Econometrics II: Econometric Modelling

Jürgen Meinecke

Research School of Economics, Australian National University

3 August, 2018

Stata

For those of you who are not familiar with Stata:

- Visit the class website and click on “Stata help”
- There you will find resources to teach yourself Stata
- Dedicate some time to teach yourself Stata
- Feel free to stop by my office if you need help
- I’m also happy to use the weekly computer tutorials to answer your Stata related questions

Problem Solving Tutorial

Unlike in EMET2007, there is no separate dedicated problem solving tutorial in EMET3004

During the first half of the semester, I will use the last 15 or 20 minutes during lectures to work through some problem solving exercises

Roadmap

Introduction

The Big Ideas from STAT1008 and EMET2007

- Statistical Inference
- Our Goal: Learning about the Causal Effect
- Identifying Causal Effects with the Linear Regression Model
- Brief Review of OLS
- Estimator Goodness: Bias, Efficiency and Consistency

Endogeneity: When OLS fails

- Internal Validity

The problem of statistical inference can be expressed like this:

- we want to learn something about the population
- but we do not observe the population
- instead we only observe a random sample drawn from the population
- the random sample is a subset of the population
- we need to use that random subset to approximate the population

Definition

The problem of **statistical inference** consists of using a random sample to learn about statistical parameters of the unobserved population.

What do we mean by ‘statistical parameters’?

- mean
- variance
- moments

In at least 80% of all cases we are interested in the mean

Example: What is the mean weight of Tidbinbilla roos?

Suppose the park rangers want to know the answer to that question and hire us to come up with an answer

They give us permission to randomly collect 30 roos

(It is out of the question to collect ALL roos, we therefore do not observe the entire population)

Wouldn't it seem reasonable to use the average weight in our sample as our best guess of the mean weight of Tidbinbilla roos?

The roo example illustrates common terminology

- We want to learn about the population mean $\mathrm{E}[Y]$
- We have no hope of knowing this mean
  b/c we do not observe the entire population
- the population mean is *unobserved*
- we do, however, observe the sample average $\bar{Y}$
- We use $\bar{Y}$ as an *estimator* of the population mean
- Given our particular random sample of 30 roos,
  the sample average takes on the value of, say, 70kg
- That value is our *estimate* of the population mean

Roadmap

Introduction

The Big Ideas from STAT1008 and EMET2007

- Statistical Inference
- Our Goal: Learning about the Causal Effect
- Identifying Causal Effects with the Linear Regression Model
- Brief Review of OLS
- Estimator Goodness: Bias, Efficiency and Consistency

Endogeneity: When OLS fails

- Internal Validity

In regression analysis, we study the relationship between several variables

At the very minimum, we study the relationship between X and Y

Actually, we study how X affects Y
(not the other way around)

More precisely, we want to *quantify* the *causal effect* of X on Y

In each application, we posit that we are interested in some causal effect

We use OLS as our tool to estimate that causal effect

The concept of causal effect is key to EMET2007 and EMET3004

It is important that you understand it well

Behind every estimation that we run is the hope that we estimate some interesting causal effect

But let's be more precise about what we mean by causal effect...

Recall from EMET2007 (lecture 3) that we think of the relationship between X and Y as follows:

$$Y_i = f(X_i, u_i),$$

where

- $f(\cdot)$ is the response function
- Y_i is the dependent variable
- X_i is the independent variable
- u_i is the error term

X_i and Y_i are observed in the data; u_i is unobserved

Remember: u_i captures all other things that explain Y_i (over and above X_i)

Digression: why do we need to include u_i?

If we did not include u_i as part of the function $f(\cdot)$ then we would presume that the relationship b/w Y_i and X_i was *deterministic*

It would mean that once we know X_i we also know Y_i

This is almost like saying that they are one and the same thing

Deterministic relationships often make sense in the natural sciences, example: relationship b/w Celsius and Fahrenheit

In economics, relationships b/w variables are never deterministic but are subject to some degree of randomness and the presence of the *error term* u_i allows for that

We are now ready to define what we mean by causal effect:

Definition

Let x be some real number. The **individual causal effect** of X on Y is given by

$$\Delta(x, u_i) := f(X_i = x+1, u_i) - f(X_i = x, u_i).$$

Intuition:
change in the function value as X_i is increased by 1 from x, keeping all unobserved factors u_i constant

The causal effect depends on the starting value x and on unobserved factors u_i

From EMET1001 you should know that mathematically speaking, $\Delta(x)$ is the *difference quotient* of the function $f(\cdot)$

The difference quotient measures the finite slope of a function

The causal effect therefore is the slope at $X = x$ of the response function $f(\cdot)$ with respect to X

That should be intuitive:
the slope captures the notion of responsiveness of a function with respect to its argument

This is what the notion of causal effect is supposed to capture

My definition of causal effect is based on a discrete random variable X_i

Alternatively, I could have defined the causal effect based on a continuous random variable

Then I would have defined the causal effect as the derivative with respect to X_i

This is slightly more complicated but does not yield any extra insights

Example

Let's say we are interested in the relationship between earnings and heights (we have looked at that in EMET2007, remember?)

