Course Outline DS925 (formerly SI915)

Causal Inference for Management Research (Fall 2016)
Course Meets: Tuesdays 2:00-5:00 PM, Questrom School of Business, Room 615 / 658

Instructor: Timothy Simcoe
E-Mail: tsimcoe@bu.edu
Homepage: http://people.bu.edu/tsimcoe
Phone (Fax): 510.685.2020
Office Hours: Schedule by email

Objectives

How should you respond when a referee says your paper has an endogeneity problem, or a seminar participant asks about your identification strategy? This course will teach you to apply methods for estimating causal relationships using observational (i.e. non-experimental) data, and in the process, how to address issues of endogeneity and identification that commonly arise in applied empirical research. We will discuss how to establish what relationships exist in the data, when you can interpret these relationships as causal, and how you can convince your audience of your results (without overselling). We will also discuss the importance of careful theoretical thinking, and a detailed knowledge of relevant institutions for this type of empirical research, particularly as applied to management topics.

The course has two main goals. The first is to develop a clear understanding of the conceptual difficulties associated with establishing causality in empirical research. In particular, I hope to de-mystify several concepts that are frequently invoked as “problems” or “issues” in empirical research, but seldom clearly explained. A good grasp of these concepts will lead to improved research design and a sharper understanding of the relative strengths and weaknesses of various statistical methods. The second goal is to describe some of the practical problems that arise in the application of these methods. I will place a special emphasis on testing the key underlying assumptions. The ultimate goal is for you to leave prepared to undertake your own empirical research.

Preparation and Prerequisites

This course is designed to complement a graduate sequence in econometrics, but it should be accessible to students with basic knowledge of probability and statistics. We will emphasize intuition and application over proof. However, the readings and class discussion will cover technical material. Most problem sets and in-class examples will be taught using the Stata statistical software package, and students will find that a working knowledge of Stata is quite valuable for following the class discussion. However, I have started to translate some course materials into R, and will accept assignments that are completed using that open source alternative.

Auditing

You are welcome to audit this class as long as you show up prepared to contribute to class discussion. I will only grade your assignments (including comments on the final paper) if you enrol for credit.
Assignments & Grading

Grades will be based on class participation (15%), two problem sets (15% each), a post-publication review (10%), a short simulation (5%), a replication exercise (20%), and a final paper (20%).

Class participation (15%): Read the materials, come to class (on time), and participate in the discussion. I will strive to create opportunities for everyone to contribute. Asking a good question is often more valuable than having the best answer.

Problem Sets (15 each%): I will provide two empirical exercises that allow you to practice the methods we learn in class within a fairly controlled environment. The data will be relatively clean, and the tasks should be clearly described (so feel free to ask questions if they are not). Students may work on these assignments in teams, but should write up their final submissions individually.

Post-publication review (10%): Choose a highly-cited empirical paper that was published within your field during the last ten years. Write a 2-3 page review of the paper that discusses and/or critiques its empirical methods using concepts from class. The review can be positive, negative or neutral about the overall merits of the paper. You are permitted, but not required, to obtain the underlying data and check the paper’s key results. Examples can be found at the SRI “On Further Review” web site (http://onfurtherreview.blogspot.com/), and I hope – with your permission – to submit several of your assignments to that outlet.

Simulation assignment (5%): A short assignment distributed in the first class to get you thinking about identification and test your Stata/R programming skills.

Replication/Reproduction Exercise (20%): Choose an existing empirical paper to replicate, and discuss/critique the robustness of the results using concepts from class. Several journals (e.g. Management Science, American Economic Review, American Economic Journals, Journal of Applied Econometrics, Journal of Business Economics and Statistics) and various researchers post data from published papers. While there is no page limit, my guess is that the reports will contain 1-4 pages of text, 2-5 tables and/or figures, and a Stata “.log” output file. One to three tables would likely replicate results from the existing paper and one or two more would present results that are not shown in the paper. If you can communicate the core ideas in less space, no problem. If you need more space, that’s fine too. The key is to show that you could reproduce the main results and that you tried some additional specifications (informed by what we do in class) to check robustness. Please confirm with me that your chosen paper is appropriate before starting to gather data, and don’t hesitate to ask questions at any point in the process!

Final Paper (20%): At the end of the semester, you should submit a written “research design.” This 4 to 6 page document will describe how you plan to implement an empirical study. Your research design should read like the “Data and Methods” section of a high-quality empirical paper. I expect to see a description of your data, a specification for the regressions you will perform, and (most importantly) discussion and justification of the assumptions that your reader must maintain in order to believe that your analysis constitutes an answer to the proposed research question.

For this assignment, I strongly recommend that you choose a question you are actually working on. Preferably, you have the data in hand. However, an acceptable alternative is to choose a research question that leads to a regression specification (or set of hypotheses) developed as part of a previous
class assignment. In either case, you should submit a 1 or 2 page summary of the theory / hypotheses along with your research design, for a total of 5 to 8 pages.

