" group) before they went to the hospital
  - So when we test to see if a hospital visit made people feel better, we were comparing people who went to the hospital (because they were sick) to people that didn't go to the hospital (because they *weren't* sick!)

- So what happened?
- Why didn't we get the "right" (true) treatment effect when we didn't run an experiment?
- Preview of next time:
    - The people who went to the hospital were sicker than the rest of the population (the "control" group) before they went to the hospital
    - So when we test to see if a hospital visit made people feel better, we were comparing people who went to the hospital (because they were sick) to people that didn't go to the hospital (because they *weren't* sick!)
    - This is a bogus comparison
    - We want to compare people that go to the hospital (because they are sick) with what would have happened if they hadn't been able to go to the hospital

- We've seen how, when we run a simple randomized experiment and do a simple t-test, we can recover what we know to be the correct treatment effect
- But we don't often use simple t-tests in practice
- More often we use regression analysis, because it allows us to simultaneously test multiple hypotheses
- We can hold multiple variables "constant" and conduct multiple t-tests (and many other tests) in one fell swoop

Modeling tools

- We used a simple Stata command `ttest health, by(hospital)` to test the effect of the hospital treatment on health status
- This is like saying that the `health` variable is the dependent variable, and the `hospital` variable is the independent variable
- If you're following along in Stata, make sure to re-set your data so that we are in the experimental case

Stata code

```
* Run an OLS regression of health on hospital
regress health hospital
```

Modeling tools

- You should get identical results when you run a simple t-test, and when you run a single-variable regression (with a constant)
- An OLS regression of a single variable is isomorphic to a simple t-test of the hypothesis that there is no difference between the treated group and the control group

Modeling tools

- You should get identical results when you run a simple t-test, and when you run a single-variable regression (with a constant)
- An OLS regression of a single variable is isomorphic to a simple t-test of the hypothesis that there is no difference between the treated group and the control group
- The experiment is the ideal way to prove a causal effect — it is the ideal tool for testing hypotheses
- But (in part) because we cannot (usually) do experiments, and so because we are usually reliant upon observational data, we often have to use more sophisticated modeling techniques to get at the statistical tests we are interested in

Modeing tools

- We are now set to begin reviewing (probably in a unique way for most of you) how the OLS estimator works
- First, the technical stuff, then the conceptual stuff
- Unless you really like matrix algebra, brace yourself for boredom

Modeling

- Model the dependent variable as a function of observables and unobservables
- For a given observation in our data (the i^{th} observation):

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k + \epsilon_i$$

Modeling

- Model the dependent variable as a function of observables and unobservables
- For a given observation in our data (the i^{th} observation):

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \ldots + x_{ik}\beta_k + \epsilon_i$$

- If the i^{th} observation is modeled as above, then all the data can be collected and expressed like this:

$$\begin{gathered} y_1 = \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \ldots + x_{1k}\beta_k + \epsilon_1 \\ \vdots \\ y_n = \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{nk}\beta_k + \epsilon_n \end{gathered}$$

Modeling with matrix notation

- We express our model (in terms of our data) in matrix notation like this:

$$\boldsymbol{y} = \boldsymbol{\iota}\beta_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{\iota} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Modeling with matrix notation

- Put it all together to see how matrix notation

$$\boldsymbol{y} = \boldsymbol{\iota}\beta_0 + \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

- ...expands to give us:

$$y_1 = \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + \ldots + x_{1k}\beta_k + \epsilon_1$$
$$\vdots$$
$$y_n = \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 + \ldots + x_{nk}\beta_k + \epsilon_n$$

Modeling with matrix notation

- In matrix notation we often combine the vector of ones ($\boldsymbol{\iota}$) with the $\boldsymbol{X}$ matrix, and the constant (β_0) with the $\boldsymbol{\beta}$ vector so that we can express things even more compactly:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{1k} \\ \vdots & & & \vdots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Modeling with matrix notation

- Verify that the matrix expression below multiplies-out to give you

$$y_1 = \beta_0 + x_{11}\beta_1 + x_{12}\beta_2 + ... + x_{1k}\beta_k + \epsilon_1$$
$$\vdots$$
$$y_n = \beta_0 + x_{n1}\beta_1 + x_{n2}\beta_2 + ... + x_{nk}\beta_k + \epsilon_n$$

