PUBL0050
Advanced Quantitative Methods

Department of Political Science
University College London
2018–19

Instructor
Dr Jack Blumenau
B.13, 29-30 Tavistock Square
j.blumenau@ucl.ac.uk
Office Hours: 10am–12pm, Friday

Teaching Fellow
Philipp Broniecki
Committee Room 2 (G.01), 31 Tavistock Square
philipp.broniecki.14@ucl.ac.uk
Office Hours: 3pm–5pm, Wednesday

1 Introduction

This course provides an introduction to statistical methods used for causal inference in the social sciences. We will be concerned with understanding how and when it is possible to make causal claims in empirical research. In particular, using the potential outcomes framework of causality, we will focus on understanding which assumptions are necessary for giving research a causal interpretation, and on learning a range of approaches that can be used to establish causality empirically. The course will be practical – in that you can expect to learn how to apply a suite of methods in your own research – and theoretical – in that you can expect to think hard about what it means to make claims of causality in the social sciences.

We will address a variety of topics that are popular in the current political science literature, including experiments (laboratory, field, and natural); matching; regression; weighting; fixed-effects; difference-in-differences; regression discontinuity designs; instrumental variables; and synthetic control. Examples are drawn from many areas of political science, including political behaviour, institutions, international relations, and public administration.

Every week will involve the following sessions:

Lecture: **Tuesday 15:00–17:00 in Taviton (16) 433**
Class (Group 1): **Friday 10:00–11:00**
Class (Group 2): **Friday 11:00–12:00**
Class (Group 3): **Friday 12:00–13:00**

2 Pre-requisites

Students should have a working knowledge of the methods covered in typical introductory quantitative methods courses (i.e. to the level of PUBL0055 or equivalent). At a minimum, this should include hypothesis testing, multiple linear regression and multiple logistic regression.

If you don't have the required prerequisites but still wish to take the module, you will need to review material on these methods. You could consult either of the following textbooks:

- Agresti, Alan and Finlay, Barbara. 2009. *Statistical Methods for the Social Sciences, Fourth Edition*, Pearson International.
- Stock, James and Watson, Mark. 2015. *Introduction to Econometrics, Updated Third Edition*. Pearson.

Students without these pre-requisites may, under rare circumstances, be allowed to take the course, but you will need to convince me first!

3 Learning, Assessment and Feedback

3.1 Learning Outcomes

By the end of the course, students will be able to:

1. Understand the concept of causation as it is typically discussed in the social sciences
2. Make distinctions between observational and experimental studies
3. Define the assumptions required to make causal claims from quantitative data
4. Construct research designs that would yield credible causal effects
5. Implement a range of statistical methods which aim to estimate causal effects, including: experiments, matching, regression, weighting, fixed-effects, difference-in-differences, regression discontinuity; and synthetic control
6. Use the R statistical software in applied research
7. Critically evaluate the use of causal inference designs used in published work

These outcomes are indicative of what kinds of knowledge should be demonstrated on formative and summative assessments.

3.2 Assessment

3.2.1 Research Paper

Students will be evaluated through a 3000-word research paper applying the methods from the course to a research question chosen by the student.

The research paper should follow the basic elements of a novel research project. The paper should address a specific research question, identify the theoretical contribution, provide testable hypotheses, and implement a suitable design based on one of the methods that we study in the course. The paper should focus narrowly on a topic of the student's choice and display a depth of understanding of one of the approaches discussed on the course, rather than a survey of all methods.

A one-page summary of the proposed research paper will be due on Friday 7th December, 2018. This summary will not be assessed (it will not count towards the degree for this module), but you will receive feedback.

The final coursework is due on Wednesday January 9th, 2019 (2pm). The coursework will be submitted online via Moodle.

Plagiarism is taken extremely seriously and can disqualify you from the module (for details of what constitutes plagiarism see this document). If you are in doubt about any of this, ask me. No late work will be accepted and, given the long-term nature of the research paper, extensions will only be offered in exceptional circumstances.

3.2.2 Problem sets

For formative work, students will also complete short "problem set" assignments each week, which allow them to apply material from the course to concrete political science examples. While these formative assessments do not count toward the final mark, they provide an opportunity for peer and instructor feedback.

All assignments will be available on the course website the day before the relevant computer class (i.e. each Thursday) and annotated solutions will be released (also via the course website) the following Monday.

4 Course Materials

4.1 Online resources

- **Course website**: The main source of information for problem sets, class assignments, and readings will be the course website, which can be found at `https://uclspp.github.io/PUBL0050/`.