Specifically:

- X_i: height of person i
- Y_i: hourly wage of person i

Then $f(X_i, u_i)$ is the function that maps heights and unobserved stuff u_i into earnings

What is the causal effect?

I could be interested in the following causal effect

$$\Delta(172, u_i) := f(X_i = 173, u_i) - f(X_i = 172, u_i)$$

Intuition:
By how much would earnings for person i, who is 172 cm tall and has unobservables u_i, increase if s/he had been 1 cm taller?

(Why would you think that this is an interesting research question?)

Alternatively, I could be interested in the causal effect

$$\Delta(182, u_i) := f(X_i = 183, u_i) - f(X_i = 182, u_i)$$

Intuition:
By how much would earnings for person i, who is 182 cm tall and has unobservables u_i, increase if s/he had been 1 cm taller?

The causal effect depends on the starting value x, so generally $\Delta(172, u_i) \neq \Delta(182, u_i)$

It is an *individual* causal effect because it depends on a person's individual unobserved value u_i

More examples of causal effects we have studied in EMET2007

- Causal effect of smoking during pregnancy on birth outcomes of babies
- Causal effect of trade openness of countries on real gdp growth rate
- Causal effect of race on job interview success rates in Australia
- Causal effect of lead consumption on infant mortality

Roadmap

Introduction

The Big Ideas from STAT1008 and EMET2007

- Statistical Inference
- Our Goal: Learning about the Causal Effect
- Identifying Causal Effects with the Linear Regression Model
- Brief Review of OLS
- Estimator Goodness: Bias, Efficiency and Consistency

Endogeneity: When OLS fails

- Internal Validity

What have we just learned?
Causal effects are the things we are after

But how do we learn about them?

It turns out that once we impose a specific structure on the function $f(\cdot)$, the causal effect can easily be identified

You know from EMET2007 that we impose linearity on $f(\cdot)$

Let $f(X_i, u_i) = \beta_0 + \beta_1 X_i + u_i$

Then it is easy to derive that

$$\Delta(x, u_i) = \beta_1$$

This results looks innocuous at first, but actually is a bit more profound than it looks:

- Once you impose linearity, the causal effect is not effectively a function in x and u_i anymore
- The causal effect does not depend on the particular value that X_i takes on
- The causal effect is the same for all subjects (e.g. people)

For that reason we now refer to it as the *average* causal effect

Linearity imposes the notion of ’averageness’:
the population parameter β_1 is the causal effect for the average person in the sample

This result is so important that we dedicate a theorem to it:

Theorem

In the linear model, the coefficient β_1 captures the ***average causal effect*** *of X on Y.*

Going back to the example of earnings and heights, we looked at the two individual causal effects

$$\Delta(172, u_i) := f(X_i = 173, u_i) - f(X_i = 172, u_i)$$

$$\Delta(182, u_i) := f(X_i = 183, u_i) - f(X_i = 182, u_i)$$

Once we impose linearity (that is, $f(X_i, u_i) = \beta_0 + \beta_1 X_i + u_i$,) this collapses to

$$\Delta(172, u_i) = \Delta(182, u_i) = \beta_1$$

The causal effect of growing by one centimeter is effectively assumed to be the same for every person in the sample

Another way to think of this is that β_1 measures the effect of growing by one centimeter for the average person in the sample

Roadmap

Introduction

The Big Ideas from STAT1008 and EMET2007

- Statistical Inference
- Our Goal: Learning about the Causal Effect
- Identifying Causal Effects with the Linear Regression Model
- Brief Review of OLS
- Estimator Goodness: Bias, Efficiency and Consistency

Endogeneity: When OLS fails

- Internal Validity

The parameter β_1 is the population regression coefficient

Because we do not observe the population, we do not know β_1

How can we learn about it?