- Do the matrix addition/multiplication for yourself:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{1k} \\ \vdots & & & \vdots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

OLS estimator

- If our model of the *real* process is $\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$
- then our (imperfect) expression of that model is denoted $\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{e}$
- That is, we would like to find a good approximation of β, which we'll call $\boldsymbol{b}$
- and since we can't observe ϵ, we'll call the difference between our observed dependent variable $\boldsymbol{y}$ and our estimate $\boldsymbol{Xb}$ the "error" $\boldsymbol{e}$

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
  - The OLS estimator is the estimator that ...

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
    - The OLS estimator is the estimator that . . . minimizes the sum of squared errors.

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
  - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
    - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

1. Express the sum of squared errors in matrix notation

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
  - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

1. Express the sum of squared errors in matrix notation
   - There is an easy way to do this.

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
  - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

1. Express the sum of squared errors in matrix notation
   - There is an easy way to do this. To square a vector, you simply multiply it by its transpose:

OLS estimator

- Derive the OLS estimator ($\boldsymbol{b}$) using matrix notation
- Steps:
  - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

1. Express the sum of squared errors in matrix notation
   - There is an easy way to do this. To square a vector, you simply multiply it by its transpose:
   - $\boldsymbol{e}'\boldsymbol{e}$

OLS estimator

- Derive the OLS estimator (***b***) using matrix notation
- Steps:
    - The OLS estimator is the estimator that . . . minimizes the sum of squared errors. So we:

1. Express the sum of squared errors in matrix notation
    - There is an easy way to do this. To square a vector, you simply multiply it by its transpose:
    - $\boldsymbol{e}'\boldsymbol{e}$
    - You can see that this is true if you look at the expansion:

$$\begin{pmatrix} e_1 & e_2 & \dots & e_n \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

OLS estimator

- So we are going to minimize the sum of squared errors: $\boldsymbol{e}'\boldsymbol{e}$
- But first, we need to express the minimization problem in terms of our choice vector $\boldsymbol{b}$

OLS estimator

- So we are going to minimize the sum of squared errors: $\boldsymbol{e}'\boldsymbol{e}$
- But first, we need to express the minimization problem in terms of our choice vector $\boldsymbol{b}$
- If $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$, then $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$ and:

OLS estimator

- So we are going to minimize the sum of squared errors: $\boldsymbol{e}'\boldsymbol{e}$
- But first, we need to express the minimization problem in terms of our choice vector $\boldsymbol{b}$
- If $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$, then $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$ and:
- $\boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$

OLS estimator

- So we are going to minimize the sum of squared errors: $\boldsymbol{e}'\boldsymbol{e}$
- But first, we need to express the minimization problem in terms of our choice vector $\boldsymbol{b}$
- If $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$, then $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$ and:
- $\boldsymbol{e}'\boldsymbol{e} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$

2. To find the minimum of the SSE, we take its derivative with respect to $\boldsymbol{b}$
   - $\frac{d}{d\boldsymbol{b}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$

OLS estimator

Taking the derivative of SSE w.r.t $\boldsymbol{b}$

$$\frac{d}{d\boldsymbol{b}}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$$

Expand

$$= \frac{d}{d\boldsymbol{b}} \{\boldsymbol{y'y} \text{ - } \boldsymbol{y'Xb} \text{ - } \boldsymbol{b'X'y} + \boldsymbol{b'X'Xb}\}$$

Take derivative and set to zero

$$\text{- } (\boldsymbol{y'X})\text{'} \text{ - } \boldsymbol{X'y} + 2\boldsymbol{X'Xb} = 0$$

Rearrange

$$2\boldsymbol{X'Xb} = (\boldsymbol{y'X})\text{'} + \boldsymbol{X'y}$$

$$2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = (\boldsymbol{y}'\boldsymbol{X})' + \boldsymbol{X}'\boldsymbol{y}$$

Consolidate

$$2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = 2\boldsymbol{X}'\boldsymbol{y}$$

Divide by 2

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$$

Pre-multiply by $(\boldsymbol{X}'\boldsymbol{X})^{-1}$

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

OLS estimator

- What are the properties of this estimator?

