Applied Microeconometrics I

Lecture 12: Recap of some of the course topics

Stefano Lombardi

Aalto University

October 12, 2023

Lecture Slides

Recap of some of the course topics

Three parts:

1. Example 1: **effect of training on re-employment**
2. Example 2: **returns to schooling**
3. Some words on the **exam**

Please **stop me** along the way if *anything* is unclear!

Example 1: training program for the jobseekers

- **Setting**:
  - **Unemployed people** registered at Public Employment Service
  - They are entitled to unemployment insurance (UI) benefits
  - The law requires them to be *actively searching* for a new job (conditionality of benefits)
  - **Training programs** are a key tool to activate unemployed people, build human capital, and ultimately **find a job**
  - But they are costly!
- **Question:**
  - *Does the training program considered **causally** affect labor market outcomes?* (labor income, employment)
- **Notation**:
  - $D_i \in \{0, 1\}$: **participation** status in the training program for i
  - Y_i: labor income 1 year after the program ends
  - Question: does D_i causally affect Y_i?

Example 1: training programs for the jobseekers

Potential outcomes framework

- Suppose we focus on **one jobseeker** i, who has $D_i = 1$.
- **Question**: did the training cause an increase in Y_i?
  - is Y_i when $D_i = 1$ larger than Y_i when $D_i = 0$?
- Equivalently, "what if" i did not take the treatment?
  - *what would have happened* to Y_i in the case of $D_i = 0$?
  - what is the *counterfactual* Y_i when $D_i = 0$?
- With counterfactual/potential outcomes: is $Y_{1i} - Y_{0i} > 0$?
  - for this jobseeker, **only** Y_{1i} **is observed** $(Y_i = Y_{1i})$
  - we can still **think of** a "what if" scenario where Y_{0i} occurred
  - *fundamental problem*: $Y_{1i} - Y_{0i}$ cannot be observed/estimated
- The *statistical solution* to this is to move away from individual-level effects and reason in terms of **average effects**

Example 1: training programs for the jobseekers

Comparing the average outcomes in *observational* data

- That is, we focus on the average effect: $E[Y_{1i} - Y_{0i}|D_i = 1]$
- Again, for those with $D_i = 1$ we only observe Y_{1i}
- Can we use information on those with $D_i = 0$ to **build a valid counterfactual** for the jobseekers with $D_i = 1$?
- Natural starting point: **observed data**
  $E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$
  - these are simply the outcome averages in the two groups
  - both are readily observed!
- However, in general this is **not the ATET**:
  $E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = ATET +$ selection bias
  - selection bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$
  - what does even mean **selection bias**?

Example 1: training program for the jobseekers

Selection into/out of the training program

- Unless otherwise specified, even in the presence of UI rules:
  - the way the training is assigned/taken is a **black box**
  - i can **decide** whether to participate into the program
  - there is **selection** into (or out of) the program
- The training participation D_i is the result of *preferences, motivation, ability, ...*
  - of both the jobseeker i...
  - and of the caseworker assigned to i
- **Selection is part of life**: we decide to do things because we think that it is good for us
  - e.g., some jobseekers might think that D is a waste of time
- Participation incentives of those who have $D_i = 1$ vs. $D_i = 0$ are potentially *very different*: $E[Y_{0i}|D_i = 1] \neq E[Y_{0i}|D_i = 0]$

Example 1: training program for the jobseekers

Comparing the average outcomes in *experimental* data

- Suppose that we randomize the **participation** status D_i
  - *Full compliance*: everyone randomized to have $D_i = 1$ gets the training, and everyone randomized to have $D_i = 0$ don't
- What does **randomization of** D_i mean/imply?
  - randomization makes D_i independent of *all* characteristics
  - hence, *every* characteristic has the same distribution in the two groups (e.g., preferences, motivation, ability, ...)
  - potential outcomes distributions are equal in the two groups
  - hence, also $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$: **no selection bias**
  - *Randomization decides who gets the training*, not jobseekers (who would have chosen D_i according to their preferences)
- Now a simple comparison of averages in the two groups $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$ is the ATET

Example 1: training program for the jobseekers

Regression with *observational* data

- Suppose we regress Y_i on D_i with **observational data**.
- The model that relates earnings to training participation is:

$$Y_i = \alpha + \beta D_i + \delta A_i + \varepsilon_i$$

  ε_i is pure random noise (uncorrelated with A_i and D_i), while the unobserved ability A_i is correlated with D_i (and with Y_i)

