EDPE 6023: Causal Methods for Education Policy Research*

TEACHERS COLLEGE • COLUMBIA UNIVERSITY • FALL 2014
TUESDAYS 3-5 PM

Prof. Judith Scott-Clayton
Office hours: Thursdays 2:30-4:30 (please make an appointment); other times by appointment
Office location: Thorndike 428
(212) 678-3478 – scott-clayton@tc.columbia.edu

Course overview
This doctoral course covers the design, implementation and interpretation of econometric methods used for evaluating causal relationships in education research. We will read and discuss applied methodological texts as well as journal articles using advanced causal methods. We will cover randomized experiments, natural experiments, differences-in-differences, instrumental variables, regression discontinuity, and propensity score matching. Goals of the course are for students to understand the conceptual underpinnings of each type of study design; to be able to critically evaluate particular studies utilizing each approach; to gain first-hand experience in formulating causal questions and implementing a causal method; and to develop skills in communicating research designs and findings (in both written and presentation form). Students will be expected to complete all readings and prepare answers to discussion questions in advance of each class. Students will work in teams to replicate and extend a paper using these causal methods and will present findings to the class. The course is designed for second-year PhD students in the Economics and Education program; other doctoral students and advanced masters candidates with appropriate preparation are also welcome, space permitting.

Prerequisites
- Graduate-level statistics (at a minimum, multiple regression analysis, familiarity with concepts of statistical bias and precision)
- Microeconomics (e.g. exposure to concepts of consumer theory, producer theory, equilibrium analysis, market failure, welfare analysis, choice under uncertainty)
- Students with any questions about their preparation after the first day of class are encouraged to contact the professor for further guidance.

Course objectives
- To understand the conceptual underpinnings of current methods for causal inference
- To be able to read and critically evaluate papers that utilize these methods
- To gain first-hand experience formulating causal questions and implementing causal methods
- To develop skills in communicating research designs and findings in both written and oral form
- To encourage and facilitate collaborative learning and teaching between students

Required texts
- Articles and working papers available online (I will provide links, but please search for yourself if link is broken)
- Methods Matter by Richard Murnane and John Willett
- Mostly Harmless Econometrics by Joshua Angrist and Jorn-Steffen Pischke

* This course is adapted from similar courses taught by Richard Murnane and John Willett at Harvard (HGSE S-290), and Susan Dynarski at Michigan (EDU 820). I thank these professors for generously sharing their course materials.
Statistical software

- In order to complete course requirements, you must have access to and be able to use statistical analysis software. Stata is the tool most commonly used by economists (it includes useful commands not readily available in other packages) and can be purchased via http://www.columbia.edu/acis/software/licenses/stata/. Note that there are several purchasing options, ranging from about $50 and up (though the cheapest option will not allow you to work with large datasets). If you plan to use Stata for your dissertation you probably should purchase a perpetual license of Stata/IC. If you will use it only for this course you can purchase a 6-month or one-year license at lower cost; for very large datasets (Census, etc.) you may need to access Stata using the computer lab at SIPA.
  - For Stata FAQs and extensive tutorials, see http://www.ats.ucla.edu/stat/stata/. Check out the movies & notes here: http://www.ats.ucla.edu/stat/stata/notes/default.htm

Course requirements

- Pop quizzes and participation: 15%. Students are expected to attend all class sessions, to prepare answers to discussion questions (which are not turned in), and to be engaged during lecture and discussion. We will have 3-4 short pop quizzes throughout the semester. At the end of the term, the lowest quiz grade will be dropped and replaced with the general class participation score (assuming the participation score is higher).
- Referee report (2-3 pages): 15%. Read and critique an empirical paper under consideration for the replication paper. This is an individual assignment, due Sept. 23.
- Causal inference in the media (1 page, 2-3 minutes): 15%. Each week, 2-3 students will identify an example of causal inference in the media, submit a 1-page discussion brief, and present the case to the class. Sign up on first day of class.
- Team replication paper (20-25 pages) and related progress reports: total of 45%.
  - The major project of the class is to replicate and extend the analysis from a previous article. I will provide a list of articles for which the underlying data is readily available. You can replicate an article that is not on this list, but the choice must be confirmed with me in advance. Possible types of extensions include changing the identification strategy, testing sensitivity to alternative sample restrictions or covariates, subgroup analyses, examining additional outcomes, bounding exercises, and so on. **You must collaborate on this project in teams of at least two and no more than three students.** You will be graded on each component of the process:
    - Oct. 15 – Data memo (10%)
    - Nov. 5 – Project proposal and initial results (10%)
    - Nov. 26 - First draft (ungraded but will be taken into account in final draft grade)
    - Dec. 16 - Final draft (25%)
- Team presentation (15-20 minutes): 10%. Near the end of the course, you will present your research-in-progress. Co-authors should split the presentation so each gets a chance to speak.

