Research Design
Carlos III-Juan March Institute
Spring 2017

Syllabus in progress and subject to changes. This version: February 3, 2017. The most recent version can be found here.

Instructor:

Simon Weschle
Email: sweschle@inst.uc3m.es
Office: 18.2.E.07
Office Hours: Thursday, 16:30-18:00, or by appointment
Class Time and Location: Monday, 9:00-12:00, Room 18.0.A.01

Course Description:

The goal of this course is to introduce you to state of the art research designs in quantitative political science. The course serves as a bridge between your substantive classes, where you learn about theoretical developments in the field, and your methodological classes, where you learn the techniques to conduct scientific analyses. A good research design is essential to connect these two elements: What is the best way to empirically answer the theoretical question you are interested in?

After a couple of weeks laying the groundwork, each week will cover different ways to design one's research. We will learn the general idea behind large-N, small-N, experimental, and quasi-experimental designs, and critically discuss examples of research that apply them. We will also discuss the advantages and disadvantages of the different research designs in furthering our understanding of the world more broadly.

This course is designed with an eye towards helping you to develop your own research. You will develop a research design proposal that identifies a question of interest to you, proposes an answer, and lays out a design to empirically test it. I encourage you to have a question as early in the semester as possible. This will allow you to think about how you could answer your question using the kind of research designs we talk about in a given week. Although the readings in class will focus on political science, you are welcome to investigate a topic in sociology or economic history as well. Feel free to talk to me during office hours about your project.

Course Requirements:

Research Proposal (60%):

The core requirement is to apply the things we learn in class by writing a research design paper. This consists of two parts:

- **Discussion and Set-Up:** The first step in a research project is to identify a research topic and to motivate it. For example, it could be that you observe a gap or inconsistency in the literature; that existing scholarship predicts empirical patterns at odds with reality; or that the literature makes predictions about a certain phenomenon that contradict each other. The choice of a research topic is more of an art than a science, and it requires considerable thought and effort. In particular, you need to (1) establish the unexplained empirical pattern that your research will address, and (2) master the literature to which you seek to contribute and demonstrate that it does not explain the empirical puzzle you

have identified. In the first part of your paper, you will accomplish these two things. The paper is due in class on **March 13**.

- **Argument and Empirical Strategy:** Identifying an empirical phenomenon and showing that the literature does not adequately explain it is the first (necessary but not sufficient) step of a research project. In a second step, you then need to craft a strategy to provide an answer to the puzzle, both theoretically and empirically. In the second part of the research design paper, you will lay out your theoretical argument and link it to a proposed empirical strategy that can test the observable implications of your theory. Be as specific as possible: What are the hypotheses that link your argument to the empirical data? What data will you use? Which is the variable you seek to explain? Which is/are the variable(s) that do the explaining? What are potential confounding factors, and how will you seek to control for their influence? What are the advantages of your empirical strategy, and what are its weaknesses and limitations? Note that you do not actually have to assemble the data and do the analyses. But you should be realistic about the kind of research design you are proposing – it should be something that could plausibly be done by a researcher subject to real-world constraints. In fact, it is a good idea to treat this research proposal as a first step towards your master thesis. You will make a short presentation of your research design in the last session of the class on **May 8**. The final paper (containing the potentially revised discussion and set-up as well as the argument and empirical strategy) is due on **June 23** per e-mail.

One of the most important things for an academic career is good (or at least not terrible) writing. If you are unable to convey your point in a clear and concise manner, it is unlikely that many people will care to read your work, no matter how insightful it is. For some advice about how to write academic papers, have a look at the following:

- Stimson, James A. (2006): "Professional Writing in Political Science: A Highly Opinionated Essay." `https://dl.dropboxusercontent.com/u/14523416/Writing.pdf`
- Weingast, Barry (2010): "Caltech Rules for Writing Papers: How to Structure Your Paper and Write an Introduction." `https://web.stanford.edu/group/mcnollgast/cgi-bin/wordpress/wp-content/uploads/2013/10/CALTECH.RUL_..pdf`
- Munger, Michael C. (2010): "10 Tips on How to Write Less Badly." `http://www.chronicle.com/article/10-Tips-on-How-to-Write-Less/124268`

You might also want to consult the following excellent book:

- Sword, Helen (2012): *Stylish Academic Writing.* Cambridge and London: Harvard University Press.

