Advanced Microeconometrics

course notes

Kristian Olesen Larsen*

University of Copenhagen

June 2, 2018

Abstract

These are my notes for the course Applied Econometric Policy Evaluation. I will attempt to cover each lecture in one or two pages, making reading the whole note-sheet feasible in a few hours.

These course notes are written as part of my personal studies for an exam, and should be taken only to reflect my understanding of the topic. I cannot guarantee that everything in these notes is correct, much less that the explanations provided here are better than those that others have already provided. To fully understand the topics covered i suggest following a similar course yourself.

Contents

I. Introduction Lecture

The main focus of the course is the estimation of *Âľausal* effects of policy interventions. In other words the central question to answer is how people react to exogenous events. In this context the ideal setup is a randomized experiment where the effect of a treatment D is measures w.r.t some outcome y. To achieve this a population should be randomly split in two, with one half getting the treatment $D = 1$ and the other not getting it $D = 0$. The size of interest is then the effect size $\delta = E[y_1] - E[y_0]$. In an OLS equation we will estimate δ as

$$y_i = \alpha + \beta D_i + u_i \tag{1}$$

where the critical assumption is that $E[D|u] = 0$. Throughout the course there are three core questions which should also be asked and answered,

1. What is the causal relation of interest?
2. What is the ideal experiment that would capture the causal effect?
3. What is the identification strategy?

As well as a fourth question which is good to consider

4. What is the mode of statistical inference?

The course lectures usually covers some theory, and shows an example of how this can be applied in real research.

II. Regression Analysis

In this lecture we study an alternative interpretation of the OLS estimator. This means all of the content is already known, but the presentation should hopefully broaden the intuitive understanding of what OLS really is. Note that we could call $E[Y_i|X_i]$ the *conditional expectation function* (CEF), as it models the conditional expectation of the outcome Y_i given X_i.

*kuol@econ.ku.dk. Please share these notes with as many people as you feel like.

Mathematically this expectation is defined as

$$E[Y_i|X_i] = \int t f_y(t|X_i = x)\mathrm{d}t \qquad (2)$$

where t is simply a value of Y_i and f_y is the density of Y_i conditional on $X_i = x$. Naturally this has a discrete counterpart where $E[Y_i|X_i] = \sum_t tP(Y_i = t|X_i = x)$. Notice also that the unconditional expectation is simply the unconditional expectation of this, so

$$E[Y_i] = E[E[Y_i|X_i]] \qquad (3)$$

Another way of rewriting the equation is to formulate it with an error term as (formally we invoke the law of iterated expectations here)

$$Y_i = E[Y_i|X_i] + \varepsilon_i \qquad (4)$$

Where ε is orthogonal to X_i by construction (this is the independence assumption). So intuitively, linear regression is simply a way of implementing that the CEF is linear in the dependent variable. Consequently if the CEF is truely linear, we will estimate it precisely with linear regression.

i. Saturated regression

Imagine a variable X which is discrete. We can then create dummy variables for *every* level of X, possibly constructing a dummy for any $x > n$ to get a finite number of dummies. Now running a regression with all of these regressors (either restricting the constant or a parameter) we get a *saturated regression equation* where there is one free parameter for each possible value of X. This regression will perfectly fit the CEF at every value of X as it estimates the mean of Y at each level of X. *In fact a regular linear regression on the conditional means produced by the saturated model gives exactly the same coefficients as OLS on the whole dataset.*

III. The counterfactual Setup

The counterfactual setup is a framework for understanding causality in regression frameworks. The setup assumes there exists a counterfactual outcome for each individual. Let Y_1 be the outcome with treatment, Y_0 the outcome without treatment and as before $D = \mathbb{1}_{(treatment)}$. The size of interest is naturally the difference between the outcomes with and without treatment on an individual level. More formally there are two sizes to be interested in, the average treatment effect (ATE) and the average treatment effect on the treated (ATT), defined as

$$\begin{aligned} ATE &= E[Y_1 - Y_0] \\ ATT &= E[Y_1 - Y_0|D = 1] \end{aligned} \qquad (5)$$

Note that the ATE is in many cases meaningless, as it includes the effect on those who were not treated, meaning it will not take into account selection effects. Two technical assumptions must be made for these to hold, namely that all variables are random draws, and that the treatment of one individual does not affect the probability of treatment for another individual.

We can write the actual effect as

$$\begin{aligned} Y &= (1 - D)Y_0 + DY_1 \\ &= Y_0 + D(Y_1 - Y_0) \end{aligned} \qquad (6)$$

Using this equation it is clear that assuming that D is independent of Y_0, Y_1 gives us that $E[Y|D = 1] = E[Y_1]$, and similarly $E[Y|D = 0] = E[Y_0]$. With these results we can then show that when outcomes are independent of treatment status $ATE = ATT$. Intuitively this is simply because when D is independent of Y's, treatment is essentially random in the population meaning the treated are affected in the same way by treatment as the untreated would be.