Concluding Remarks

Jaakko Meriläinen

5304 Econometrics @ Stockholm School of Economics

Review

- We have had a full plate of lectures, seminar groups, presentations, and assignments
- Let us conclude by taking a quick overview of the course
  - Details about the exam
  - Course aims and how the various parts of the course fit together (in retrospect!)
  - A quick recap of IV, RDD, and DID
  - Fill in the survey!

Grading

- Grades for this course will be aggregated across three different sub-components

1. **Homework assignments** (20%)

2. **Presentation in a sub-group** (10%)
   - This includes both referee report and the presentation
   - Grades will be communicated alongside the final exam

3. **Final exam** (70%)
   - More on this to follow

Revisiting the Objectives of This Course

- The goals for this course are for students to:
  - **Understand current empirical research methods**
    - To understand commonly-used econometric tools
    - To be able to read primary economic research directly
    - To be able to critique statistical methods employed in papers
  - **Be able to design empirical analyses**
    - How would you go about designing econometric analyses on a given research question?
  - **Be able to execute empirical research**
    - How to use statistical software to analyze data
    - How to interpret and present results

- Thus, this was primarily an applied econometrics course (with requisite bits of theory)

The Purpose of Different Course Portions

- **Lectures:**
  - To present and explain methods: objectives, content, issues
  - To tie the different elements of the course together
- **Homework and seminars:**
  - How to apply methods in the lecture
  - Stata tasks for data work
  - How you should think of modeling, interpreting estimates, etc.
- **Presentations:**
  - How to read empirical research critically
  - To make methods less abstract
  - But also as introduction to what modern empirical research looks like (methods and also a range of topics)
  - To focus on one set of methods really closely

Final (Take-Home) Exam

- **When:** The exam will appear on Canvas on December 20, 2023, at 9:00 CET
- **Deadline:** You must submit your solutions by December 20, 2022, 13:00 CET—if you have extra time, please email me asap!
- **All material** covered in the course are up for inclusion
  - Focus on applying tools covered in lectures/readings
  - Coding on Stata/R will not be required for the exam
- **What you are allowed:**
  - You **can** consult lecture slides,notes, readings (text books and assigned papers), problem sets
  - You **may not** consult any individual, online or offline
    - I take cheating very seriously
    - Any suspicion of such methods will lead to immediate cancellation of the grades
- I will be available online at various times during the exam

Course Plans Revisited

- This course builds upon undergraduate training in econometrics
- We covered some of this material again but faster than in a Bachelor's course
- We will be extending this in several dimensions
  - A deeper look at topics you had probably seen before (OLS, IV, panel data, statistical inference)
  - Some new topics (RDD, potential outcomes/treatment effects, heterogeneous effects, experiment design, synthetic controls)
  - A lot more direct engagement with economic research and data (using Stata/R to work with real datasets, many paper discussions)
- Our focus was primarily on measuring causal effects

A Quick Recap

- In Lecture 4, we studied the potential for bias in OLS estimates
- The ZCM assumption, $E(u|x) = 0$, could break down for different reasons
  - Omitted variables bias
  - Simultaneity
  - Measurement error
- A convenient RCT is not always (or usually) available
- What to do then? This is what most of the course was about!
- Let us recap the three main methods discussed in this course: IV, RDD, and DID
- NB. The course also covered other important topics—note that everything is up for inclusion in the exam!

A Quick Recap: IV

- Given an outcome variable y_i, an endogenous variable x_i, an instrument z_i and a vector of controls A_i:
  - **First-stage regression:**

$$x_i = b_0 + b_1 z_i + A_i^{'} \gamma + e$$

  - **Second-stage (structural) regression:**

$$y_i = \beta_0 + \beta_1 \hat{x}_i + A_i^{'} \phi + u$$

  - **"Reduced-form" regression:**

$$y_i = a_0 + a_1 z_i + A_i^{'} \lambda + \xi$$

A Quick Recap: IV

- The two conditions for IV are conceptually very distinct

- **Relevance:** $Cov(z, x_1) \neq 0$
  - This is testable
  - Regressing x_1 on z should give a coefficient different from 0 on z
  - We will come back to testing for relevance later...

- **Validity:** $Cov(z, \eta) = 0$
  - This requires z to not be a direct determinant of y
  - z must be uncorrelated with any other determinants of y except x_1
  - This restriction is called the **exclusion restriction** for the instrument
  - This is not directly testable, needs justification from theory, knowledge of institutions etc.—oftentimes, a lot of story telling!
  - Even with randomization, exclusion restriction might not hold

A Quick Recap: IV

- Let us stay with the case of a binary instrument (Z_i) and a binary endogenous variable (D_i)
- Imagine X is finishing college, Z is getting a (randomized) scholarship
- We can think of four types of compliance units:
  - **Always-takers:** $D = 1$, whether $Z = 1$ or $Z = 0$
    - I always go to college, whether or not I get a scholarship
  - **Never-takers:** $D = 0$, whether $Z = 1$ or $Z = 0$
    - I never go to college, whether or not I get a scholarship
  - **Compliers:** $D = 1$ when $Z = 1$; $D = 0$ when $Z = 0$
    - Getting a scholarship makes me go to college
  - **Defiers:** $D = 0$ when $Z = 1$; $D = 1$ when $Z = 0$
    - I would have gone to college otherwise; but now that I have a scholarship, I will not go!

