SOC: 476/POLISCI 490: Mixed Methods for Causal Inference

Professor: Robert Braun
1808 Chicago Avenue, room 103
robert.braun@northwestern.edu
Office Hours: MW 5:30-6:30 PM, 5:30-6:30 PM

Fall 2016
1812 Chicago Avenue, room 101
FR 11:00-1:50PM

Course description

This seminar is intended to help students become better consumers and producers of quantitative, qualitative and mixed methods causal inference. Whereas the former will make you a better colleague to fellow social scientists, the latter improves your job prospects (and helps you get colleagues to begin with). Based on the potential outcomes framework, this class will introduce students to qualitative and quantitative techniques to isolate causal effects. It will force students to move beyond "naive" regression and case study designs and explore how less naive combinations of both methods can strengthen our confidence in causal claims. It does this by emphasizing the importance of research design and making explicit the assumptions underlying causal arguments. The class argues against three types of research:

• Quantitative work that pays no attention to research design.
• Quantitative work that pays no attention to qualitative information.
• Qualitative work that pays no attention to case selection.

Via 7 weekly assignments, you will 1) learn how to critique and improve causal claims of existing research, 2) develop your own mixed methods research design for causal inference. As such, this class serves as a dissertation prospectus/NSF-proposal boot camp.

Required Texts

The following books are available for purchase at the Norris bookstore or your favorite online vendor:
• Thad Dunning. *Natural experiments in the social sciences: A design-based approach*. Cambridge University Press, 2012


I have also put copies on reserve at the main library. Other texts are available on CANVAS, ONLINE or via the library website either as EBOOKS or EJOURNALS. In addition, I would highly recommend you to purchase:


Prerequisites

Although, this course assumes no prior course work, students will benefit from having taken a basic regression class. For two assignments we will be using STATA. However, no in-depth knowledge of this statistical software is necessary. If you are not confident about your methods background come talk to me (short story: you will be fine).

Requirements

The formal requirements and grade calculation are as follows:

• You are required to attend all class meetings, to come to seminar having read and thought about the assigned material for the week, and to participate actively in class discussions. Class participation will compose 30% of the final grade.

• Complete all 7 assignments described below, and be prepared to discuss these ideas in class. Each assignment consists of two parts. Part A evaluates an existing piece of scholarship based on criteria introduced in class. In part B you apply these criteria to develop your own work. Each assignment typically consists of 24 pages of written work. Your assignments will be assessed on the quality of thinking that went into the project and research design, rather than the proposed theory. The assignments cumulatively count for 50% of the final grade. Assignments will be graded on a five-point scale. **Assignments are due the day of class before 7 am. They should be submitted to robert.braun@northwestern.edu. Put “RDCI: in the title.**
Students rework the research design part of the previous assignments into one integrated dissertation prospectus/NSF-proposal of around 15 pages and will be asked to give a short presentation. This counts for 20% of the final grade.

Relationship to Other Methods Classes at NU

This class provides a unified framework for mixed methods research that enables social scientists to assess causal arguments. As such it bridges a diverse set of methods and deals with a wide range of methodological issues. This class should not be considered comprehensive. Luckily, other great classes are available near you! Below one can find an overview of topics related to causality that are not or only briefly discussed in class. Between parentheses it mentions the corresponding class one should take to find out more about these topics. Students are encouraged to take this class in sequence with one or more of these seminars:

- Descriptive inference (Not discussed: see Sociology 476 Comparative Methods; Political Science 490: Research Design)
- Typologies (Not discussed: see Sociology 476 Comparative Methods)
- Concept formation (Week 3: see Political Science 490 Research Design)
- Comparative historical analysis (Week 1: see Sociology 439 Comparative Historical Analysis)
- Set-theoretic methods (Not discussed: see Sociology 439 Comparative Historical Analysis)
- Lab experiments (Week 1 and 5: see Political Science 407 Experimental Political Science)
- Data Collection (Week 3: see Sociology 403 Field Methods of Social Research/Sociology 405: Research Methods)
- Ethnography (Week 4: see Sociology 403 Field Methods of Social Research)
- Difference-in-difference and panel designs (Week 8: see Sociology 476 Statistical Methods for Hierarchical/Panel Data)
- Estimation of causal models (Week 5, 6 and 7: see Political Science 406 Causal Inference)
Week 1: Causality

This week we will go over the syllabus, talk about your research interests and discuss what causality and random assignment are. We will identify the strengths and weaknesses of different methods in assessing causality and explore how they can be combined.

GENERAL


DEBATE


- OR: Laura T Hamilton. More is more or more is less? parental financial investments during college. *American Sociological Review*, 78(1):70–95, 2013

ASSIGNMENT

No assignment.
Week 2: Questions and Answers

Finding a good question is by far the most important and difficult task of the social scientist. Where do good questions come from? There is no real answer but several eminent scholars have made suggestions. We will read some of these below. In addition, we will debate several successful pieces (in terms of placement and citations) of social science and try to figure out where their questions and answers came from. In the assignment you will try to formulate your own research puzzle. This is an extremely frustrating and demanding task that costs a lot time and sleep. Do not panic if you do not get it right the first time. Hardly anyone does.

