Quantitative Methods 3
Government 703
Fall 2015

Instructor: Marko Klašnja, ICC 302-N, marko.klasnja@georgetown.edu

Time and place: MW 2:00–3:15PM, New North 412

Office hours: M 4-6PM

Overview

In this course, we will cover two major topics: causal inference and maximum likelihood estimation. The first half of the course – or a bit more – will be devoted to studying the canonical approaches under the general umbrella of causal inference, including the potential outcomes framework, the design and analysis of experiments, the challenges of analysis of observational data, and a suite of approaches commonly labeled "natural experiments." Causal inference approaches are rapidly becoming the mainstream of political science, and even encroaching on some of the existing approaches to research. This part of the course aims to introduce you (in more detail) to this paradigm.

The second part of the course will cover maximum likelihood estimation (MLE), primarily to revisit and make up for some lost time in Quant 2. We will begin with the basics of MLE theory and mechanics, and then move into the most common applications: the models for binary, multichotomous, ordered, and count dependent variables. Time permitting, we will also discuss the basics of time duration models.

The two parts of the course do not conflict with each other conceptually, but do not fit hand-in-glove either. Causal inference approaches place considerably more emphasis on research design than aspects such as the nature of the dependent variable – a point of departure for most of the MLE applications we will cover. We will take note of tensions like these as we go through the course. At the same time, I want to be clear that the focus of this class rests more heavily with the first part of the course. In this respect, a larger share of the course assignments will focus on causal inference (more details are below).

Grading

Problem Sets	50%
Final research proposal	50%

Problem sets

There will be approximately 6-7 problem sets. The lectures will cover the *conceptual* material. That is, given the amount of material we will try to cover, the lectures will *not* focus on how a particular approach is implemented in practice (say, with `Stata` or `R`). This will primarily be the focus of the problem sets, which will typically include working with simulated or real data.

You are encouraged to work on the problem sets together. However, you must submit your own write-up. When there is coding involved, you will be required to submit an annotated code as well. Please use either `Stata` or `R`. One problem set will likely ask you to write a referee report on an article employing one of the techniques we have learned. This assignment must be done individually and not as a group.

You can submit the first half of problem sets in any format you want (e.g. Microsoft Word or LaTeX). The second half of problem sets *must* be submitted in LaTeX. For better or worse, your knowledge of LaTeX often serves as a signalling device to your (academic) peers, and so if you have not done it already, it is a good investment to learn how to use it. Of course, when the problem set involves creating estimation output, you are well advised to have your statistical software export LaTeX-ready output tables. For both LaTeX and `Stata/R` questions, your peers and the internet are your best friends, and you should rely on both extensively. I will also occasionally provide some example code or pointers.

Research proposal

There is a paper requirement for this course. However, I do *not* expect you to write a complete research paper in one semester. Instead, you will be expected to write a research proposal, which should include the following: (1) a statement of a research question or puzzle, (2) how and why it may advance our current knowledge, (3) a *brief* review of the literature related to the topic, (4) a theoretical argument that can be used to provide an answer to the question, (5) testable hypothesis or hypotheses drawn from the theory, and (6) preliminary description of what data and methods may be used to test these hypotheses, including how the hypotheses could be falsified.

There are two specific requirements with respect to item (6) in the list above. First, my view is that political science is increasingly putting a premium on originally generated data, such as surveys, survey experiments, field experiments, coding of hard-to-get data, etc., even if paired with off-the-shelf datasets (e.g. Polity, Correlates of War, or the World Bank Development Indicators). This is especially the case in the political science job market for ABDs looking for jobs with their dissertation work. Therefore, I want you to propose one such original data collection effort, hopefully in the direction of your planned dissertation work. You will be expected to discuss in some detail the particulars of such a planned data collection (e.g. sampling frame for a survey, a randomization

scheme, archival sources for coding, etc.). Second, you will be required to identify (at least) one method from the first part of the class that you could plausibly apply to your question of interest, such as an original experiment or a natural experiment. You will be asked to discuss the "identifying assumptions" in detail.

We will devote one class, roughly in the seventh week of the course, to collectively brainstorming how to design your research proposals (see below for more details).