To see how your writing holds up, evaluate a page or two using the following tool: `http://writersdiet.com/?page_id=4`

Class Participation (40%): The quality of the seminar depends heavily on student participation. There will be some parts where I will lecture, but most of the class will be conducted seminar-style, that is heavily discussion-based. You are therefore expected to come to class having completed *all* the required readings and be prepared to talk about them. To ensure this, please send me half a page or so of questions and discussion points related to the week's readings by **noon the day before each class (only Sessions 3-9).** For example, you can raise questions about the papers' theory, research designs, the connection between the two, connections or conflicts between different readings or readings from prior weeks, and so on. You can also raise specific questions about things you did not understand. The comments can focus on all papers, a few, or just one. I will compile the questions/comments and send them to everyone.

Seminar Schedule and Reading Assignments

Session 1 (February 13): Introduction, Overview of Research Design

We begin with an overview session, in which we discuss the basic elements of the research process, debate the ways in which social science is a science (and maybe ways in which it is not), and talk about what good research questions are and how we can connect them to empirical evidence.

- *Social Science as a Science:* King, Gary, Robert O. Keohane, and Sidney Verba (1994): *Designing Social Inquiry. Scientific Inference in Qualitiative Research.* Princeton University Press. Chapter 1.
- *Overview of Research Design:* Trochim, William and James P. Donnelly (2007): *The Research Methods Knowledge Base.* Atomic Dog Publishing. Chapter 1. Skip sections 1.4b and 1.5.
- *Research Questions, Connecting Theory and Empirics:* Barakso, Maryann, Daniel M. Sabet, and Brian F. Schaffner (2014): *Understanding Political Science Research Methods. The Challenge of Inference.* Routledge. Chapters 2 and 3.

Session 2 (February 20): The Problem of Inference in the Social Sciences

As social scientists, we are interested in discovering causal relationships. What does that mean, how can we achieve it, and what are problems we commonly encounter?

- *Easy:* Barakso, Maryann, Daniel M. Sabet, and Brian F. Schaffner (2014): *Understanding Political Science Research Methods. The Challenge of Inference.* Routledge. Chapter 1.
- *A Bit Harder:* King, Gary, Robert O. Keohane, and Sidney Verba (1994): *Designing Social Inquiry. Scientific Inference in Qualitiative Research.* Princeton University Press. Chapter 3.
- *Another Bit Harder:* Holland, Paul W. (1986): "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396): 945-960. Read only Sections 1-4.

Session 3 (February 27): Large-N Inference I

After the first two sessions set the scene, we now start discussing different approaches to conduct empirical analyses. We begin with large-N statistical analyses, which have been the dominant approach in comparative politics over the last few decades. Researchers compile a data set that contains variation in the variable they are interested in, and use statistical analyses to explain this variance. In this session, we discuss the sources of the variation to be explained, focusing on two dimensions: *place* and *time*. For the former, studies in comparative politics can explain variation within a country (or part of a country), or between countries. For the latter, studies can explain differences at a certain point in time, over the course of time, or both.

- *Cross-Sectional Variation, Within-Country:* Chen, Jowei (2013): "Voter Partisanship and the Effect of Distributive Spending on Political Participation." *American Journal of Political Science* 57(1): 200-217.
- *Cross-Sectional Variation: Cross-Country* Pierskalla, Jan H. and Florian M. Hollenbach (2013): "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107(2): 207-224.
- *Time-Series Cross-Sectional Variation:* Ansell, Ben W. (2008): "University Challenges: Explaining Institutional Change in Higher Education." *World Politics* 60(2): 189-230. Skip the sections "The Microfoundations of Higher Education Policy" and "Extending the Model".

- *Panel Variation:* Cox, Gary W., and William Terry (2008): "Legislative Productivity in the 93d-105th Congresses." *Legislative Studies Quarterly* 33(4): 603-16.

