Overview

This course provides a current perspective on estimating causal effects in social science research. The approach is rooted in non-parametric and semi-parametric frequentist methods. We will emphasize research design, causal identification, and robust estimation and inference.

Prerequisites

The course requires that students have working knowledge of probability theory, matrix algebra, and calculus at the level of POL-GA 1250, “Quant I.” The course is intended to provide foundational methodological training to PhD students in politics as part of their required sequence of courses toward their degree. Only PhD students in Politics will be allowed to take the course for a grade. (We do not have adequate teaching assistant and other resources to service students from other departments taking this for a grade, unfortunately.) Students from other programs may audit if space permits.

Texts

The course will draw mostly from the following two textbooks:


The course will mostly follow MHE, using CCI to provide more intuitive background and illustrations. I will also supplement the textbooks with notes, sections from other textbooks, and journal articles. I have listed “further reading” for each topic, and my lectures will sometimes draw on these. Readings will be available in a public Dropbox (see course website).
The books are each detailed, up to date, and they complement each other well. The books are relatively inexpensive, so students are recommended to acquire both. MHE can be mathematically difficult at times, but you are strongly encouraged to dive in, replicate proofs, and work hard to understand it. CCI provides intuition to keep you sane and grounded.

While these two textbooks are technically sound, they do gloss over details on estimation mechanics. An excellent textbook that provides the proper statistical background for this course is the following:


It is freely available on Hansen’s website, and there is also a PDF in our readings dropbox.

For those of you that feel like you need to refresh/improve your mathematical chops, the following is a nice workbook style development of rudiments for regression and statistical inference, with a critical eye:


My dear favorite graduate-level intro textbook, and the one that I used in ye olden days, is the following:


Goldberger develops the theory of “agnostic” linear regression step-by-step from first principles. It’s a classic, but out of print (you can get cheap used versions on Amazon though). An updated take is given in the following:


You can get it here: http://aronow.research.yale.edu/aronowmiller.pdf

The following textbooks provide some deeper theoretical insights on statistics for causal inference and I will sometimes reference them:


Software

You will have the choice to work in R or Stata, however you wish. Each program has its advantages, and so it is useful to obtain fluency in both. R is great for programming estimators and algorithms “from scratch,” programming simulations, and making graphics. Some assignments will ask you to do that. However many of the packaged estimation routines are flawed and it does not handle very large datasets well. Stata’s pre-programmed estimation routines are more reliable, and some assignments could be done using them; however Stata is very clumsy (for me) for programming, simulations, or graphics. Use whatever you want to get the tasks done in a manner that is sound and reproducible.
Requirements and policies

Homework
You will receive homework about every week. You will have to submit your completed assignment by the following week; exact deadlines will be made clear on the assignment. You can work with others, but to receive credit, your homework must comply with the following guidelines:

- You must turn in a hard copy of your own homework by the stated deadline.
- The assignment that you turn in must clearly reflect your own thinking. Sets of verbatim copies of homework will not receive any credit.
- Homework assignments may be hand written or typed, but they must be clearly legible.
- Estimates obtained from data analysis programs (e.g., Stata or R) must be formatted properly into tables resembling journal presentation styles. You should use a table formatting function (e.g., outreg or esttab in Stata, or apsrtab or stargazer in R). Use a reasonable (2 or at most 3) number of digits after decimal points, report standard errors along with coefficients, clarify what are the dependent variables in each table, and explain in footnotes to your tables what kinds of estimators or adjustments have been used. Print outs of “raw” screen output or commented logs will not receive any credit. However, you may include such output as an appendix so that the grader can troubleshoot.
- Mathematical derivations should include all key steps with explanations for important techniques.

Homework will be graded for points as indicated on each assignment and count toward 50% of your grade.

Mid-term exam
An in-class mid-term exam is provisionally scheduled during class on March 9 (to be confirmed). The mid-term serves the purpose of evaluating individual progress, which in turn helps me to understand where to place emphasis for the remainder of the semester. If you are unable to make it to the exam, you must provide notice at least a week prior so that we can arrange an alternative time. The mid-term will count toward 15% of your grade.

Final exam
An in-class final exam will be scheduled during the final examination period. The final also serves the purpose of evaluating individual progress, which in turn allows me to provide individualized recommendations on where students should apply effort to strengthen their methodological foundations. If you are unable to make it to the exam, you must provide notice at least a week prior so that we can arrange an alternative time. The final will count toward 25% of your grade.