**Assigned Reading**

Each class will have several assigned readings. There are three types of reading:

1) Conceptual readings deal with tools and methods for causal inference. Some conceptual readings are academic papers, and others are chapters from *Mostly Harmless Econometrics: An Empiricist's Companion*, by Joshua Angrist and Steve Pischke. This is an excellent handbook for applied empirical research, and I highly recommend that you purchase a copy. Since MHE is not a complete reference, you may also wish to get a copy of William Greene’s *Econometric Analysis*, Jeffrey Wooldridge’s *Econometric Analysis of Cross Section and Panel Data*, or Cameron and Trivedi’s *Microeconometrics*.

2) Applied Readings are research papers that use the tools and methods we learn in class. For each Applied Reading, you should arrive in class ready to answer to the following questions: What is the research question? What is the unit of observation? What are the sources of variation? What are the key estimating equation(s)? What are the results and interpretation? In some cases, you will have access to the underlying data used in an Applied Reading, and I strongly encourage you to play around with it.

3) Optional readings provide additional detail on topics related to those covered in class.

**Calendar**

<table>
<thead>
<tr>
<th>Room</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>615</td>
<td>Sep 6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>615</td>
<td>Sep 13</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>615</td>
<td>Sep 27</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>615</td>
<td>Oct 4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>615</td>
<td>Oct 20</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>658</td>
<td>Oct 18</td>
<td>Simulation</td>
</tr>
<tr>
<td>7</td>
<td>615</td>
<td>Oct 25</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>615</td>
<td>Nov 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>615</td>
<td>Nov 8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>615</td>
<td>Nov 15</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>615</td>
<td>Nov 22</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>615</td>
<td>Nov 29</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>658</td>
<td>Dec 6</td>
<td>Post-pub review</td>
</tr>
</tbody>
</table>

**No Class on October 11**

<table>
<thead>
<tr>
<th>Room</th>
<th>Date</th>
<th>Topic</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>658</td>
<td>Oct 18</td>
<td>Problem Set 1</td>
</tr>
<tr>
<td>7</td>
<td>615</td>
<td>Oct 25</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>615</td>
<td>Nov 1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>615</td>
<td>Nov 8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>615</td>
<td>Nov 15</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>615</td>
<td>Nov 22</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>615</td>
<td>Nov 29</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>658</td>
<td>Dec 6</td>
<td>Final Paper Due</td>
</tr>
</tbody>
</table>

**Detailed Reading List**

KEY: C = Required conceptual reading; A = Required application reading; O=optional reading.
**Session 1: What is Identification?**


C: Mostly Harmless Econometrics, Chapters 1 & 2.


**Session 2: Classic problems & field-experiments**

Go to this web site, and find examples that correspond to Omitted variables self-selection, simultaneity and reverse causality. Be prepared to discuss:

http://jfmueller.faculty.noctrl.edu/100/correlation_or_causation.htm

Part 1: Classic Identification Problems

*(Self) Selection*


*Simultaneity & Reverse Causality*


State-dependence vs. Heterogeneity


Co-linearity


The Reflection Problem


Part 2: Field Experiments

C: Mostly Harmless Econometrics, Chapter 3.

C: Abhijit Banerjee, Sylvain Chassang and Erik Snowberg, “Decision Theoretic Approaches to Experiment Design and External Validity” Working Paper, April 2016. SECTIONS 1 1, 3.1.1 and 3.2,1 ONLY


Session 3: Selection on Observables


**Session 4: Matching Methods**


**Session 5: Instrumental Variables**

C: Mostly Harmless Econometrics, Chapter 4.


Comment” American Economic Review, forthcoming. (http://www-personal.umich.edu/~albouy/)


**Session 6: IV Continued**


**Session 7: Regression Discontinuity**

C: Mostly Harmless Econometrics, Chapter 6.


Session 8: Panel Data and Fixed Effects

C: Mostly Harmless Econometrics, Chapters 5 and 8.


Session 9: Difference-in-Differences


**Session 10: Unification and Extension**

Read the following post to prepare for the replication assignment: http://www.env-econ.net/2016/03/replication-vs-duplication-whats-the-difference.html


**Session 11: Non-linear models**


Logit, Probit and LPM

http://davegiles.blogspot.co.uk/2012/06/another-gripe-about-linear-probability.html
http://davegiles.blogspot.co.uk/2012/06/yet-another-reason-for-avoiding-linear.html
http://www.mostlyharmlesseconometrics.com/2012/07/probit-better-than-lpm/

Non-linear interactions


Count Data


Duration Data


Session 12: “Structural” Models


C: Journal of Economic Perspectives (Spring 2010). Symposium: Con Out of Economics (Skim)


Session 13: Conclusions


C: Section I of the Strategy Research Initiative “Strategy Reader” available at (http://strategyresearch.net/strategy_reader)


O: Journal of Economic Literature (June 2010). Forum on the Estimation of Treatment Effects