Learning about an unknown population coefficient is the goal of statistical inference (as we have seen last week)

When the unknown population coefficient is part of a linear model, then there is one dominant method to learn about the unknown population coefficient: estimation via OLS

The whole point of EMET2007 was to expose you to OLS estimation

OLS is one of many ways to estimate β_1

Let's do a brief review of OLS and its properties …

(EMET2007 lecture 4)

Definition

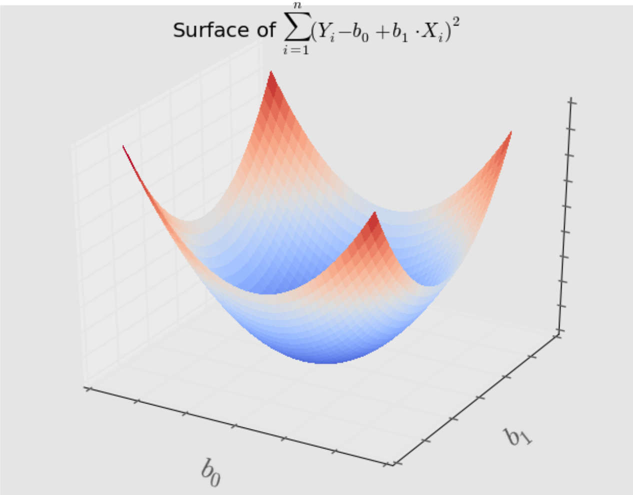
The **Ordinary Least Squares (OLS) estimators** are defined by

$$\hat{\beta}_0, \hat{\beta}_1 := \operatorname*{argmin}_{b_0, b_1} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

In words

- we look at the rhs as a function in b_0 and b_1
- that function happens to be quadratic
- we find the values of b_0 and b_1 that minimize that function
- the values that minimize that function are called solution
- we give the solution a specific name: $\hat{\beta}_0$ and $\hat{\beta}_1$

Geometry of the minimization problem



The single point at the very bottom (the unique minimum) is denoted $(\hat{\beta}_0, \hat{\beta}_1)$

Digression

In today’s lecture I only use the *simple* linear regression model

I could have used the *multiple* linear regression model (with alltogether *k* regressors) instead

Using the simple model instead of the multiple model is without loss of generality

Using the simple model makes the notation a little bit easier

The key ideas and the math carries through straightforwardly

The mathematics of finding the solution

The basic approach is *multivariate calculus* which you know from high school or EMET1001 or both

First step: differentiate $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2$ wrt b_0, b_1

Second step: set derivatives equal zero (obtain the foc)

Third step: solve

Fourth step: clean up

End result:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

The OLS estimator of the slope is equal to the ratio of sample covariance and sample variance!

The OLS estimators are functions of the sample data only

Given the sample data (X_i, Y_i) we can first compute the rhs for $\hat{\beta}_1$ and then we can compute the rhs for $\hat{\beta}_0$

Computer programs such as Stata easily calculate the rhs for you

Roadmap

- Introduction
- The Big Ideas from STAT1008 and EMET2007
  - Statistical Inference
  - Our Goal: Learning about the Causal Effect
  - Identifying Causal Effects with the Linear Regression Model
  - Brief Review of OLS
  - Estimator Goodness: Bias, Efficiency and Consistency
- Endogeneity: When OLS fails
  - Internal Validity

Given an infinitely large set of possible estimators for β_1, why would we use this complicated looking OLS procedure?

As you know already, it turns out that the OLS estimator has some desirable properties

We assess ‘goodness’ of an estimator by three properties:

1. bias
2. variance
3. consistency

Let’s look at these in turn

Definition

An estimator $\hat{\theta}$ for an unobserved population parameter θ is **unbiased** if its expected value is equal to θ, that is

$$\mathrm{E}[\hat{\theta}] = \theta$$

If we draw lots of random samples of size n we obtain lots of estimates $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \ldots$

If the estimator $\hat{\theta}$ is unbiased, then the mean of these estimates will be equal to θ

Note that this is only a thought exercise, in reality we will not draw lots of random samples (we only have one available)

Definition

An estimator $\hat{\theta}$ for an unobserved population parameter θ has **minimum variance** if its variance is (weakly) smaller than the variance of any other estimator of θ. Sometimes we will also say that the estimator is **efficient**.

In EMET2007 I gave this definition of consistency:

Definition

An estimator $\hat{\theta}$ for an unobserved population parameter θ is **consistent** if it converges in probability to θ.