Attendance and participation
Attendance and participation in class discussions is required and counts toward 10% of your grade.
Special needs
Students with special needs should come to office hours or schedule an appointment with the instructor to discuss possible accommodation.

Topics
Topics listed below will be covered in around 1-2 class sessions each. Required reading sometimes corresponds directly to material covered in the sessions and sometimes builds up background needed for future sessions. Most of the required reading comes from MHE and CCI, although the topics covered toward the end of the semester will draw on other texts that will be made available.

1 Identification, estimation, and inference


Required reading: MHE Ch. 1-2; CCI Ch. 1-3.

Further reading: Angrist and Krueger (1999); Angrist and Pischke (2010); DiNardo and Lee (2011); Freedman (1991); Heckman and Vytlacil (2007); Holland (1986); Imbens and Wooldridge (2009); Imbens and Rubin (2011, Ch. 1-3); Manski (1995, Ch. 1); Pearl (2009, Ch. 1); Rothman (1976); Rosenbaum (1999); Rosenbaum (2002, Ch. 1-3); Rubin (1974); Rubin (1978); Rubin (1986).

2 Regression mechanics

Core decompositions and their relation to regression; Frisch-Waugh-Lovell; conditional independence/exogeneity; omitted variable bias formula; effect heterogeneity and nonlinearity; regression and matching; regression and simple contrasts; influence and leverage; testing restrictions on coefficients.

Required reading: MHE Ch. 3; CCI Ch. 4,6; Aronow and Samii (2016).

Further reading: Angrist and Krueger (1999); Freedman (2008a); Freedman (2008b); Imbens and Rubin (2011, Ch. 7); Lin (2013); Schochet (2010).

3 Notions of bias

Confounding; post-treatment bias; sample selection bias; aggregation/effect heterogeneity biases; misspecification/extrapolation bias and model dependence; measurement error biases.

Required reading: CCI Ch. 8; Bound et al. (2000, pp. 1-39); Heckman (1979); Lalonde (1986); Rosenbaum (1984); Samii (2016).
Further reading: Frangakis and Rubin (2002); Heckman et al. (1998); Hyslop and Imbens (2001); Imai et al. (2008); King and Zeng (2006); Pearl (2009, Ch. 3, 6).

4 Control and balance via matching and weighting

Identification under conditional exogeneity; alternative matching and weighting algorithms; estimation and inference after matching.

Required reading: CCI Ch. 5,7.

Further reading: Abadie and Imbens (2006); Abadie and Imbens (2008); Abadie and Imbens (2011); Busso et al. (2014); Dehejia and Wahba (2002); De Luna et al. (2011); Hainmueller (2011); Hirano and Imbens (2004); Ho et al. (2007); Iacus et al. (2011); Imai and van Dyk (2004); Imbens (2000); King and Nielsen (2016); Lu et al. (2001); Rosenbaum and Rubin (1983); Sekhon (2009); Todd (2008).

5 Robust statistical inference

Clustering, autocorrelation, and spatial dependence; Moulton’s problem; heteroskedasticity and cluster robust standard errors; bootstrapping; estimating the exact randomization variance; permutation tests.

Required reading: MHE Ch. 8; Young (2015a).

Further reading: Aronow et al. (2015); Barrios et al. (2012); Bell and McCaffrey (2002); Bertrand et al. (2004); Cameron et al. (2008); Cameron et al. (2009); Chung and Romano (2013); Conley (1999); Efron and Tibshirani (1993); Freedman (2009, Ch. 8); Hansen (2007); Imbens and Kolesar (2012); Janssen (1997); Moulton (1986); Romano (1990); Samii and Aronow (2012); Young (2015b).

6 Instrumental variables

Exclusion restriction; valid first stage; principal strata; local average treatment effect (LATE); weak instrument; sensitivity analysis.

Required reading: MHE Ch. 4; CCI Ch. 9.

Further reading: Abadie (2003); Angrist et al. (2000); Angrist et al. (1996); Baum et al. (2003); Bazzi and Clemens (2013); Bound et al. (1995); Conley et al. (2010); Deaton (2010); Heckman and Urzua (2009); Imbens (2010); Imbens and Rosenbaum (2005); Kolesar et al. (2011); Sovey and Green (2011); Staiger and Stock (1997); Stock et al. (2002).

7 Repeated observations

Adjusting for unobserved heterogeneity via fixed effects and difference-in-differences; synthetic control; event studies.