Students with disabilities

The College will make reasonable accommodations for persons with documented disabilities. Students are encouraged to contact the office of Access and Services for Individuals with Disabilities for information about registration (166 Thorndike Hall). Services are available only to students who are registered and submit appropriate documentation. I am happy to discuss specific needs with you as well.

Incompletes

The grade of Incomplete is to be assigned only when the course attendance requirement has been met but, for reasons satisfactory to the instructor, the granting of a final grade has been postponed because certain course assignments are outstanding. If the outstanding assignments are completed within one calendar year from the date of the close of term in which the grade of Incomplete was received and a final
grade submitted, the final grade will be recorded on the permanent transcript, replacing the grade of Incomplete, with a transcript notation indicating the date that the grade of Incomplete was replaced by a final grade. **If the outstanding work is not completed within one calendar year from the date of the close of term in which the grade of Incomplete was received, the grade will remain as a permanent Incomplete on the transcript.** In such instances, if the course is a required course or part of an approved program of study, students will be required to re-enroll in the course including repayment of all tuition and fee charges for the new registration and satisfactorily complete all course requirements. If the required course is not offered in subsequent terms, the student should speak with the faculty advisor or Program Coordinator about their options for fulfilling the degree requirement. Doctoral students with six or more credits with grades of Incomplete included on their program of study will not be allowed to sit for the certification exam.

**E-mail**

Teachers College students have the responsibility for activating the Columbia University Network ID (UNI), which includes a free Columbia email account. As official communications from the College – e.g., information on graduation, announcements of closing due to severe storm, flu epidemic, transportation disruption, etc. -- will be sent to the student’s Columbia email account, students are responsible for either reading email there, or, for utilizing the mail forwarding option to forward mail from their Columbia account to an email address which they will monitor.

**Religious holidays**

It is the policy of Teachers College to respect its members’ observance of their major religious holidays. Students should notify instructors at the beginning of the semester about their wishes to observe holidays on days when class sessions are scheduled. Where academic scheduling conflicts prove unavoidable, no student will be penalized for absence due to religious reasons, and alternative means will be sought for satisfying the academic requirements involved. If a suitable arrangement cannot be worked out between the student and the instructor, students and instructors should consult the appropriate department chair or director. If an additional appeal is needed, it may be taken to the Provost.

**Academic integrity**

Students who intentionally submit work either not their own or without clear attribution to the original source, fabricate data or other information, engage in cheating, or misrepresentation of academic records may be subject to serious penalties. Sanctions may include course failure and dismissal from the college for violation of the TC principles of academic and professional integrity fundamental to the purpose of the College.

**Grades**

I have received feedback from students that I am a “hard grader.” I hope this is not a discouragement from taking the course but rather a signal of my high expectations for both my students and my own teaching. Grades are (loosely) on a curve though I do not set quotas/limits on specific grades. I follow the TC Catalogue in setting my grading standards:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A+</td>
<td>Rare performance. Reserved for highly exceptional, rare achievement.</td>
</tr>
<tr>
<td>A</td>
<td>Excellent. Outstanding achievement.</td>
</tr>
<tr>
<td>A-</td>
<td>Excellent work, but not quite outstanding.</td>
</tr>
<tr>
<td>B+</td>
<td>Very good. Solid achievement expected of most graduate students.</td>
</tr>
<tr>
<td>B</td>
<td>Good. Acceptable achievement.</td>
</tr>
<tr>
<td>B-</td>
<td>Acceptable achievement, but below what is generally expected of graduate students</td>
</tr>
<tr>
<td>C+</td>
<td>Fair achievement, above minimally acceptable level.</td>
</tr>
<tr>
<td>C</td>
<td>Fair achievement, but only minimally acceptable.</td>
</tr>
<tr>
<td>C-</td>
<td>Very low performance. The records of students receiving such grades are subject to review.</td>
</tr>
</tbody>
</table>
# Schedule of topics and deadlines