- Since ability is left in the error term, $u_i = \delta A_i + \varepsilon_i$, and we can do OLS of the (short) regression:

$$Y_i = \alpha + \beta D_i + u_i$$

- **OVB**: since u_i is correlated with Y_i, the OLS estimator is *not* consistent for the true β, but will converge to $\beta + \delta \frac{Cov(A_i, D_i)}{Var(D_i)}$

Example 1: training program for the jobseekers

Regression with *observational* data

- Perhaps we have information on **rich set of pre-determined characteristics** X_i (past labor market history, SES, etc.)
- Then we could do OLS of:

$$Y_i = \alpha + \beta D_i + \gamma X_i + u_i$$

- **Conditional independence assumption**: the presumption is that *conditional on* X_i, D_i is as good as randomly assigned
- Then $E[Y_{0i}|D_i = 1, X_i] = E[Y_{0i}|D_i = 0, X_i]$ and $E[Y_i|D_i = 1, X_i] - E[Y_i|D_i = 0, X_i]$ is ATET condit. on X_i
- **Issues**: CIA unrealistic, measurement error, bad controls

Example 1: training program for the jobseekers

Regression with *experimental* data

- Suppose we regress Y_i on D_i with **experimental data**.
- The model looks identical as before:

$$Y_i = \alpha + \beta D_i + \delta A_i + \varepsilon_i$$

  but now A_i **is uncorrelated with** D_i thanks to randomization (it's still correlated with Y_i, we don't care)

- We estimate, as before:

$$Y_i = \alpha + \beta D_i + u_i$$

- u_i is uncorrelated with Y_i: OLS estimator is consistent for β
  - $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$
    $= (\alpha + \beta + E[u_i|D_i = 1]) - (\alpha + E[u_i|D_i = 0]) = \beta$

Example 1: training program for the jobseekers

Different types of experiments

- Remember that in this example we randomized **participation**
- Often we can only randomize **eligibility** (call it Z_i): participation D_i is (endogenously) chosen by the jobseeker
  - Some will participate $(D_i = 1)$ when offered $(Z_i = 1)$
  - But others with $Z_i = 1$ will not show up $(D_i = 0)$
  - Many ineligible jobseekers $(Z_i = 0)$ will not show up $(D_i = 0)$
  - But others will manage to convince their caseworker to make them participate to the program $(D_i = 1)$ even if $Z_i = 0$.
- In an **RCT with imperfect compliance** we cannot force people to follow randomization: D_i is taken based on *motivation, ability, etc.*
- A simple comparison of outcome averages (i.e., regressing Y_i on D_i) *will not work* (even if we are using experimental data!)

Example 1: training program for the jobseekers

RCT with imperfect compliance

- However, remember that we have randomized *eligibility* (Z_i)
- Then we can **instrument** training participation with training eligibility
- We saw that the **instrumental variables** estimation of the effect D_i on Y_i requires additional assumptions (other than randomization of Z_i)
- The average effect identified is also different from ATET:
  - LATE: *local* in the sense that the effect is for *compliers* only
  - people *at the margin* of taking the treatment; induced/pushed by Z_i to take D_i (but they would not if Z_i was 0)
  - $D_i = 1$ if $Z_i = 1$, and $D_i = 0$ if $Z_i = 0$

Some general points about IV

- **Why is LATE different from ATET?**
  - ATET is for *all treated units*
  - the ATET conditions on $D_i = 1$. Who are these people?
  - assuming monotonicity holds, *always-takers* and *compliers*
  - by exclusion restriction, LATE is not informative of the treated who are always-takers
  - LATE is only for the subset of *treated who are compliers*
- **First stage** only requires randomization of Z_i
  - causal effect of treatment *eligibility* Z_i on participation D_i
  - It measures the compliance rate
- **Reduced form** also only requires randomization of Z_i
  - causal effect of treatment *eligibility* Z_i on Y_i
  - it occurs via D_i and (possibly many!) other channels
- **IV estimator** is RF/FS
  - If we are willing to additionally assume that Z_i only affects Y_i via D_i, then we can adjust the RF by the compliance rate

Some general points about IV

Nomenclature can be confusing!

