Public Policy 571: Applied Econometrics
Fall 2020 Syllabus

Instructor: Jonathan Hanson
jkhanson@umich.edu
4223 Weill Hall, 615-1496
Office Hours: Tues. 12:00–1:30 (Zoom), Fri. 11:30–1:00 (Zoom), or by appointment

The objective of this course is to help gain students proficiency in advanced techniques that are employed in the service of causal inference, including instrumental variables, selection models, regression discontinuity, and matching methods. Students will understand how to design, execute, and interpret the results from advanced estimation techniques beyond the standard OLS model. Students will learn to work with data structures in which the observations are not independently and identically distributed: time-series data, panel data, time-series-cross-sectional data, multi-level data, etc.

Prerequisite: it is assumed that students in this course will have previously taken Public Policy 529 and Public Policy 639 (or equivalent coursework).

Class Meeting Schedule

Unless otherwise noted, lectures are Mondays and Wednesdays from 10:00–11:20 pm in 1110 Weill Hall as well as via Zoom at `https://umich.zoom.us/j/98921999920`.

Textbooks

There is no single textbook for this course, and readings will be available on the Canvas website. You can log into Canvas at `http://canvas.umich.edu`. The following books and articles, however, will be very good references for course material.

- Joshua D. Angrist and Jörn-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.
- J. Scott Long. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications.
- Paul J. Gertier, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2016. *Impact Evaluation in Practice*, 2nd ed. Washington, DC: World Bank. `https://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-0779-4`.
- Alberto Abadie and Matias D. Cattaneo. 2018. "Econometric Methods for Program Evaluation." *Annual Review of Economics* 10: 465–503. `https://economics.mit.edu/files/14922`.

Assignments and Grading

Your grade for this course will be determined by the following:

Problem sets	50%
Midterm exam	25%
Final exam	25%

There will be six major problem sets in this course. Additionally, there will be two exams – one midterm and one final – covering the first-half and second halves of the course respectively. Each exam will have two components: a take-home part and an in-class part.

You are encouraged to help each other figure out the answers to the problem sets, but it is expected that you write up your answers independently. The take-home exams are exams. You are expected to work on these without the assistance of any other person. You may, however, consult any textbook or internet-resources.

Assignment and Exam Calendar

Problem Set 1	September 21
Problem Set 2	October 5
Problem Set 3	October 12
Midterm Exam	October 19
Problem Set 4	November 2
Problem Set 5	November 16
Problem Set 6	December 7
Final Exam	December 17

The final course letter grade reflects the Ford School's guidelines. An A is awarded for work that is Excellent, an A- for work that is Very Good, a B+ for work that is Good, a B for work that is Acceptable, and a B- for work that is below expectations for graduate work. You should know I do not have a predetermined formula to convert numeric point totals into these categories. It would be a mistake, for instance, to assume that a grade of 75% on an exam translates into a C, since exams vary in their difficulty.

Software

Students can use either R or Stata statistical software to complete homework assignments. These applications are available on computers in the Ford School computer lab, as well as the larger computer labs on campus. Additionally, students can remotely log in to the university's Virtual Sites (see information at `https://documentation.its.umich.edu/node/312`) to access Stata when not on campus. Windows users can use the AppsAnywhere system (`https://its.umich.edu/computing/computers-software/campus-computing-sites/appsanywhere`) to run Stata on their local machine, which is much faster.

R is an open-source program that is freely downloadable from `https://cran.r-project.org`. Students who use R are strongly encouraged to download the free RStudio Desktop companion application (https://rstudio.com/products/rstudio/download/) to serve as their interface with R.

There are several resources for learning Stata available on Canvas, including a handbook that I compiled for Public Policy 567. If you wish to purchase a book, consider the following:

- Alan C. Acock. 2016. *A Gentle Introduction to Stata*, 5th edition, Stata Press.

Some other Stata and R resources:

- Econometrics Academy: https://tinyurl.com/y56tj9wl.
- J-PAL Stata resources: https://www.povertyactionlab.org/research-resources.

Academic Integrity

It is expected that students are familiar with the Ford School's expectations for academic integrity as described at http://fordschool.umich.edu/academics/expectations, which adhere to the academic integrity policies for Rackham Graduate School. Violations of these policies will be taken seriously.