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topics</th>
<th>Deadlines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sept. 9</td>
<td>Intro to causal inference</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Sept. 16</td>
<td>Rubin Causal Model</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Sept. 23</td>
<td>RCT Basics (1)</td>
<td>Referee Report Due, Teams formed</td>
</tr>
<tr>
<td>4</td>
<td>Sept. 30</td>
<td>RCT Complications (2)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Oct. 7</td>
<td>Power and clustered RCTs (3)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Oct. 14</td>
<td>Instrumental variables</td>
<td>Team: data memo due Oct 15</td>
</tr>
<tr>
<td>7</td>
<td>Oct. 21</td>
<td>Lotteries as instruments</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Oct. 28</td>
<td>Regression discontinuity</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Nov. 4</td>
<td>Differences in differences</td>
<td>Team: Proposal due Nov 5</td>
</tr>
<tr>
<td>10</td>
<td>Nov. 11*</td>
<td>Self-study</td>
<td>Self-study write up due Nov. 14</td>
</tr>
<tr>
<td>11</td>
<td>Nov. 18</td>
<td>Fixed Effects</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Nov. 25</td>
<td>Propensity scores</td>
<td>Team: 1st draft due Nov 26</td>
</tr>
<tr>
<td></td>
<td>Nov 27</td>
<td>Thanksgiving holiday</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Dec. 2</td>
<td>Presentations (3-4)</td>
<td>Team presentations</td>
</tr>
<tr>
<td>14</td>
<td>Dec. 9**</td>
<td>Presentations (6-8)</td>
<td>Team presentations and course wrap-up</td>
</tr>
<tr>
<td></td>
<td>3-7 PM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Dec. 16**</td>
<td>No class</td>
<td>Final Papers Due</td>
</tr>
</tbody>
</table>

*I will be traveling the week of Nov. 10-14 so there will not be a regular class that week and I will not be holding office hours. The course assistant will be available as usual, and I will have limited availability via e-mail during that week.

**We typically schedule a “double-class” session on the second-to-last week of class so that in the final week, everyone can focus on meeting with their teams to finish the final paper. If many students in the class have conflicts, we can explore alternative scheduling options.
Week 1 (September 9): Causal Inference and the “Credibility Revolution.”
Motivation for the course, requirements and expectations.

Readings:

Additional resources:
- Cook, Thomas (2001). “Sciencephobia.” Education Next (Fall), pp. 63-68. [Accessible article on resistance to experimentation in education.]

Questions:
1. Consider a causal question that has arisen in your prior work/school/life experience. How was the question approached, and with what results? Is there an “ideal experiment” that could test this causal relationship?
2. What does Angrist view as the most important challenge to the usefulness of randomized trials? Explain the concern and his example. What does he conclude, and do you agree?
3. True or false: Angrist and Pischke believe that an observational study can never be as credible as a randomized trial. Why or why not?

Week 2 (September 16): Causal Questions, Causal Problems
Asking causal questions; Rubin’s Causal Model, the Fundamental Problem of Causal Inference, Linking regression and causality.

Readings:
- Angrist & Pischke (2009). Mostly Harmless Econometrics, Ch 1-2, 3-3.1.2, 3.1.4-3.2.2, pp. 3-40, 48-64.

Additional resources:
- Ani Katchova (2013-14) “Linear Regression,” Econometrics Academy on YouTube. [30-minute review of linear regression and 15-minute example, also 20-minute example with Stata.]

Questions:
1. Explain the notation used in Rubin’s model of causal inference (described in the Holland article). Specifically, what do \( Y_d(u) \) and \( Y_c(u) \) refer to?
2. What is the Fundamental Problem of Causal Inference, and what are the two types of solutions Holland describes? Give an example of each (other than those in the article). How do randomized trials address the Fundamental Problem of Causal Inference?

3. Explain Angrist & Pischke’s four “FAQs.” The Murnane & Willett reading suggests there may be another important question to ask when beginning a research project. What is it?

4. What is the CEF and how does OLS regression relate to it? What is the difference between the three theorems described on pages 37-38 of MHE?

Week 3 (September 23): Randomized Experiments (1) - Causal Inference in RCTs
***REFEREE REPORT DUE VIA MOODLE BY 3:00 PM TUESDAY.***

The basics: how RCTs work for causal inference; challenges in designing and implementing randomized experiments; role of covariates in experiments. Binary outcomes, conditional-on-positive outcomes.