OLS estimator

- What are the properties of this estimator?
- B.L.U.E.

OLS estimator

- What are the properties of this estimator?
- B.L.U.E.
  - Best, Linear, Unbiased Estimator
- The ”Linear” part is pretty obvious
- Let’s examine ”unbiasedness”

OLS estimator

- What are the properties of this estimator?
- B.L.U.E.
  - Best, Linear, Unbiased Estimator
- The "Linear" part is pretty obvious
- Let's examine "unbiasedness"
  - $\boldsymbol{b}$ is a random variable

OLS estimator

- What are the properties of this estimator?
- B.L.U.E.
  - Best, Linear, Unbiased Estimator
- The "Linear" part is pretty obvious
- Let's examine "unbiasedness"
  - $\boldsymbol{b}$ is a random variable
  - That means that if we took a new sample of $\boldsymbol{y}$, we'd get a different $\boldsymbol{b}$

OLS estimator

- What are the properties of this estimator?
- B.L.U.E.
  - Best, Linear, Unbiased Estimator
- The ”Linear” part is pretty obvious
- Let’s examine ”unbiasedness”
  - $\boldsymbol{b}$ is a random variable
  - That means that if we took a new sample of $\boldsymbol{y}$, we’d get a different $\boldsymbol{b}$
  - So what we’d like to know is: What, *on average* is the value of $\boldsymbol{b}$?

What, on average, is the value of $\boldsymbol{b}$?

OLS estimator
Unbiased

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}]$$

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + \epsilon)]$$

OLS estimator
Unbiased

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + \epsilon)]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon]$$

OLS estimator
Unbiased

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + \epsilon)]$$

$$= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\beta] + E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon]$$

$$= E[\beta] + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(E[\epsilon)]$$

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}]$$

$$= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{X}\beta + \epsilon)]$$

$$= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}\beta] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon]$$

$$= E[\beta] + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(E[\epsilon)]$$

$$= \beta + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(0)$$

What, on average, is the value of $\boldsymbol{b}$?

$$E[\boldsymbol{b}] = E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'y}]$$

$$= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{X}\beta + \epsilon)]$$

$$= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}\beta] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon]$$

$$= E[\beta] + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(E[\epsilon)]$$

$$= \beta + (\boldsymbol{X'X})^{-1}\boldsymbol{X'}(0)$$

$$= \beta$$

OLS estimator

- OK, so we have a point estimator, and it's unbiased. What is the variability of the estimator?

OLS estimator

- OK, so we have a point estimator, and it's unbiased. What is the variability of the estimator?
- That is, how much can we expect this estimate to vary from the "true" value, in repeated samples?

Recall:

OLS estimator

- OK, so we have a point estimator, and it's unbiased. What is the variability of the estimator?
- That is, how much can we expect this estimate to vary from the "true" value, in repeated samples?

Recall:

1. Formula for variance of a random variable z is $E[(z - E[z])^2]$

OLS estimator

- OK, so we have a point estimator, and it's unbiased. What is the variability of the estimator?
- That is, how much can we expect this estimate to vary from the "true" value, in repeated samples?

Recall:

1. Formula for variance of a random variable z is $E[(z - E[z])^2]$
2. Formula for variance-covariance of a vector of random variables $\boldsymbol{z}$ is $E[(\boldsymbol{z} - E[\boldsymbol{z}])(\boldsymbol{z} - E[\boldsymbol{z}])']$

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

$$E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon\epsilon'\boldsymbol{X}(\boldsymbol{X'X})^{-1}]$$

OLS estimator

See what the variance-covariance of ***b*** is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

$$E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\epsilon\epsilon'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E[\epsilon\epsilon']\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

$$E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon\epsilon'\boldsymbol{X}(\boldsymbol{X'X})^{-1}]$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}E[\epsilon\epsilon']\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