Students with special needs

If you believe you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities to help us determine appropriate accommodations. I will treat any information you provide as private and confidential.

Student Mental Health and Wellbeing

The University of Michigan is committed to advancing the mental health and wellbeing of its students. We acknowledge that a variety of issues, such as strained relationships, increased anxiety, alcohol/drug problems, and depression, directly impacts students' academic performance. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) and/or University Health Service (UHS). For a listing of other mental health resources available on and off campus, visit: http://umich.edu/~mhealth/.

Inclusivity

Members of the Ford School community represent a rich variety of backgrounds and perspectives. We are committed to providing an atmosphere for learning that respects diversity. While working together to build this community we ask all members to:

- share their unique experiences, values and beliefs
- be open to the views of others

- honor the uniqueness of their colleagues
- appreciate the opportunity that we have to learn from each other in this community
- value one another's opinions and communicate in a respectful manner
- keep confidential discussions that the community has of a personal (or professional) nature
- use this opportunity together to discuss ways in which we can create an inclusive environment in Ford classes and across the UM community

Please refer to http://fordschool.umich.edu/academics/expectations for a full statement on the Ford School's academic expectations.

COVID-19 Statement

In order to participate in any in-person aspects of this course, including meeting with other students to study or work on a team project, you must follow all safety measures mandated by the State of Michigan, the University of Michigan and the Ford School. This includes maintaining physical distancing of six feet from others and properly wearing a face covering at all times while on campus. In addition, it is expected that you will protect and enhance the health of everyone in the Ford School community by staying home and following self-isolation guidelines if you are experiencing any symptoms of COVID-19, have been exposed to someone with COVID-19, are awaiting a test result, or have engaged in a higher-exposure activity such a flying or attending an indoor social gathering of more than 10 people. If you are unable or unwilling to adhere to all prescribed safety measures, you will be accommodated through remote access to all aspects of this course. Additional information on public health safety measures is described in the Wolverine Culture of Care and the University's Face Covering Policy for COVID-19.

Advanced Estimation Methods for Cross-Sectional Data

August 31: Principles of Maximum Likelihood Estimation

In the classical linear regression model and its variants, we find the parameters of interest (i.e. the regression coefficients) though analytic formulas. With Maximum Likelihood Estimation (MLE), these parameters are often found by a search algorithm which tours the parameter space and finds the values of the parameters that are "most likely" given the data. The power of MLE its flexibility. Many different functional forms are possible, facilitating analysis of many kinds of data for which linear regression is not suitable.

- M S Sasidhar. 2018. "Introduction to Maximum Likelihood Estimation." https://tinyurl.com/y2hb45vv
- Sherry Towers. "Maximum Likelihood Estimation." http://www.sherrytowers.com/mle_introduction.pdf.
- Online simulation by Shawna Metzger: https://mybinder.org/v2/gh/MetzgerSK/shinyElement/major?urlpath=shiny/mleLogit/.

September 2 & 9: Dichotomous Dependent Variables

When our dependent variables are dichotomous, OLS might be problematic. For one thing, OLS can in some cases produce predicted values of the dependent variable that fall outside the range the 0-1 range. Other problems arise with non-normality of disturbances and non-linear effects of independent variables. With Maximum Likelihood Estimation, we can use functional forms that are designed for these data: probit and logit models.

- J. Scott Long. 1997. "Binary Outcomes: The Linear Probability, Probit, and Logit Models" In *Regression Models for Categorical and Limited Dependent Variables*, chapter 3.
- Sheheryar Banuri, Stefan Dercon, and Varun Gauri. 2019. "Biased Policy Professionals." *The World Bank Economic Review* 33(2): 310–327.
- Steven B. Caudill. 1988. "An Advantage of the Linear Probability Model over Probit or Logit." *Oxford Bulletin of Economics and Statistics* 50(4): 425–427.

September 14: Ordinal Categories as Dependent Variables

When the dependent variable consists of ordinal categories, the list of possible problems grows. We can use ordered probit and ordered logit models for this kind data. A key objective is to learn how to interpret and convey the results.