- **Moodle**: Other material relevant to the course will be uploaded to the course Moodle site, which can be found at: https://moodle-1819.ucl.ac.uk/course/view.php?id=6932.
- **Piazza**: We will be using Piazza as a discussion forum for the course. To register for this course on Piazza, you can visit this link: http://piazza.com/ucl.ac.uk/fall2018/publ0050. Piazza can also be accessed via a link on the course's Moodle page. I encourage you to use Piazza for both student-to-student and student-to-tutor communication, meaning that you should be a) posting questions for other students to answer and b) attempting to answer questions posted by other students. If you ask me a substantive question about the course via e-mail, I will ask you to post it on Piazza so that other students may also benefit

4.2 Textbooks and Readings

We primarily use the following textbooks on the course:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press, 2008.
- Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation.* WW Norton, 2012.

Both books provide excellent introductions to the potential outcomes framework which forms the conceptual core of this course. These books also cover the majority of the methodological approaches that we will study. That said, we will often focus on readings from other books/papers when necessary. The reason that we do not always follow a single textbook is that the field of causal inference is rapidly evolving, and there is no single canonical volume that would cover of all the interesting topics we focus on in this course.

For a less technical approach to some of this material, students might also consult:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect.* Princeton University Press, 2014.

though this should be in addition to, rather as a replacement for, the 2008 Angrist and Pischke book.

You may also find the following books useful as references for topics on which you have a particular interest:

- Stephen L Morgan and Christopher Winship. *Counterfactuals and causal inference.* Cambridge University Press, 2015.
- Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press, 2015.

In addition to these textbook treatments, it is essential that students read the articles set as "required" reading each week, and it is worth familiarising yourself also with the "recommended" reading. The required reading will often contain material that is not covered in the textbooks, partly because the methods on this course are at the cutting edge of the discipline and so are (sometimes) too new to have received coverage in textbooks and (often) it is more interesting to read the papers than the book.

The "recommended" readings will typically cover recent or important implementations of the methods we will learn about, and will be helpful in (at least) two regards. First, reading these articles will provide you with an understanding of when the methods we study can provide interesting answers to previously thorny empirical questions. Second, these articles will be helpful templates for the research paper you will write at the conclusion of the course.

4.3 Software

Throughout the course we will use the free and open source statistical analysis software R. You can download and install R on your personal computer from `https://cran.r-project.org/`. You should also install RStudio (`https://www.rstudio.com/products/rstudio/download3/`), which is a user-interface to R. Please ensure that both R and RStudio are installed on your personal computers before coming to the first computer class. The UCL machines will already have this software installed.

Students are not expected to have programming knowledge before starting class, and the computer labs will be centred around bite-size assignments which will help build knowledge of and intuition for coding in R. My hope is that students will enjoy the coding component of the course, rather than be scared by it!

We will cover the basics of using R in the first computer class, and you will be provided with all the relevant code necessary for completing the class assignments and problem sets each week (you will also be provided with solution code for the problem sets). That said, what you get out of your experience with R in this course really will be a function of what you put into learning it. With that in mind, I'd recommend that everyone who is serious about doing well on the course spend at least a little time familiarising yourself with R in advance.

You may therefore find the following resources useful:

- `https://www.datacamp.com/courses/free-introduction-to-r`
- `http://cran.r-project.org/doc/manuals/R-intro.pdf`

5 Schedule

The general schedule for the course is as follows. Details on topics covered and the readings for each week are provided on the following pages. The exact reading recommendations each week may change throughout the term, particularly if there are interesting new papers to read.

Note: There will be no lecture or class during reading week (November 5 – 11).

5.1 Causal Inference and Potential Outcomes (Oct 2)

In the first lecture we will introduce the topic of causal inference. We will outline a specific definition of causality using the potential outcomes framework, and will describe the fundamental problem of causal inference. We will highlight the persistent threat of selection bias in observational data and we will discuss differences between statistical inference and causal inference.

Required reading:

- Angrist & Pischke Ch. 1 and Ch. 2 p. 11-15
- Geber and Green Ch. 2
- Paul W Holland. Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(3), 1986.

Recommended reading:

- David A Freedman. Statistical models and shoe leather. *Sociological methodology*, pages 291–313, 1991.
- Stephen L Morgan and Christopher Winship. *Counterfactuals and causal inference.* Cambridge University Press, 2015. p. 1 – 24

5.2 Randomised Experiments (Oct 9)

In this lecture we will review the logic that underpins a research design that has become a mainstay of political science research: randomised experiments. We will focus on why randomisation is such a powerful force for making causal inferences (spoiler: internal validity), and will discuss the trade-offs implicit in experimental research (spoiler: external validity). In learning how to analyse experimental data, we will review the t-test and also cover regression as a tool for analysing experiments.