You should be thinking about your paper and discussing the topic with me by the end of the fifth week of the course. Note that if your research interest lies primarily in topics covered later in the class, you should be prepared to read ahead. Also, for the papers related to your chosen topics, I recommend reading them not just for their substance, but also for how they present the aspects you are required to include in your proposal.

This requirement can be combined with a paper requirement in other class(es) you will be taking concurrently. In my view, the goal is to have you designing research proposals that will lead to published academic articles, and if cross-fertilization across classes helps with that, all the better. However, there are two prerequisites to doing so: (1) the faculty in the other class will have to agree, and (2) you will have to inform me in advance and discuss with me the details of how you will address the specific requirements in both classes.

The proposal is due by the last week of the course. No incomplete grades will be allowed, except in exceptional circumstances.

Materials and readings

There will be readings for each class. You are expected to read *ahead* of the lecture. The readings will mostly come from these four textbooks:

Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2008.

Morgan, Stephen L., and Christopher Winship. *Counterfactuals and Causal Inference*. Cambridge University Press, 2014 (2nd Edition).

Gerber, Alan S., and Donald P. Green. *Field Experiments: Design, Analysis, and Interpretation*. W.W. Norton & Company, 2012.

Long, J. Scott. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, 1997.

I have not ordered copies for the bookstore since they are usually more expensive than what you can find when you shop around on the web.

We will also read selected chapters from the following books, which will be made available in electronic copy on the course website:

Dunning, Thad. *Natural Experiments in the Social Sciences: A Design-Based Approach.* Cambridge University Press, 2012.

King, Gary. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. University of Michigan Press, 2001.

Box-Steffensmeier, Janet M., and Bradford S. Jones. *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press, 2004.

Cameron, A. Colin, and Pravin K. Trivedi. *Microeconometrics: Methods and Applications*. Cambridge University Press, 2005.

Finally, we will also read several articles, which will be made available on the course website.

Topics

Topics listed below will be covered in about 2 class sessions each (i.e. for one week). The outline of the structure is tentative and may change somewhat over the course of the semester. Readings are indicated in *italic*.

Part I: Causal inference

1. Potential outcomes framework, identification
   Angrist and Pischke, chapters 1-2; Morgan and Winship, chapters 1-2

2. Design and analysis of experiments; difficulty of regression control
   Gerber and Green, chapters 3-4; Morgan and Winship, chapters 3-4

3. Matching methods
   Morgan and Winship, chapter 5; Ho et al. (2007)

4. Natural experiments, instrumental variables
   Dunning, chapters 1-2; Angrist and Pischke, chapter 4; Gerber and Green, chapters 5-6

5. Repeated observations, fixed effects, diff-in-diff, synthetic control
   Angrist and Pischke, chapter 5; Abadie et al. (2015)

6. Regression discontinuity design
   Angrist and Pischke, chapter 6; Dunning, chapter 3, Caughey and Sekhon (2011)

7. Causal inference wrap-up
   John Huber Monkey Cage blog post; Imbens and Gelman (2013); Dunning, chapter 11

Class session on research proposals

To facilitate the discussion, everyone in the class will be required to circulate one week in advance a document with one research idea, in no more than a few paragraphs. Preferably, this idea will

represent the topic of your eventual research proposal. We will choose several of the research ideas, and spend time in class discussing the process of turning those ideas into the outline of a research proposal.

Part II: Maximum likelihood

1. MLE theory and mechanics
   Long, chapter 4; King, chapters 1-4

2. Binary outcome variables, logit, probit, rare events
   Long, chapter 3; King and Zeng (2001)

3. Models for ordered and unordered dependent variables
   Long, chapters 5-6

4. Event counts, Poisson, negative binomial models, zero-inflated count models
   Long, chapter 8; Cameron and Trivedi, chapter 20

5. Duration models
   Box-Steffensmeier and Jones, chapters 1-4

Wrap-up

We will devote the last week of class to two things. In the first meeting, all students will present their research proposals with a *poster*. We will discuss the format in class.

In the second meeting, we will have an informal discussion about how the things you learned may be applied to real-world issues of doctoral research, the use of methods in political science (and elsewhere), and your research career. We will also briefly discuss important things we did not cover in this course that are worth exploring either later in the program, or elsewhere (other departments in D.C., ICPSR, EITM summer schools, etc.).