Readings:
- Angrist & Pischke (2009). Mostly Harmless Econometrics, Ch 3.2.3 and 3.4.2, pp. 64-68, 94-102

Additional resources:

Questions:
1. Consider Table 4.1 (p. 49) in Methods Matter. Comparing the results from Strategy 2 and Strategy 3, what can you say about the difference in PRE_ACH (prior achievement) between the treatment and control groups?
2. Why is it useful to control for covariates even in the context of a randomized experiment? What kinds of variables should not be included as covariates? If including covariates significantly changes an impact estimate from a randomized experiment, is this cause for concern or not?
3. What tradeoffs do researchers make when determining the point at which randomization should occur? Describe with an example.
4. What is A&P’s argument for using OLS to analyze binary outcomes? Do you find it convincing or not?

Week 4 (September 30): Randomized Experiments (2) – Classic Example & Complications

Non-experimental estimates in experimental studies; missing data; multiple testing; Krueger’s classic class-size experiment; mechanism experiments.

Readings:

Additional resources:

Questions:
1. How can missing data undermine causal inference even in the context of an experiment?
2. Is it possible to experimentally examine the impact of Career Academies on participants’ wages? Why or why not?
3. How can we verify that randomization “worked” in a given RCT? What concerns might one have about the initial randomization in the TN Star experiment, and what evidence is provided to allay these concerns?
4. What is the difference between a mechanism experiment and a policy evaluation?

Week 5 (October 7): Randomized Experiments (3) – Significance, Power and Interpretation.
Statistical power and sample size; clustering; ethical considerations in RCTs

Readings:
• Murnane, Richard, and John Willett (2010). Methods Matter, Chapter 6-7, pp. 82-134.

Additional resources:
• Abdul Latif Jameel Poverty Action Lab, Making Schools Work for Marginalized Children: Evidence from an Inexpensive and Effective Program in India. (very short and accessible, written for practitioners)

Questions:
1. What is “effect size”? Does Cohen’s definition (described in Murnane and Willett) of small, medium, and large effects make sense? Think of at least one other way that coefficients might be sensibly scaled.
2. What is the difference between statistical significance and statistical power? How are they related? What is the difference between finding that an intervention has no statistically significant impact, and finding a “precisely-estimated zero” impact?
3. At what level are standard errors clustered in the Burde/Linden experiment, and why?
4. Would you consider the Burde & Linden experiment this week to be a mechanism experiment? Why or why not?

Week 6 (October 14): Instrumental Variables
***DATA MEMO DUE BY 5:00PM WEDNESDAY OCTOBER 15***

Readings:
• Angrist & Pischke (2009). Mostly Harmless Econometrics Ch 4 to 4.1.2 and 4.4, pp. 113-133, 150-172

Additional resources:
• Murnane, Richard and John Willett (2010). Method Matters, Chapter 10 (73 pages!!)

Questions:
• What is the difference between the “structural equation” or “econometric” approach to IV, versus the “potential outcomes” approach to IV (see Angrist, Imbens & Rubin 1996)?
• What are the key assumptions needed to justify a causal interpretation of an IV estimate?
• What are compliers, defiers, always-takers and never-takers in the context of IV?
• In words, define what Angrist means by the “first stage” equation/relationship and the “reduced form” equation/relationship (see p. 10). The first stage gives the effect of ____ on ____? The reduced form gives the effect of ____ on ____?
• Angrist explains that the IV/2SLS estimator can be interpreted as a “re-scaling” of the reduced form estimate to account for imperfect compliance (p. 11). How much difference will this re-scaling make if the difference in treatment delivery between the two groups is 100 percentage points? 50 percentage points? 0 percentage points? What is the actual difference in treatment delivery in this experiment (see Table 2, pp. 11-12)?

Week 7 (October 21): Lotteries as Instruments

Readings:

Additional readings:
• Murnane, Richard and John Willett (2010). Methods Matter, Chapter 11

Questions:
• In each of these two studies, what is the instrument? What is the endogenous variable (treatment) of interest? What is the control condition? (Note: Deming focuses on “reduced form” findings and doesn’t mention IV until page 2110; why do you think this is?)
• In each study, what schools/students are the estimates “local” to? In other words to what kinds of schools/students may these results be generalized?
• What can each study say (or not say) about the mechanisms driving the effects they observe?
• In what ways does the empirical setup in Deming’s paper differ from the other two? What are the implications for internal/external validity?
Week 8 (October 28): Regression Discontinuity

Readings:

Additional readings:

Questions:
1. What are the key assumptions that justify causal inference with regression discontinuity? What are some standard ways to check for these violations?
2. What does “bandwidth” refer to and how do researchers decide on the “right” bandwidth to use?
3. What is a “sharp” versus a “fuzzy” regression discontinuity, and how does IV relate to this?
4. Why don’t Scott-Clayton & Rodriguez use assignment to remediation as an IV for actual remedial coursework?
5. In the Scott-Clayton & Rodriguez paper, what difference does it make (if any) whether the placement test is a highly accurate measure of incoming ability, versus a very noisy one?