Required reading:

- Ch. 2 from Angrist & Pischke
- Ch. 2 and 3 from Gerber & Green
- Alan S Gerber, Donald P Green, and Christopher W Larimer. Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1):33–48, 2008.

Recommended reading:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect*. Princeton University Press, 2014. Chapter 1
- Raghabendra Chattopadhyay and Esther Duflo. Women as policy makers: Evidence from a randomized policy experiment in india. *Econometrica*, 72(5):1409–1443, 2004.
- Joshua L Kalla and David E Broockman. Campaign contributions facilitate access to congressional officials: A randomized field experiment. *American Journal of Political Science*, 60(3):545–558, 2016.
- Laura Haynes, Ben Goldacre, David Torgerson, et al. Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials— Cabinet Office. 2012.
- Jason Barabas and Jennifer Jerit. Are survey experiments externally valid? *American Political Science Review*, 104(2):226–242, 2010.
- Alexander Coppock, Thomas J. Leeper, and Kevin J. Mullinix. Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 2018.

5.3 Selection on Observables, part 1 (Oct 16)

Randomisation is a powerful tool because it means that confounders can be safely ignored by researchers as, in expectation, they will be balanced across treatment and control groups. Sadly, some of the most interesting social science questions cannot be addressed using randomised experiments (Why? First, because experiments are costly, and second, because it would be bad form to randomly assign, for instance, the institutions that govern a country's electoral system, or whether you get a distinction in your degree). When it is not possible to randomise, how can we make valid causal inferences? In the next two lectures, we discuss methods for non-experimental data which assume that selection into treatment groups is based on observable factors. This week we focus on subclassification and matching.

Required reading:

- Ch. 3 from Angrist & Pischke (we will use material from chapter 3 both this week and next week)
- Elizabeth A Stuart. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1, 2010.
- Andrew C Eggers and Jens Hainmueller. MPs for sale? Returns to office in postwar British politics. *American Political Science Review*, 103(4):513–533, 2009.

Recommended reading:

- William G Cochran. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, pages 295–313, 1968.
- Gilligan, Michael J and Sergenti, Ernest J. Do UN interventions cause peace? Using matching to improve causal inference. *Quarterly Journal of Political Science*, 3(2):89–122, 2008.
- Rajeev H Dehejia and Sadek Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448):1053–1062, 1999.
- Donald B Rubin. For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, pages 808–840, 2008.
- Stephen L Morgan and Christopher Winship. *Counterfactuals and causal inference.* Cambridge University Press, 2015. Chapter 5

5.4 Selection on Observables, part 2 (Oct 23)

This week we continue to consider methods that rely on a selection-on-observables assumption for making causal inferences from non-experimental data. We focus on regression and the propensity score.

Required reading:

- Ch. 3 from Angrist & Pischke

Recommended reading:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect*. Princeton University Press, 2014. Chapter 2
- Stephen L Morgan and Christopher Winship. *Counterfactuals and causal inference*. Cambridge University Press, 2015. Chapter 6
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- Daniel E Ho, Kosuke Imai, Gary King, and Elizabeth A Stuart. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15(3):199–236, 2007.
- Aronow, Peter M and Samii, Cyrus. Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1):250–267, 2016.

5.5 Panel Data and Difference-in-Differences (Oct 30)

When we can observe and measure potentially confounding factors, we can recover causal effects by controlling for these factors. Often, however, confounders may be difficult to measure or impossible to observe. If this is the case, we need alternative strategies for estimating causal effects. One approach is to try to obtain data with a time dimension, where one group receives a treatment at a given point in time but the other group does not. Comparing the differences between pre- and post-treatment periods for these two groups allows us to control for unobserved omitted variables that are fixed over time. Under certain assumptions, this can produce valid estimates of causal effects.

Required reading:

- Ch. 5 from Angrist & Pischke

Recommended reading:

- David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania. Technical report, National Bureau of Economic Research, 1993.
- Jonathan McDonald Ladd and Gabriel S Lenz. Exploiting a rare communication shift to document the persuasive power of the news media. *American Journal of Political Science*, 53(2):394–410, 2009.
- David Card. The impact of the mariel boatlift on the miami labor market. *ILR Review*, 43(2):245–257, 1990.
- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect*. Princeton University Press, 2014. Chapter 5
- Elias Dinas, Konstantinos Matakos, Dimitrios Xefteris, and Dominik Hangartner. Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-right Parties? *Political Analysis*, Forthcoming. Available here

5.6 Synthetic Control (Nov 13)

Don't have a good control unit to use in a difference-in-differences design? Don't panic; just synthesise one. Synthetic control approaches allow for causal inferences based on similar assumptions to difference-in-differences, but are particularly well suited for situations in which the treatment occurs for a single unit. By providing a systematic way to choose comparison units, synthetic control is a good method for application to comparative case studies.