OLS estimator

See what the variance-covariance of ***b*** is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

$$E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon\epsilon'\boldsymbol{X}(\boldsymbol{X'X})^{-1}]$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}E[\epsilon\epsilon']\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

$$\sigma^2(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X})^{-1}$$

OLS estimator

See what the variance-covariance of $\boldsymbol{b}$ is:

$$E[(\boldsymbol{b} - E[\boldsymbol{b}])(\boldsymbol{b} - E[\boldsymbol{b}])']$$

$$E[(\boldsymbol{b} - \beta)(\boldsymbol{b} - \beta)']$$

$$E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\epsilon\epsilon'\boldsymbol{X}(\boldsymbol{X'X})^{-1}]$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}E[\epsilon\epsilon']\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

$$(\boldsymbol{X'X})^{-1}\boldsymbol{X'}\sigma^2\boldsymbol{I}_n\boldsymbol{X}(\boldsymbol{X'X})^{-1}$$

$$\sigma^2(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}(\boldsymbol{X'X})^{-1}$$

$$\sigma^2(\boldsymbol{X'X})^{-1}$$

OLS estimator

- What does this amount to if we are in the simplest possible case?
- If we have a collection of data $\boldsymbol{y}$, and we regress it on a vector of ones $\boldsymbol{\iota}$, then we get a single coefficient estimate b, of the mean of $\boldsymbol{y}$
- Then the variance of b is simply equal to σ^2
- But wait, what is σ^2 again? σ^2 is the variance of ϵ
- We don't observe ϵ, so we don't observe σ^2
- But we can *estimate* it
- If we want to estimate σ^2, we do it using our observations of e

OLS estimator

- Our estimate of σ^2, which we will denote $\hat{\sigma}^2$, is equal to the average squared error
- If we have our estimate of β (b), then we have a vector of errors $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\beta$
- We want to know the empirical variance of the error:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2$$

OLS estimator

- So we know the point estimate *b* (our estimate of the average of *y*), we know the estimate of the variability of *b*, $(\boldsymbol{X'X})^{-1}\hat{\sigma}^2$ . . .
- How do we use these things to test hypotheses?
- t-statistics
  - (we can do other things, but this is the first test almost everybody learns)

- A t-test of the null hypothesis that $y = 0$ can be carried out simply by dividing the empirical mean of y, $\bar{y}$, by the standard error of the mean of random variable y:

$$\frac{\bar{y}}{std.err.} \sim t$$

- This is the same thing we get when we use the OLS estimator b and observe the t-statistic based on $\hat{\sigma}^2(\boldsymbol{X'X})^{-1}$

- To see that this is true, just think about what we get when we calculate $(\boldsymbol{X'X})^{-1}$
- Remember, $\boldsymbol{X}$ is just a vector of n ones (it's just ι)

$$\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$= 1 + 1 + \ldots + 1 = n$$

- So $\boldsymbol{X'X}$ is just n
- and $(\boldsymbol{X'X})^{-1}$ is $\frac{1}{n}$
- and the variance-covariance matrix just simplifies to $\frac{\hat{\sigma}^2}{n}$

OLS estimator

- If you divide the estimate of the mean of y by $\frac{\hat{\sigma}}{\sqrt{n}}$, you'll get a number that is identical to the t-statistic that you get when you regress y on a constant
- In sum:
  - When you want to get the coefficient estimate of the mean of y, as well as the standard error and the t-statistic of the coefficient estimate, you simply:
    1. Estimate the mean of y from your data, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
    2. Find the standard error of your estimate of the mean, $std.err. = \frac{\hat{\sigma}}{\sqrt{n}}$
    3. Divide your estimate of the man by the standard deviation $\frac{\bar{y}}{std.err.}$, and you get the t-statistic

OLS estimator

- So we've established that regressing a dependent variable y on a vector of ones is essentially identical to running a t-test to test the hypothesis that $y \neq 0$
- Using OLS regression techniques, we can examine several t-tests (to test several independent hypotheses) all at once

OLS estimator

- So we've established that regressing a dependent variable y on a vector of ones is essentially identical to running a t-test to test the hypothesis that $y \neq 0$
- Using OLS regression techniques, we can examine several t-tests (to test several independent hypotheses) all at once
    - If $(\boldsymbol{X'X})^{-1}$ is a $k \times k$ matrix instead of just the scalar n, then when we calculate $\hat{\sigma}^2(\boldsymbol{X'X})^{-1}$ we get the necessary ingredients to test hypotheses about several coefficients at once