In general, the **classification is somewhat arbitrary**.
What I used in class is:

- *Instrumental variables* is the overall identification approach
- *IV estimator*: one instrument, one endogenous variable; it's equal to the ratio of Z_i coefficients from RF and FS
- *Wald estimator*: special/simplest IV estimator with binary Z_i
- *2SLS estimator* is the most general one:
  - it comprises the two above cases (numerically *identical*)
  - it also allows for multiple instruments and endogenous variables (e.g., more instruments than endogenous variables)
  - i) predict the treatment status D_i with Z_i (FS); ii) replace D_i in the model for Y_i with the FS fitted values and do OLS again

Example 2: Returns to schooling

- Whether education really increases earnings is one of the classic questions in economics
- *Education*: in years, completing a degree, degree type, etc.
- Subject on intensive study since Jacob Mincer's 1960's work
- Methods used: DD, IV, RDD

Example 2: Returns to schooling

Identification based on observables

- Early work on returns to schooling relied on identification based on observables
- Typical model would look like this (*Mincer equation*):

$$\log Y_i = \alpha + \rho S_i + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$$

  where $\log Y_i$ is the logarithm of annual earnings, S_i is years of education, and X is potential work experience

- How credible is the CIA assumption here?

Example 2: Returns to schooling

Identification based on observables

- Which factors are we omitting, when estimating returns to schooling relying on identification based on observables?
- Denote "ability" with A_i
- Suppose that real model of Y_i on S_i looks like this (ignore experience X for convenience):

$$\log Y_i = \alpha + \rho S_i + \gamma A_i + \epsilon_i$$

Example 2: Returns to schooling

Identification based on observables

- If A_i is not observable and is omitted from the regression, our estimates are biased:

$$\hat{\rho} = \rho + \gamma \frac{Cov(S, A)}{Var(S)}$$

- $\gamma \frac{Cov(S,A)}{Var(S)}$ is the "ability bias"
- What is the likely sign of this bias?

Example 2: Returns to schooling

Identification based on observables

- Many early studies tried to control for ability with proxies (IQ)
- Suppose $D_i = 1$ if i has graduated from university
- Then the observed earnings difference between university graduates and non-graduates conditional on IQ is:

$$E[Y_i|D_i = 1, \mathrm{IQ}] - E[Y_i|D_i = 0, \mathrm{IQ}] =$$
$$E[Y_{1i} - Y_{0i}|D_i = 1, \mathrm{IQ}] + \{E[Y_{0i}|D_i = 1, \mathrm{IQ}] - E[Y_{0i}|D_i = 0, \mathrm{IQ}]\}$$

- Under CIA, the second term (selection bias) disappears
- Two serious problems with this strategy:
  1. IQ may not capture all relevant abilities
  2. IQ may be a bad control

Example 2: Returns to schooling

Identification based on observables

- With CIA is vital to control for rich X vector But introducing controls can introduce several problems
- **Bad controls**: Control variables that are themselves outcomes caused by our causal variable of interest
- For example think of **controlling for white collar status**
- Even if college status is randomly assigned, controlling for occupation will induce selection bias

$$E[Y_{1i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]$$
$$=E[Y_{1i}-Y_{0i}|W_{1i}=1]+\{E[Y_{0i}|W_{1i}=1]-E[Y_{0i}|W_{0i}=1]\}$$

- Conditioning on occupation changes the composition of the treatment and control group: the white collar workers with college degree have a different Y_{0i} than those without college

Example 2: Returns to schooling

Difference-in-differences

- Suppose we controlled for some pre-determined covariates, but we are still concerned about **omitted family-level determinants** of income A_i (e.g. family inputs)
- Can we use **fixed effects/differences-in-differences** to estimate returns to education when we have data on twins?

$$\text{Twin 1: } Y_{1f} = \alpha + \rho S_{1f} + \gamma A_f + \epsilon_{1f}$$
$$\text{Twin 2: } Y_{2f} = \alpha + \rho S_{2f} + \gamma A_f + \epsilon_{2f}$$

f denotes the family, the outcome is log-earnings.