Required reading:

- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505, 2010.

Recommended reading:

- Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017.
- David Hope. Estimating the effect of the EMU on current account balances: A synthetic control approach. *European Journal of Political Economy*, 44:20–40, 2016.
- Benjamin Born, Gernot J Müller, Moritz Schularick, and Petr Sedlacek. The Costs of Economic Nationalism: Evidence from the Brexit Experiment. *Working Paper*, 2018. – Available here.

5.7 Instrumental Variables, part 1 (November 20)

Aside from experiments, all of the strategies covered up to this point rely on the researcher being able to control for confounding factors when estimating causal effects. For the next two weeks, we focus on a strategy – instrumental variables – which can be used to address unobserved confounding factors in the context of cross-sectional data (i.e. when we can't use the panel data methods discussed in previous weeks). This week, we will motivate instrumental variable (IV) methods, by discussing how this strategy can be useful in the context of experimental data where some units fail to comply with the treatment.

Required reading:

- Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.
- Ch. 4 from Angrist & Pischke
- Ch. 5 & 6 from Gerber and Green

Recommended reading:

- Allison J Sovey and Donald P Green. Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, 55(1):188–200, 2011.

5.8 Instrumental Variables, part 2 (November 27)

In this lecture we focus on the logic of instrumental variables in the context of observational studies. We will also discuss a number of applied examples that use an IV strategy, paying attention to how they work, and how they can go wrong.

Required reading:

- Ch. 4 from Angrist & Pischke
- Ch. 5 & 6 from Gerber and Green

Recommended reading:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect*. Princeton University Press, 2014.Chapter 3
- Holger Lutz Kern and Jens Hainmueller. Opium for the masses: How foreign media can stabilize authoritarian regimes. *Political Analysis*, 17(4):377–399.
- Andreas Madestam, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. Do political protests matter? Evidence from the tea party movement. *The Quarterly Journal of Economics*, 128(4):1633–1685, 2013.
- Daron Acemoglu, Simon Johnson, and James A Robinson. The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401, 2001.
- Elias Dinas, Konstantinos Matakos, Dimitrios Xefteris, and Dominik Hangartner. Waking Up the Golden Dawn: Does Exposure to the Refugee Crisis Increase Support for Extreme-right Parties? *Political Analysis*, Forthcoming. Available here
- Steven D Levitt. Using electoral cycles in police hiring to estimate the effect of policeon crime. *American Economic Review*, 87(3):270–290, 1997.

5.9 Regression Discontinuity Designs (December 4)

A regression discontinuity design (RDD, for short) arises when the selection of a unit into a treatment group depends on a covariate score that creates some discontinuity in the probability of receiving the treatment. In this lecture we will consider both 'sharp' and 'fuzzy' RDDs.

Required reading:

- Ch. 6 from Angrist & Pischke

Recommended reading:

- Joshua D Angrist and Jörn-Steffen Pischke. *Mastering 'metrics: The path from cause to effect*. Princeton University Press, 2014.Chapter 4
- Andrew B. Hall. What happens when extremists win primaries? *American Political Science Review*, 109(1):18–42, 2015.
- Andrew C Eggers and Jens Hainmueller. MPs for sale? Returns to office in postwar British politics. *American Political Science Review*, 103(4):513–533, 2009.
- Charlotte Cavaillé and John Marshall. Education and anti-immigration attitudes: Evidence from compulsory schooling reforms across western europe. *American Political Science Review*, Forthcoming.
- Erik Meyersson. Islamic rule and the empowerment of the poor and pious. *Econometrica*, 82(1):229–269, 2014.

5.10 Overview and Review (December 11)

We end with a schematic overview of the course. We will discuss the extent to which the methods we cover on the course sacrifice external validity at the expense of internal validity, and whether this matters. We will also have a Q and A on the final coursework.

Recommended reading:

- Cyrus Samii. Causal empiricism in quantitative research. *The Journal of Politics*, 78(3):941–955, 2016.
- Christopher J Ruhm. Shackling the identification police? Working Paper 25320, National Bureau of Economic Research, November 2018.
- Kosuke Imai, Luke Keele, Dustin Tingley, and Teppei Yamamoto. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, 105(4):765–789, 2011.