Consistency is difficult to understand

Here is a useful way to look at it:
If, in a thought experiment, you observe the entire population and apply your estimator to it, you want the resulting estimate to be equal to θ

Let's be slightly more technical (and therefore more precise)

Definition (Convergence in Probability)

Let $\hat{\theta}$ be an estimator of θ. We say that $\hat{\theta}$ **converges in probability** to θ if

$$\Pr(|\hat{\theta} - \theta| > \varepsilon) \to 0 \text{ for all } \varepsilon > 0.$$

We write $\hat{\theta} \xrightarrow{p} \theta$ and say that $\hat{\theta}$ is a **consistent** estimator of the population parameter θ.

Unbiasedness and consistency can seem like the same thing

But they aren’t
(admittedly, they ’feel’ similar)

Key distinction:

- consistency is a probabilistic statement about the estimator
- consistency asks, what would you be estimating if you had an infinitely sized random sample available?
- unbiasedness is a statement about the expected value of the estimator

We’ll study that distinction at the end of today’s lecture

Important result about the ’goodness’ of the OLS estimator: (EMET2007 lecture 6)

Theorem

Under OLS Assumptions 1 through 4a, the OLS estimator

$$\hat{\beta}_0, \hat{\beta}_1 := \underset{b_0,b_1}{argmin} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

is BLUE.

The Gauss-Markov theorem provides a theoretical justification for using OLS

Digression: do you remember the 4 Assumptions?

1. conditional mean independence (CMI):

   $$\mathrm{E}[u_i|X_i] = E[u_i]$$

2. sample data are i.i.d. draw from population distribution
3. finite fourth moments (large outliers are unlikely)
4. homoskedasticity

Putting together the pieces of the puzzle:

- our research objective is the causal effect of X_i on Y_i
- generically there exists a functional relationship between the two: $Y_i = f(X_i, u_i)$
- to make our lives easier, we assume that $f(\cdot)$ is linear
- then the causal effect boils down to the parameter β_1 and can be interpreted as the *average* causal effect
- to estimate that parameter we use OLS
- we obtain the estimate $\hat{\beta}_1$ which is our estimate of the causal effect β_1
- by the Gauss-Markov theorem, the estimate $\hat{\beta}_1$ is 'good' as long as the four OLS Assumptions are satisfied

Roadmap

Introduction

The Big Ideas from STAT1008 and EMET2007

- Statistical Inference
- Our Goal: Learning about the Causal Effect
- Identifying Causal Effects with the Linear Regression Model
- Brief Review of OLS
- Estimator Goodness: Bias, Efficiency and Consistency

Endogeneity: When OLS fails

- Internal Validity

But what if some of the OLS Assumptions are not satisfied?

Then there is the risk that the OLS estimator $\hat{\beta}_1$ does not correctly estimate the average causal effect β_1

Statistical inference using OLS may be flawed in that case

In the worst case, we may make completely false inferences about the average causal effect

We are in a bad place: our big goal was to learn something about the average causal effect but we have now arrived at a point at which all our efforts could be in vain

What exactly do we mean when we say that the OLS estimator may not correctly estimate the average causal effect?

Definition (Internal Validity)

A statistical analysis is **internally valid** if statistical inferences about causal effects are valid for the population being studied.

The associated estimator is said to have **internal validity**.

For the next few weeks, we concern ourselves with situations in which the OLS estimator is not internally valid

In these cases, we cannot use OLS to learn about the average causal effect

But when exactly is the OLS estimator internally valid?
(and when is it not?)

Theorem

The OLS estimator is internally valid if it is unbiased and consistent.

But when is the OLS estimator unbiased and consistent?

Theorem

The OLS estimator is unbiased and consistent only under OLS Assumption 1.

Connecting the dots:

If OLS Assumption 1 is indeed ’true’, then the OLS estimator is internally valid. Otherwise it is not internally valid.

In the next few weeks we will become experts at understanding when OLS Assumption 1 is ’true’ and when it is not

We will learn what to do when it is not true

Problem Solving Exercises

1. Let $Y_i \sim$ i.i.d.(μ, σ^2). You know from EMET2007 that $\bar{Y}$ is an unbiased and consistent estimator for the population mean μ. Are the following estimators also unbiased or consistent for μ? Discuss!

   (i) $\hat{\mu}_2 := 42$ ('answer to everything' estimator)

   (ii) $\hat{\mu}_3 := \bar{Y} + 3/n$

   (iii) $\hat{\mu}_4 := (Y_1 + Y_2 + Y_3 + Y_4 + Y_5)/5$