GENERAL


DEBATE

SKIM:


**ASSIGNMENT**

**A.** Find one well-cited article in your area of interest. First, make an outline of the key points made in the introduction. Where does the research question come from? How do/does the author(s) find an answer? What is it about the question (if anything) that resulted in its publication and eventual citation by others? Describe in under 1 page. Be prepared to talk about this in class.

**B.** For your own project: find a puzzle or topic that (a) you think is important and interesting; (b) you think our understanding of is inadequate; and (c) you believe it is possible to collect data or information that would help us better understand it. A puzzle could be variation in an outcome across time or space, a pattern that seems suboptimal or inefficient, or a phenomenon that simply does not make sense and you dont get. This could come from something you read in another class or from a particular case. It could be a case that does not fit with what we think we know. Try not to formulate your question as a version of I want to look at the effect of x on y. In one page, describe your puzzle without using jargon or suggesting a possible explanation. You might present your puzzle as a brief account of an event, a motivating comparison, an anomaly, a regularity or relationship in statistical data, or some other way. Describe in under 1 page.

**Week 3: Theories, Concepts and Measurement**

This week we will briefly talk about theory construction, conceptualization and measurement. We will read some common criteria for the evaluation of measurement and conceptualization and apply these to some examples of successful scholarship in political sociology.

**GENERAL**


DEBATE


ASSIGNMENT

A. Find one well-cited article in your area of interest that makes a causal argument. Briefly describe the causal argument, concepts and measurement. Then deploy Gerrings criteria to evaluate all three. Be prepared to talk about this in class (<2 pages).

B. For your own project: identify and briefly summarize one main and two plausible alternative explanations that might resolve the puzzle you formulated last week. There are different ways to do this. You could develop explanations off the top of your head or you might collect several articles or books that examine roughly the same puzzle that you have posed (i.e., have roughly the same dependent variable) and draw out conjectures from those works. You may need to read about a particular case (or a small set of cases) that is an example (or are examples) of the phenomenon that you seek to explain. Try to think in terms of variables. State your three explanations as clear, testable, and falsifiable causal propositions. A falsifiable proposition is one that can be shown to be wrong. If you cannot think of any evidence that would convince you that the proposition is wrong, you need to reformulate the proposition. If it helps, you may render the propositions (and their falsifiability) in diagrams or equations or both. Be clear about the mechanism(s) that link the cause (independent variable(s)) and effect (outcomes you are trying to explain).
Week 4: Case Study Before Quantitative Analysis

This week we will discuss research designs in which case studies precede quantitative analysis. Case studies often allow for internally valid measurement, the identification of fine-grained mechanisms and theory generation. By combining case studies with quantitative research external validity can be improved.

GENERAL


DEBATE


ASSIGNMENT

A. Find one well-cited article in your area of interest that relies on case studies only to make a causal argument. How does this paper use case studies to understand processes and what inferential leverage is gained through a close examination of these cases. Outline all the steps in the causal argument. What evidence is brought to bear for each step? Is the evidence compelling? Does the
author say anything about how additional insights gleaned from processtracing apply to the general question beyond this particular case (<2 pages)?

B. For your own project: identify several direct observable implications of each proposition that you formulated last week. Try to think of observable implications at different levels or units of observation (e.g., individuals, groups, spatial units). Construct at least three hypotheses about the observable implications of your possible explanations. Explicitly define your key independent and dependent variables. Clearly state how your variables vary. Second, describe how you will operationalize these variables consistently and discuss the tradeoffs, if any, involved in the operationalization that you propose. If helpful, follow the steps outlined by Adcock and Collier/Gerring. Make a list of the information you need to collect to (a) know whether your explanation is wrong (to disconfirm your hypotheses), (b) make a compelling case that your argument is correct, and (c) dismiss the major alternative plausible hypotheses. Think about what would be ideal data with which to assess each conjecture (<2 pages).

Week 5: Regression and Case Studies

When analyzing observational data, regression with control variables is often used to reduce omitted variable bias due to non-random assignment. This is done through conditioning on observable confounders. Not all controls however improve causal inference. Moreover, not all confounders can be observed. Combining regression with detailed case knowledge can improve our causal estimates in three ways 1) control through research design 2) the detection of omitted variables 3) establishing causal linkages.

GENERAL


• Jason Seawright. *Multi-method social science: Combining qualitative and quantitative tools*. Cambridge University Press, 2016. Chapter 4. PURCHASE.

DEBATE


ASSIGNMENT

A. Go back to the article you discussed last week. Is there anyway in which the paper could benefit from an additional quantitative analysis of large-n data? If so, develop a research strategy for building this dataset, and provide a few sample rows. If possible, describe coding rules. State the specific hypotheses amenable to quantitative tests and how the data gathered would allow you to assess those hypotheses.