Session 4 (March 6): Large-N Inference II

In the second session on large-N inference, we discuss different kinds of data that can be used for analysis. They differ along a number of dimensions, such as the unit of analysis, the time frame, the way the data is collected, and so on. The goal is not to provide an exhaustive list of data sources, but to provide you with some ideas about common kinds that researchers use, and to discuss the advantages and limitations of each of them.

- *Survey Data and Macro-Level Country Data:* Rehm, Philipp (2011): "Social Policy by Popular Demand." *World Politics* 63(2): 271-299.
- *Archival/Historical and Geographic Data:* Stasavage, David (2010): "When Distance Mattered: Geographic Scale and the Development of European Representative Assemblies." *American Political Science Review* 104(4): 625-643.
- *Administrative and Census Data:* Asher, Sam, and Paul Novosad (2016): "Market Access and Structural Transformation: Evidence from Rural Roads in India." Unpublished Paper. Skim section 2.
- *Text as Data:* King, Gary and Jennifer Pan and Margaret E. Roberts (2013): "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107(2): 326-343.

Session 5 (March 13): Small-N Inference

Instead of studying a large number of cases, many studies instead focus on one of a few cases, and study them in great detail. This is often, but not always, done with qualitative methods. In this session, we discuss different approaches to small-N inference and examine its advantages and disadvantages compared to large-N research.

- *Case Study:* Harding, David J., Cybelle Fox and Jal D. Mehta (2002): "Studying Rare Events Through Qualitative Case Studies: Lessons from a Study of Rampage School Shootings." *Sociological Methods Research* 31(2): 174-217.
- *Comparative Case Study:* Lapuente, Victor and Bo Rothstein (2014): "Civil War Spain Versus Swedish Harmony: The Quality of Government Factor." *Comparative Political Studies* 47(10): 1416-1441.
- *Synthetic Control Method:* Abadie, Alberto, Alexis Diamond and Jens Hainmueller (2015): "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2): 495-510.
- *Ethnographic Research:* Goffman, Alice (2009): "On the Run: Wanted Men in a Philadelphia Ghetto." *American Sociological Review* 74(3): 339-357.

March 20 – No Class (Festivo)

Session 6 (March 27): Experiments

Discussion and Set-Up due in class

So far, we have discussed studies that take some variation that naturally exists in the world and try to explain it. This is usually done by focusing on one specific mechanism, and by controlling for other factors that potentially affect the outcome of interest. Such "observational" studies have important drawbacks. First, we can never be sure that we have accounted for all

confounders, so there always is the chance that the outcome is really driven by some other mechanism that we failed to control for (omitted variables). In addition, it may also be the case that the mechanism we use to explain an outcome is in turn driven by another mechanism that we failed to take into account (endogeneity), or that the dependent variable in fact explains the independent one (reverse causality). Experiments get around these problems by randomly assigning observations to a treatment and control group. If the random assignment works, the values for all potential confounders should be the same in the treatment and control group, so they cannot affect the outcome and any difference between the two groups is due to the treatment. In other words, instead of taking some existing variation and try to explain it, experiments actively assign the variation to the study population in a controlled manner. Today, we discuss different forms of experiments.

- *Overview:* Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia, Eds. (2011): *Cambridge Handbook of Experimental Political Science.* Cambridge University Press. Chapters 1 and 2.
- *Lab Experiment:* Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein (2007): "Why Does Ethnic Diversity Undermine Public Goods Provision?" *American Political Science Review* 101(4):709725.
- *Survey Experiment:* Hainmueller, Jens and Michael J. Hiscox. "Attitudes toward Highly Skilled and Low-Skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review* 104(1): 61-84.
- *Field Experiment:* Wantchekon, Leonard (2003). "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin." *World Politics* 55(3): 399-422.

Session 7 (April 3): Natural Experiments I

In the last session, we have seen how experiments solve the crucial problems of endogeneity and omitted variable bias. However, we have also seen that it is not feasible to run experiments to answer many social science questions we are interested in. A way to get the benefits of randomization without actively running an experiment is to find and analyze so-called natural experiments. Here, "nature" has created situations where assignment to a treatment is as good as random. In this and the next section, we discuss approaches to study such natural experiments.