- If A_f is common to the pair of twins, then differencing yields:

$$Y_{1f} - Y_{2f} = \rho(S_{1f} - S_{2f}) + (\epsilon_{1f} - \epsilon_{2f})$$
$$\bar{Y}_f = \rho \bar{S}_f + \bar{\epsilon}_f$$

- If nothing left in the differenced error term is correlated with $S_{1f} - S_{2f}$, then estimating ρ with the differenced equation gives us the causal effect of schooling on earnings

Example 2: Returns to schooling

Difference-in-differences (Ashenfelter and Rouse, 1998)

OLS estimates in the population and in the twin sample

TABLE II
OLS ESTIMATES OF THE (MEAN) RETURN TO SCHOOLING USING THE CPS AND TWINS DATA

	CPS[a]	Identical twins
	OLS (1)	OLS (2)
Own education	0.085	0.110
	(0.0003)	(0.009)
Age	0.071	0.104
	(0.0004)	(0.010)
Age^2 (÷100)	−0.074	−0.106
	(0.0005)	(0.013)
Female	−0.253	−0.318
	(0.001)	(0.040)
White	0.087	−0.100
	(0.002)	(0.072)
Sample size	476,851	680
R^2	0.332	0.339

Standard errors are in parentheses. All regressions include a constant.

a. The Current Population Survey (CPS) sample is drawn from the 1991–1993 Outgoing Rotation Group files. The sample includes workers age 18–65 with an hourly wage greater than $1 per hour in 1993 dollars; the regression is weighted using the earnings weight. We converted the 1992 and 1993 education categories into a continuous measure according to the categorization suggested by Park [1994].

Example 2: Returns to schooling

Difference-in-differences (Ashenfelter and Rouse, 1998)

GLS (OLS) and first difference estimates

	GLS (1)	GLS (2)	3SLS (3)	First-difference (4)	First-difference by IV (5)
Own education	0.102	0.066	0.091	0.070	0.088
	(0.010)	(0.018)	(0.024)	(0.019)	(0.025)
Avg. education $[(S_1 + S_2)/2]$		0.051	0.033		
		(0.022)	(0.028)		
Age	0.104	0.103	0.103		
	(0.013)	(0.013)	(0.013)		
Age^2 ($\div$100)	−0.107	−0.104	−0.104		
	(0.015)	(0.015)	(0.015)		
Female	−0.315	−0.309	−0.306		
	(0.049)	(0.049)	(0.049)		
White	−0.106	−0.105	−0.101		
	(0.090)	(0.091)	(0.091)		
Covered by a union					
Married					
Tenure (years)					
Sample size	680	680	680	340	340
R^2	0.262	0.264	0.267	0.039	

Example 2: Returns to schooling

Quarter of birth instrument

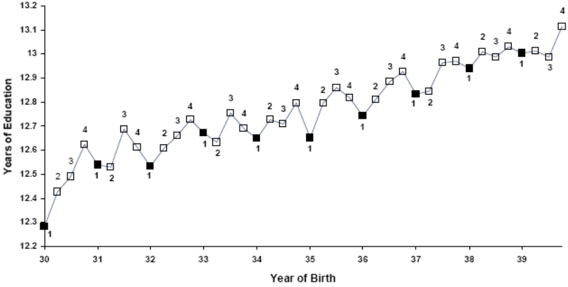
- In Lecture 7, we saw how one could use instrumental variables to estimate the returns to schooling
- Angrist and Krueger: Quarter of birth as an instrument for schooling
- Students enter schooling in the September of the calendar year in which they turn 6
- And compulsory school law requires them to remain in school until they become 16
- Hence people *born late* in the year are more likely to *stay at school longer*

Example 2: Returns to schooling

Quarter of birth instrument

First stage:

A. Average Education by Quarter of Birth (first stage)

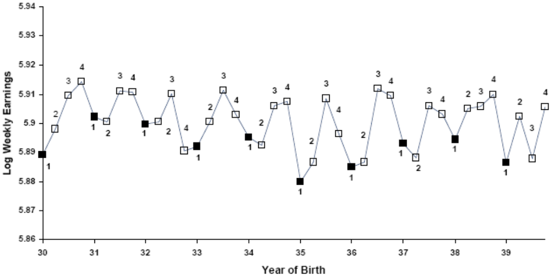


Example 2: Returns to schooling

Quarter of birth instrument

Reduced form for earnings

B. Average Weekly Wage by Quarter of Birth (reduced form)



Example 2: Returns to schooling

Quarter of birth instrument

Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

Notes: Adapted from a re-analysis of Angrist and Krueger (1991) by Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930-39 birth cohorts in the 1980 Census 5 percent file. The sample size is 329,509.