A Quick Recap: IV

1. **Independence:** $\{Y_i(D_{1i}, 1), Y_i(D_{0i}, 0), D_{1i}, D_{0i}\} \perp Z_i$
   - This means that Z_i is as good as randomly assigned
2. **Exclusion restriction:** $Y_i(d, 0) = Y_i(d, 1) \equiv Y_{di}$ *for* $d = 0, 1$
   - The instrument Z_i only affects Y_i through D_i
3. **First stage:** $E[D_{1i} - D_{0i}] \neq 0$
   - This is the equivalent of what we have called the relevance condition
   - The instrument has some explanatory power for D_i
4. **Monotonicity:** $D_{1i} - D_{0i} \geq 0$ *for all* i or vice versa
   - This means that the instrument affects the probability of $D_i = 1$ in the same direction for all individuals
   - This is also called the "no-defiers" assumption

A Quick Recap: IV

- Under the four assumptions on the previous slide, it can be shown that the IV estimator

$$\beta = \frac{E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

  identifies the effect of treatment on those individuals whose treatment status has been changed by the instrument (i.e. the compliers)

- This parameter is called the **Local Average Treatment Effect**

- In general, LATE is not ATE, ATT, or ATU (see slides from Lecture 8 for examples)

A Quick Recap: IV

- It is also possible to have multiple endogenous variables and instruments—two additional conditions must be satisfied
- **Order condition:** We need as many instruments excluded from the structural equation as endogenous variables
- **Rank condition:** There is a first-stage relationship for every variable in the structural equation
- IV identifies a weighted average of instrument-specific LATEs

A Quick Recap: RDD

- The basic idea is simple
- There is some underlying variable (X_i—known as a running variable) that determines whether you get access to a program based on a threshold
- On crossing the threshold, the probability of treatment jumps discontinuously
- If potential outcomes vary continuously over the threshold, we can get consistent estimates
  - Formally, $E(Y_i^0|c)$ and $E(Y_i^1|c)$ should evolve smoothly at the cutoff
  - Absent the treatment, in other words, the expected potential outcomes would not have jumped
  - At the cutoff, it is as-good-as-random whether you end up to the left-hand or the right-hand side of the cutoff—i.e., whether a unit is treated or not

A Quick Recap: RDD

- The clearest application of this approach is in cases where...
  - ...nobody gets the treatment up to some value c
  - ...and everyone gets it after

- The treatment is then defined as

$$W_i = 1\{X_i \geq c\}$$

- We can then look at the jump in the conditional expectation of the outcome (Y_i) given the covariate (X_i) at the threshold c

- This gives us the ATE at the cutoff (assuming treatment is as-good-as-randomly assigned at the cutoff)

A Quick Recap: RDD

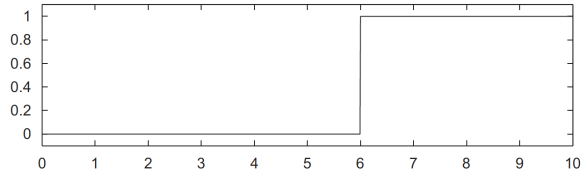

Fig. 1. Assignment probabilities (SRD).

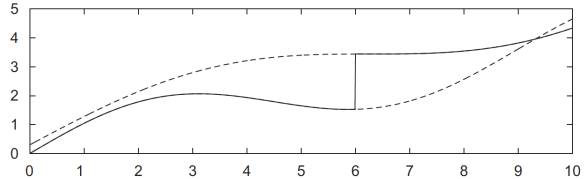

Fig. 2. Potential and observed outcome regression functions.

A Quick Recap: RDD

- We can express a sharp regression discontinuity design as

$$Y_i = \alpha + \beta 1(X_i \geq c) + f(X_i) + \epsilon_i$$

- Here α is a constant, $1(X_i \geq c)$ is an indicator for the running variable or assignment variable crossing the cutoff c, $f(X_i)$ is a control polynomial (typically linear, estimated separately for each side of the cutoff), and ϵ_i is the error term

A Quick Recap: RDD

- To estimate a causal effect, we do not require the probability of treatment to go from zero to one at the discontinuity
- We just need it to be discontinuous ("jump")
- As long as that is the case, we can get consistent estimates from the Wald Estimator:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E(Y_i | X_i = x) - \lim_{x \uparrow c} E(Y_i | X_i = x)}{\lim_{x \downarrow c} E(W_i | X_i = x) - \lim_{x \uparrow c} E(W_i | X_i = x)}$$

- It is the reduced-form (ITT) estimate divided by the first-stage—**fuzzy RDD** is IV! (LATE at the cutoff)