Week 9 (November 4): Differences-in-Differences

*** PROPOSAL DUE BY 5:00PM WEDNESDAY NOVEMBER 5TH ***

Readings:
- Chapter 8 of Murnane and Willett, pp. 135-164.

Questions:
- What are the key assumptions justifying causal inference with a difference-in-difference approach?
- What is the plausibly exogenous source of variation that Dynarski utilizes to estimate effects in each paper?
- What are the two dimensions of difference in each study? What does her difference-in-difference analysis add, above and beyond a simple first difference approach?
- Describe who is eligible for the SSSB program. Why does Dynarski focus on estimating effects only for a subset of eligible students – those with deceased fathers?
Week 10 (November 11): Self-Study. SELF-STUDY REVIEWS DUE FRIDAY NOV. 14TH.

***Assignment: This week, instead of lecture I would like you to explore free online resources relating to Stata, statistics, econometrics, and causal inference. Please spend 1-2 hours working through an online resource relevant to the course, and write a brief (1-page) review for your classmates (present and future) that will count toward your participation grade. In the review please include such information as:
- Name and URL of the resource
- Credentials of the author/content provider (is the source trustworthy?)
- Structure of the resource and time to complete
- Quality and overall usefulness of the resource – would you recommend it for students in this class? Why or why not?

Possible resources (you may use others, as long as they are free and available online):
- Econometrics Academy on YouTube (breaks a full econometrics course into many videos, from econometrics software, to linear regression, to specific econometric models): https://sites.google.com/site/econometricsacademy/home
- Ben Lambert’s graduate course in econometrics (esp. videos 54 and beyond): https://www.youtube.com/playlist?list=PLwJRxp3blEvaxmHgI2iOzNP6KGLSyd4dz
- SREE 2014 Conference Videos (I recommend the opening address by John Ioannidis on “Reproducible Research”): https://www.sree.org/conferences/2014s/pages/video.php
- Stata tutorials: http://www.ats.ucla.edu/stat/stata/notes/default.htm
- 2-hr lesson on regression: https://www.khanacademy.org/math/probability/regression/regression-correlation/v/correlation-and-causality

Week 11 (November 18): Fixed Effects as an Identification Strategy

Readings:
- Angrist & Pischke (2009). Mostly Harmless Econometrics Ch 5 through 5.3

Additional examples:

Questions:
- What are the key assumptions underlying fixed effects as an identification strategy? Think in terms of a specific research question and consider whether the assumptions are likely to hold.
- To what populations/contexts do fixed effects estimates generalize? I.e. whom are these estimates “local” to? It may help to consider – what cases/observations will be excluded when you switch from a basic OLS to a fixed effects approach?
• Why are measurement error and omitted variables bias of particular concern in fixed effects methods (and indeed all methods based on matching cases)?

Week 12 (November 25): Propensity Scores and Matching
*** FIRST DRAFT DUE BY 5:00 PM WEDNESDAY NOVEMBER 26 ***

Readings:

Additional resources:

Questions:
1. How is matching on a propensity score different from directly matching on covariates? Why would anyone do one versus the other?
2. In what ways is propensity score matching different from OLS? In what ways is it similar?
3. What does it mean to “check for common support”? Is it obvious that having plenty of “common support” is always a good thing? What can be done if the common support condition fails?
4. For each of the two studies this week, what are possible answers to the question, “What explains why individuals with the same propensity score received different values of the ‘treatment’?”

November 28-29: Thanksgiving Break

Weeks 13-15 (December 2-16): Presentations

Readings:
• Murnane & Willett, Ch. 13, pp. 332-349.
• Students are also expected to skim the papers that other teams have chosen to replicate. These will be announced in class no later than one week prior to the relevant session.

Questions:
1. How are “forward causal inference” and “reverse causal questions” related? How are they distinct?
2. Gelman & Imbens describe Why versus What If questions. Do you think one is more important than the other in education or other social science research? Are there other basic questions social scientists should attend to in their research?
3. What, for you, have been the most notable “takeaway messages” from this course? What questions did you expect to be answered that were not answered?