SI915 Course Outline

Models & Methods for Causal Inference in Strategy Research (Spring 2012)
Course Meets: BU School of Management

Instructor: Timothy Simcoe
E-Mail: tsimcoe@bu.edu
Homepage: http://people.bu.edu/tsimcoe
Phone (Fax): 617.358.5725
Office Hours: By Appointment, or just knock on my door

Objectives

What should you do when a referee says your paper has an endogeneity problem or a seminar participant asks how your estimates are identified? This course introduces a “toolkit” for attempting to estimate causal relationships using non-experimental data. 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 think carefully about the interaction between large-sample empirical research, qualitative institutional data, and theory, especially the importance of careful theoretical thinking (in the context of the institutional details) for empirical research.

There are two main goals. The first is to develop a clear understanding of the conceptual difficulties associated with establishing causality in empirical work. In particular, I hope to de-mystify several concepts that are frequently invoked as “problems” or “issues” in applied 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 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. Problem sets and in-class examples will be taught using the Stata statistical software package.

Assignments & Grading

Grades will be based on class participation (10%), a problem set (30%), a replication exercise (30%), and a research paper/presentation (30%).

Class participation (10%): Read the materials; come to class; participate.
Problem Set (30%): I will provide a set of empirical exercise and data that will allow you to practice the methods we learn in class within a fairly controlled environment.

Replication Exercise (30%): Choose an existing empirical paper to replicate, and discuss/critique the robustness of the results using concepts from class. Several journals (e.g. American Economic Review, American Economic Journals, Journal of Applied Econometrics, Journal of Business Economics and Statistics) and researchers (Bronwyn Hall, Bruce Hansen, Justin Wolfers, Dan Trefler, myself) 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 perhaps show 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 replicate the main results and that you tried some other specifications (informed by what we do in class) to check robustness. Please confirm with me that your chosen paper is appropriate before you work on it, and don’t hesitate to ask questions at any point in the process!

Research Paper / Presentation (30%): 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 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. An alternative is to use a 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.

Class Calendar

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Jan 25</td>
<td>What is identification?</td>
<td></td>
</tr>
<tr>
<td>2 Feb 1</td>
<td>Field experiments</td>
<td></td>
</tr>
<tr>
<td>3 Feb 8</td>
<td>Selection on observables</td>
<td></td>
</tr>
<tr>
<td>4 Feb 15</td>
<td>Matching etiquette</td>
<td></td>
</tr>
<tr>
<td>5 Feb 22</td>
<td>Instrumental variables</td>
<td></td>
</tr>
<tr>
<td>6 Feb 29</td>
<td>IV Etiquette</td>
<td></td>
</tr>
<tr>
<td>7 Mar 7</td>
<td>Panels and fixed effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPRING BREAK</td>
<td></td>
</tr>
<tr>
<td>8 Mar 21</td>
<td>Diff-in-diffs etiquette</td>
<td></td>
</tr>
<tr>
<td>9 Mar 28</td>
<td>Regression discontinuity</td>
<td></td>
</tr>
<tr>
<td>10 Apr 4</td>
<td>Non-standard data problems</td>
<td></td>
</tr>
<tr>
<td>11 Apr 11</td>
<td>Classic identification problems</td>
<td>Replication</td>
</tr>
<tr>
<td>12 Apr 18</td>
<td>Using theory for identification</td>
<td></td>
</tr>
<tr>
<td>13 Apr 25</td>
<td>Quality in empirical research</td>
<td></td>
</tr>
<tr>
<td>14 May 2</td>
<td>Student presentations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paper &amp; Prest'n</td>
<td></td>
</tr>
</tbody>
</table>
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, but most 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 (in theory/ideal and in the data)? What are the sources of (exogenous and endogenous) 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.

C = Required conceptual reading; A = Required application reading

**Session 1: What is Identification?**

C: Mostly Harmless Econometrics, Chapters 1 & 2.


**Session 2: Randomization and Field Experiments**

C: Mostly Harmless Econometrics, Chapter 3.


Evidence From India” NBER Working Paper No. 16658


**Session 3: Selection on Observables**


**Session 4: Matching Etiquette**


**Session 5: Instrumental Variables**

C: Mostly Harmless Econometrics, Chapter 4.


**Session 6: IV Etiquette**


**Session 7: Panel Data and Fixed Effects**

C: Mostly Harmless Econometrics, Chapters 5 and 8.


Session 8: Difference-in-Differences Etiquette


Session 9: Regression Discontinuity

C: Mostly Harmless Econometrics, Chapter 6.


**Session 10: Data problems**

*Count Data*


*Endogenous Dummies*


*Non-linear interactions*


*Age-Year-Cohort Problems*


*Duration Data*


**Session 11: Classic identification problems**

*Heterogeneity vs. State-dependence*

**Simultaneity**


**Identifying Clusters**


**The Reflection Problem**


**Session 12: Structural Models – Using theory for identification**


**Session 13: Views on Quality in Empirical Research**


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

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

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