Example 2: Returns to schooling

Quarter of birth instrument

- How is the effect local here? Who are the compliers?
- **Monotonicity**?
  - Barua and Lang (2010): parents delay school entrance of their child if they think child is not mature (redshirting)
  - children born late in the year are more likely to be redshirted
  - some will spend *less* time in school than if born in January
- **Other examples where monotonicity might not hold?** What about the sex-mix instrument?
  - parents' sex bias might lead to monotonicity violation
  - Dahl and Moretti (2008): US fathers have a preference for boys
  - couples who prefer two boys/two girls are defiers
- Other example: monetary incentives in encouragement design
  - Frey and Jegen (2001): for some, monetary incentives part of the design crowd-out (*reduce*) intrinsic motivation

Example 2: Returns to schooling

RDD: Do degrees matter? Clark and Martorelli (2014)

- Finally, we go over an **RDD example** on the effects of schooling
- What is the effect of the high school diploma *as such*? (literally, the "piece of paper" from a given institution)
- Sheepskin effect: The effect of diploma as a piece of paper, ceteris paribus

Example 2: Returns to schooling

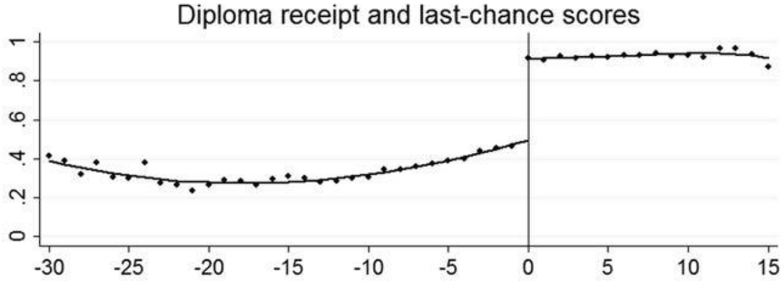
RDD: Do degrees matter? Clark and Martorelli (2014)

- In Texas, getting a high school diploma is conditional on passing an exit exam
- The probability of getting the diploma jumps discontinuously at the passing of exit exam threshold; can use this to identify the effect of diplomas on earnings
- No reasons to expect other discontinuities at the threshold, so that getting a diploma is as good as randomly assigned near c
- Results:
    - The probability of getting the diploma increases by 50 percentage points at the passing threshold (FS)
    - Yet, the earnings don't change discontinuously (RF)
    - Therefore, no evidence of sheepskin effects

Example 2: Returns to schooling

RDD: Do degrees matter? Clark and Martorelli (2014)

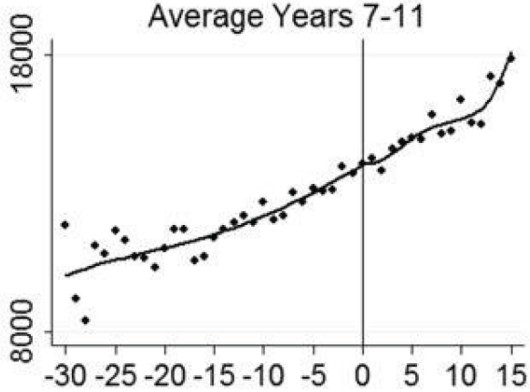
First stage



Example 2: Returns to schooling

RDD: Do degrees matter? Clark and Martorelli (2014)

Reduced form



Concluding remarks

- **For each method**, ask yourself:
  - what is randomized? (or conditionally randomized)
  - is D chosen or assigned randomly? Do people have a say in terms of taking/not the treatment?
  - what are the identifying assumptions?
  - what can I test with the data and how? Is the test “definitive”?

Concluding remarks

- **For the exam**:
  - do not memorize proofs (unless this helps you)
  - the papers we covered are part of the materials in the sense that you need to understand what has been done and why
    - I will not ask questions like "what is Card doing in this paper"?
    - but you need to understand what the papers we covered did!
  - revise the PS as well: exam questions will be similar
  - no software questions
  - no "write this equation with the proper notation", but I can ask to describe in words what one or more equations are in the context of the example (What is the dependent variable? What is on the right-hand side? Interpretation?)
  - short answers to each sub-question are fine!
  - read a full question before starting answering to the sub-points
  - If you are stuck, move on and go back to it later
  - if you can, try to write even partial answers: *I will do all that I can to give points for effort*!