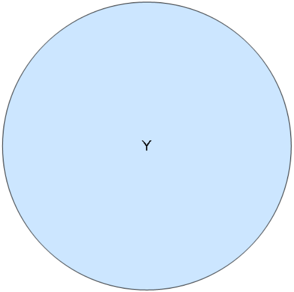
OLS estimator

- So we've established that regressing a dependent variable y on a vector of ones is essentially identical to running a t-test to test the hypothesis that $y \neq 0$
- Using OLS regression techniques, we can examine several t-tests (to test several independent hypotheses) all at once
  - If $(\boldsymbol{X'X})^{-1}$ is a $k \times k$ matrix instead of just the scalar n, then when we calculate $\hat{\sigma}^2(\boldsymbol{X'X})^{-1}$ we get the necessary ingredients to test hypotheses about several coefficients at once
  - HINT, HINT
  - You can find out how to get a vector of t-statistics by analogy
  - All the elements are here, in the lecture slides and in your notes
  - You'll need to understand this to do the homework

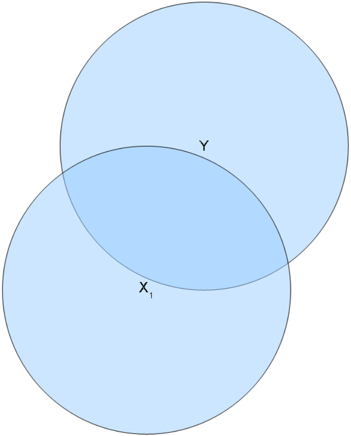
- But what about more complicated model building?
- What about compound hypotheses?
- Let's put the technical stuff on hold for a little while and think conceptually

OLS estimator

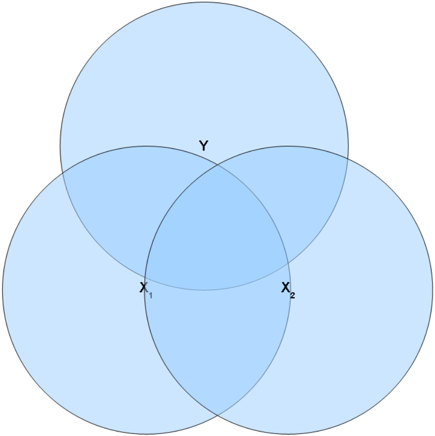
Graphically



OLS estimator
Graphically



OLS estimator
Graphically



Stata code

```
* Set seed
set seed 12345
* Create a matrix of correlations
matrix C = (1, 0.2, 0.2 \ 0.2, 1, 0.2 \ ///
0.2, 0.2, 1)
* Create a matrix of means
matrix m = (3,2,2)
* Create a matrix of standard deviations
matrix sd = (0.5,2,1)
* Draw three random variable from a
* multivariate distribution
drawnorm x1 x2 x3, n(100) means(m) ///
sds(sd) corr(C)
* Draw some "unobservable" stuff
gen eps = rnormal()
```

Stata code

```
* Create a dependent variable y
gen y = 5 + 2*x1 - 3*x2 + eps
* Regress y on x1 (by itself)
regress y x1
* Regress y on x2 (by itself)
regress y x2
* Regress y on x1 and x2
regress y x1 x2
```

Homework

- Write a Stata program that runs an OLS regression, calculates the variance-covariance matrix of $\boldsymbol{b}$, and calculates t-statistics of the individual elements of $\boldsymbol{b}$
- ALL three of these elements should be printed to the screen
- That is, your program should print to the Stata results window:
  1. coefficient estimates $b_0, b_1, \ldots, b_k$
  2. the variance-covariance matrix $\sigma^2(\boldsymbol{X'X})^{-1}$
  3. and the t-statistics corresponding to the coefficient estimates $b_0, b_1, \ldots, b_k$
- Program should take as inputs: $y, x_1, x_2, \ldots, x_k$

Homework

- To make your life easier, you don't need to write a fully general program like `regress`
- Instead, you can write a program that works using data from today's lecture:
  - Use the data we created during the "Try it" section of today's lecture (slide 82-83)
  - and write an OLS program that produces results identical to what you get if you were to type `regress y x1 x2` at the command line of Stata
- Reading to make this easier: Appendices A and B
- If you get all that done and want to be prepared for next time, start reading about IV (Ch. 6)