A Quick Recap: RDD

- We can express a fuzzy regression discontinuity design as

$$Y_i = \gamma + \delta \hat{D}_i + g(X_i) + \mu_i$$

- Here γ is a constant, $\hat{D}_i$ comes from the first stage, $g(X_i)$ is a control polynomial (typically estimated separately on each side of the cutoff, and μ_i is the error term

- The first-stage regression is

$$D_i = \lambda + \kappa 1(X_i \geq c) + h(X_i) + \xi_i$$

A Quick Recap: RDD

- **Is there a discontinuity?**
  - How does the RDD plot look like?
  - Can the underlying function be reasonably approximated by a linear or quadratic polynomial?
- **Is the RD valid?**
  - Manipulation in the running variable?
  - Heaping in the running variable?
  - Discontinuities in covariates?
  - Discontinuities in placebo regressions?
- **How sensitive are results to bandwidth choice?**
  - How far from the threshold can you go?

A Quick Recap: Difference-in-Differences

- Imagine, for instance, wanting to study the effect of minimum wages on employment
  - Two states, A and B
  - Two time periods, 1 and 2
  - State A introduces a minimum wage in period 2

$$Y_{igt} = \alpha_1 + \alpha_2 POST_t + \beta_1 Treat_g + \beta_2 (Treat_g \times POST_t) + \epsilon_{igt}$$

- Ignoring covariates, we have the following outcomes:

	State A (T)	State B (C)
Period 1 (pre)	$\alpha_1 + \beta_1$	α_1
Period 2 (post)	$\alpha_1 + \alpha_2 + \beta_1 + \beta_2$	$\alpha_1 + \alpha_2$

A Quick Recap: Difference-in-Differences

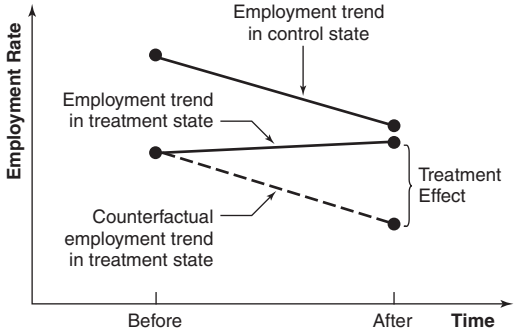

Figure 5.2.1 Causal effects in the DD model.

A Quick Recap: Difference-in-Differences

- The causal interpretation of the DiD estimator requires that (potentially conditional on X) the treatment and control groups **would have had similar changes in outcomes in the absence of treatment**
- This is the **"parallel trends" assumption**
- What this implies:
  - The treatment and control groups **can** have different levels of Y...
  - ...but not different growth rates in Y
- If the parallel trends assumption holds, DiD gives us ATT
  - If it does not hold, there will be bias from non-parallel trends
  - See Cunningham's book for an explicit expression of this bias

A Quick Recap: Difference-in-Differences

- The parallel trends assumption cannot be directly tested, but you should assess pre-treatment trends
- With just three periods of data, you could look at the change between $t = -1$ and $t = 0$ as the dependent variable
- But you might also have more data—then you can easily extend this to multiple time periods

$$Y_{it} = \alpha + \beta T_i + \sum_{s \neq S}^{T} (\gamma_s 1[t = s]_t + \delta_s 1[t = s]_t T_i) + \epsilon_{it}$$

- Another option: try controlling for group-specific time trends
- You should also visualize the data

A Quick Recap: Difference-in-Differences

- Generalized difference-in-differences controls for unit-specific fixed effects and time fixed effects (also known as a two-way fixed effects specification)

$$y_{st} = \lambda_s + \lambda_t + \beta \mathit{Treatment}_{st} + \epsilon_{st}$$

  where λ_s are the unit fixed effects, λ_t are the year fixed effects, and $\mathit{Treatment}_{st}$ is an indicator telling whether unit s was treated in year t (NB. treatment timing could vary across units!)

- Parallel trends assumption? Control for unit-specific trends or estimate a dynamic specification

Questions I have for you

- Course pace
- Split across theory and applications
- How useful were the seminar groups, assignments, paper discussions?
- Course workload
- Anything you would change?

Final Words

- The econometric methods we have seen through the course are tools
  - For good empirical economics, you need to find good questions, good data
  - The tools are very useful and applicable to many, many issues—as you have seen through the course
- Causal analyses are hard
  - There is often bias, which sometimes we are fortunate in being able to deal with
  - There are always issues of interpretation
    - How useful is the parameter we estimate?
    - How generalizable is it?
- But it can be very exciting when you manage to tackle a hard problem
  - So I hope that is what you will try for your theses!

Postscript from Angrist and Pischke (2009)

If applied econometrics was easy, theorists would do it. But it's not as hard as the dense pages of Econometrica might lead you to believe. Carefully applied to coherent causal questions, regression and 2SLS almost always make sense. Your standard errors probably won't be quite right, but they rarely are. Avoid embarrassment by being your own best skeptic - **and, especially, don't panic***!*