- *Overview:* Dunning, Thad (2012): *Natural Experiments in the Social Sciences: A Design-Based Approach.* Cambridge University Press. Chapter 1.
- *Matching:* Gilligan, Michael J. and Ernest J. Sergenti (2008): "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference." *Quarterly Journal of Political Science* 3(2): 89-122/
- *Natural Experiment, Cross-Section Regression:* Angrist, Joshua D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3): 313-336.
- *Natural Experiment, Fixed Effects/Difference-in-Differences:* Hainmueller, Jens and Dominik Hangartner (forthcoming): "Does Direct Democracy Hurt Immigrant Minorities? Evidence from Naturalization Decisions in Switzerland." *American Journal of Political Science.*
- *Regression Discontinuity Design:* Eggers, Andrew C. and Jens Hainmueller (2009): "MPs for Sale? Returns to Office in Postwar British Politics." *American Political Science Review* 103(4): 513-533.

April 10 & 17 – No Class (Semana Santa)

Session 8 (April 24): Natural Experiments II

We continue our discussion of approaches to study natural experiments. Then, we take a step back and discuss the impact of the so-called "identification revolution" in the social sciences, that is the development to prioritize research using experimental or quasi-experimental data over observational studies. In what ways does this help our quest to explain the world, and in what ways does it do harm?

- *Overview:* Angrist, Joshua D. and Jörn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press. Chapter 4 (Parts).
- *Instrumental Variable Regression:* Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369-1401.
- *Validity of Instrumental Variable Regression Assumptions:* Glaeser, Edward L. and Rafael La Porta and Florencio Lopez-de-Silanes and Andrei Shleifer (2004): "Do Institutions Cause Growth?" *Journal of Economic Growth* 9(3): 271-303. Read only sections 1 and 5.
- *Consequences of the Identification Revolution I:* Huber, John (2013): "Is Theory Getting Lost in the 'Identification Revolution'?" *The Political Economist* Summer: 2-4.
- *Consequences of the Identification Revolution II:* Samii, Cyrus (2016): "Causal Empiricism in Quantitative Research" *Journal of Politics* 78(3): 941-955.

Session 9 (May 4, 16:00-19:00): Assorted Additional Topics

We discuss some issues that often come up in empirical research. When evaluating an argument we generally focus on post-hoc explanations evaluated through statistical significance. But in the words of Daniel Kahnemann: "The ultimate test of an explanation is whether it would have made the event predictable in advance." How do social science theories fare? Then, we turn to the measurement of empirical concepts in the social realm, and how that can affect our inference. We discuss sources of model uncertainty and how they can be addressed in robustness tests. Finally, we talk about the importance of replicability in the social sciences.

- *Prediction:* Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke (2010): "The Perils of Policy by p-value: Predicting Civil Conflicts." *Journal of Peace Research* 47(4): 363-375.
- *Measurement I:* Mikhaylow, Slava, Michael Laver, and Kenneth R. Benoit (2012): "Coder Reliability and Misclassification in the Human Coding of Party Manifestos." *Political Analysis* 20(1):78-91. Skip section 3.
- *Measurement II:* Gallop, Max and Simon Weschle (forthcoming): "Assessing the Impact of Non-Random Measurement Error on Inference: A Sensitivity Analysis Approach." *Political Science Research and Methods.* Skip section "Extensions" and "Application: Leadership and Democratic Deliberation" (pp. 12-15).
- *Robustness:* Plümper, Thomas and Erik Neumeyer (2012): "Model Uncertainty and Robustness Tests: Towards a New Logic of Statistical Inference." Unpublished Manuscript. Skip sections 5-7.
- *Replication:* King, Gary (1995) "Replication, Replication." *PS: Political Science and Politics*, 28(3): 444-452. Skip the sections "Tenure and Promotion Review Committees", "Graduate Programs", "Editors and Reviewers of Books and Journals", and "Support for These Policies".

Session 10 (May 8): Research Design Presentations