B. For your own project: identify a relevant universe of cases. Be clear about the unit of analysis and what you mean by “case and observation. Select cases that you could use to assess the hypotheses of the observable implications of your theory. Explain the rationale you used for selecting this unit of analysis. Explain what additional knowledge, if any, might be gained by collecting qualitative evidence on a small number of additional cases. Describe a case (which might be a country, province, city, piece of legislation, group, individual, etc.) that you intend to use to investigate in detail and theorize empirically the translation of values on independent variables onto values on dependent variables. How will you use this case to understand process and what inferential leverage might you gain through a close examination of cases (e.g., supporting an argument or ruling out alternative explanations or both)? Finally, how will you know if additional insights gleaned from your “processtracing apply to the general question beyond this particular case?

Week 6: Matching and Case Studies

Matching is a specific way of conditioning on confounders. Scholars prefer it over the deployment of normal controls because it makes less parametric assumptions and guarantees that counterfactual cases actually exist in the data.
The downside of matching is that it often requires sample reduction and as such threatens the external validity of estimates. Matching provides opportunities for case study researchers because of its similarity with conventional case selection techniques.

GENERAL


DEBATE


ASSIGNMENT

A. Find one well-cited article in your area of interest/Pick one exemplar that deploys regression and for which replication data is available online. Word to the wise: pick a paper that uses cases you are familiar with (US states, cities, countries etc) and do not pick a paper in which the observations are individuals. Briefly summarize the paper then choose one of the case selection methods we discussed last week. According to this method what case(s) should the author have selected for in-depth case studies? What type of data would you look for to study this case? Explain how you got to this answer. Report relevant statistics (<2 pages).

B. For your own project: outline a largeN dataset that would produce some statistical test of your theory using crossnational, subnational or survey data. Develop a research strategy for building this dataset, and provide a few sample rows. If possible, describe coding rules. State the specific hypotheses amenable to quantitative tests and how the data gathered would allow you to assess those
hypotheses. Optional: Produce some basic statistical test of the theory or an observable implication of your own theory using original data or an existing dataset. Provide descriptive statistics that show the plausibility of the theory, focusing on statistical and graphical descriptions of the dependent and principal independent variables (<2 pages).

Week 7: Natural Experiments, Instrumental Variables and Qualitative Data

Given the impossibility to measure all confounding variables, scholars across the social sciences more and more turn to natural experiments to establish causality. Case studies can help one discover natural experiments and corroborate their underlying assumptions.

GENERAL

• Thad Dunning. Natural experiments in the social sciences: A design-based approach. Cambridge University Press, 2012. Chapters 2, 4 and 7. PURCHASE.


DEBATE


• Jasjeet S Sekhon and Rocio Titiunik. When natural experiments are neither natural nor experiments. American Political Science Review, 106(01):35–57, 2012. LIBRARY EJOURNAL.

ASSIGNMENT

A. Find one well-cited article in your area of interest that deploys propensity score matching and for which replication data is available online. Word to the wise: pick a paper that uses cases you are familiar with (US states, cities, countries etc) and do not pick a paper in which the observations are individuals. Briefly summarize the paper then choose one of the case selection methods we discussed last week. According to this method what cases should the author
have selected for in-depth case studies? What type of data would you look for to study this case? Explain how you got to this answer. Report relevant statistics (<2 pages).

B. For your own project: create a causal graph of your main dependent variable and the most important independent variables that potentially explain it. Also draw out connections between the different explanatory variables. On what variables would you need to condition in your final analysis? On what variables should not you condition? How would you measure these variables? (<2 pages).

Week 8: Panel Data, Regression-Discontinuity and Qualitative Data

This week we will discuss two additional tools to improve causal inference with observational data: conditioning on time-invariant confounders and regression-discontinuities.

GENERAL


DEBATE

- Matthew Adam Kocher and Nuno P Monteiro. What’s in a line? natural experiments and the line of demarcation in wWII occupied france. *Natural Experiments and the Line of Demarcation in WWII Occupied France (July 31, 2015)*, 2015. AVAILABLE ONLINE.


ASSIGNMENT

A. Prepare for presentation.

B. For your own project: write a memo describing a possible endogeneity, omitted variable bias or sample selection problem in your project and how you intend to solve it or minimize the risk of biased inference. If you believe that causal identification is not a problem for your project, write a 23 page memo explaining why endogeneity, omitted variable bias and sample selection are not issues. Consider how a lab, field, or natural experiment could be used to put some aspect of your theory or the mechanisms underlying it to an empirical test. Be specific about the treatment, the unit at which the treatment is applied (e.g., individuals, villages), and how cases will be or were assigned to treatments and controls(<2 pages).

Week 9: Group presentations

Schedule TBA