Required reading: MHE Ch. 5; CCI Ch. 11; Abadie and Gardeazabal (2003).
Further reading: Abadie (2005); Abadie et al. (2010); Athey and Imbens (2006); Beck and Katz (2001); Bound and Solon (1999); DellaVigna and La Ferrara (2010); Dube et al. (2011); Fisman (2001); Green et al. (2001); Imai and Kim (2012); Imbens and Rubin (2011, Ch. 31, 33); Mora and Reggio (2013).

8 Regression discontinuity (RD)

Forcing variables; sharp and fuzzy RD; conditional average treatment effect (CATE); local linearity, bandwidth, and non-parametric regression; kernel weighting; multiway discontinuities; checks for sorting around cut-points; endogenous forcing variables; measurement error in forcing variables.

Required reading: MHE Ch. 6; CCI Ch. 11.

Further reading: Caughy and Sekhon (2011); Froelich (2007); Green et al. (2009); Imbens and Kalyanaraman (2009); Imbens and Lemieux (2008); Lee and Card (2008); Lee and Lemieux (2010); McCrary (2008); Papay et al. (2011); Urquiola and Verhoogen (2009).

9 Moderators, mediators, and causal explanation

Moderators and effect heterogeneity; mediators and mechanisms; sequential ignorability.

Required reading: Acharya et al. (2016); CCI Ch. 10; Angrist et al. (2011); Bullock et al. (2010); Imai et al. (2011).

Further reading: Glynn (2011); Heckman et al. (1997); Jo et al. (2011); Ludwig et al. (2011); VanderWeele (2008); VanderWeele (2015).

10 Distributional effects and quantile regression

Quantile treatment effect; minimum absolute deviations; rank invariance.

Required reading: MHE Ch. 7.

Further reading: Bitler et al. (2006); Chernozhukov and Hansen (2005); Heckman et al. (1997); Koenker and Hallock (2000).

11 Multiple endpoints

Index and mean effects; multiple comparisons adjustments.

Required reading: Anderson (2008); Caughy et al. (2015); Romano and Wolf (2007).

Further reading: Casey et al. (2011); Clingingsmith et al. (2009); Farcomeni (2008); Gibson et al. (2011); Kling and Liebman (2004); O’Brien (1984); Shaffer (1995).
12 **Missing data and attrition**

Bounds; inverse probability weighting; imputation.

*Required reading*: CCI Ch. 12; Gerber and Green (2012, Ch. 7); Horton and Kleinman (2007); King et al. (2001); Lee (2009); Manski (1995, Ch. 2); Vansteelandt et al. (2010).

*Further reading*: Aronow et al. (2015); Jones (1996); Puma et al. (2009); Samii (2011).

13 **Limited dependent variable effects**

Structural versus causal estimands.


*Further reading*: Davidson and MacKinnon (2004, Ch. 10-11) Fox (2002); Freedman (2006); Greene (2004); Hubbard et al. (2010); Imbens and Rubin (2011, Ch. 8); Liang and Zeger (1986); Van der Laan and Rose (2011, Ch. 7, 11, 16-17); Wooldridge (2002, Ch. 15, 19-20).

14 **Machine learning and causal inference**

Data-driven estimation; ensemble methods; machine learning for implementing CIA, characterizing effect heterogeneity, and discovering instruments.

*Required reading*: Athey and Imbens (2015); Belloni et al. (2014); Imai and Strauss (2011); Pearl (2009, Ch. 2); Samii et al. (2015).

*Further reading*: Green and Kern (2012); Imai and Ratkovic (2012); Kleinberg et al. (2015); Van der Laan and Rose (2011); Wager and Athey (2015).

15 **Generalization and external validity**

Unconfounded location; external validity; generalizability; prediction error.

*Required reading*: Dehejia et al. (2015); Gechter (2015); Hotz et al. (2005); Imbens (2010).


16 **Structure and identification**

Policy effects versus structural parameters; spill-over and equilibrium effects.

*Required reading*: Acemoglu (2010); Angrist and Pischke (2010); Heckman (2010).

*Further reading*: Acemoglu et al. (2015); Aronow and Samii (2015); Chetty (2009); Rosenzweig and Wolpin (2000); Todd and Wolpin (2006); Wolpin (2013).
References


Aronow, P. M., A. Coppock, A. S. Gerber, D. P. Green, and H. L. Kern (2015). Combining double sampling and bounds to address non-ignorable missing outcomes in randomized experiments. Manuscript, Yale University, Columbia University, and Florida State University.