- J. Scott Long. 1997. "Ordinal Outcomes: Ordered Logit and Ordered Probit Analysis." In *Regression Models for Categorical and Limited Dependent Variables*, chapter 5.
- Zijing Lin and Wei "David" Fan. 2020. "Bicycle Ridership Using Crowdsourced Data: Ordered Probit Model Approach." *Journal of Transportation Engineering, Part A: Systems* 146(8): 1–15.

September 16: Nominal Categories as Dependent Variables

With unordered categories, estimation becomes trickier since we cannot use a latent dimension as an underlying framework. We can use multinomial logit models in these circumstances.

- J. Scott Long. 1997. "Nominal Outcomes: Multinomial Logit and Related Models." In *Regression Models for Categorical and Limited Dependent Variables*, chapter 6.
- M. Niaz Asadullah. 2018. "Madrasah for Girls and Private School for Boys? The Determinants of School Type Choice in Rural and Urban Indonesia." *International Journal of Educational Development* 62: 96–111.

September 21 & 23: Count Models

Count models are for cases in which the dependent variable is a count of the number of times something occurs.

- Stefany Coxe, Stephen G. West, and Leona S. Aiken. 2009. "The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives." *Journal of Personality Assessment* 91(2): 121–136.
- Rainer Winkelmann (2015). "Counting on Count Data Models." IZA World of Labor 2015: 148.
- Paul Kwame Nkegbe and Naasengnibe Kuunibe. 2019. "Poverty and Malaria Morbidity: A Study Using Count Regression Model." SAGE Research Methods Cases in Business and Management.

Causal Inference for Cross-Sectional Data

September 28 & 30: Selection Models

The purpose of selection models is to address situations in which the cases that make it into the sample are different in important, unmeasured ways from those who do not, and these unmeasured factors are relevant for predicting the dependent variable.

- Christopher Achen. 1986. "Quasi-Experiments with Censored Data: Why Regression and Weighting Fail." In *Statistical Analysis of Quasi-Experiments*, chapters 4 and 5. Berkeley: University of California Press.
- Claire Infante-Rivard and Alexandre Cusson. 2018. "Reflection on Modern Methods: Selection Bias – A Review of Recent Developments." *International Journal of Epidemiology* 47(5): 1714–1722.
- Shawn Bushway, Brian D. Johnson, and Lee Ann Slocum. 2007. "Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology." *Journal of Quantitative Criminology* 23: 151–178.
- Manh Hung Do and Sang Chul Park. 2019. "Impacts of Vietnam's New Rural Development Policy on Rural Households' Income: Empirical Evidence from the Heckman Selection Model." *International Review of Public Administration* 24(4): 229–245.

October 5 & 7: Multivariate Matching Methods

With observational data, we face inherent challenges in estimating the average treatment effect (ATE) of some policy intervention because the treatment is not randomly-assigned. The treatment group and control group may differ from each other, in the aggregate, on both observable and unobservable characteristics in ways other than the treatment, and we cannot adequately control for these other factors. Matching methods are intended to bring balance to the treatment groups on these other characteristics to reduce bias in estimating the ATE. In this section of the course we will learn various matching methods and develop awareness of their limitations.

- C. Lockwood Reynolds and Stephen DesJardins. 2009. "The Use of Matching Methods in Higher Education Research." In *Higher Education: Handbook of Theory and Research*, ed. John C. Smart. Springer.

- Elizabeth Stuart. 2010. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 25(1): 1–21.
- Paul R. Rosenbaum. 2010. "Basic Tools of Multivariate Matching." *Design of Observational Studies*, chapter 8. New York: Springer.
- Jonathan Bartlett. 2016. "Why You Shouldn't Use Propensity Score Matching." The Stats Geek blog.
- Alexander Pfaff, Juan Robalino, Eirivelthon Lima, Catalina Sandoval, and Luis Diego Herrera. 2014. "Governance, Location, and Avoided Deforestation from Protected Areas: Greater Restrictions Can Have Lower Impact, Due to Differences in Location." *World Development* 55: 7–20.

October 12 & 14: Regression Discontinuity Models

Regression discontinuity models are useful for scenarios in which assignment into the treatment group is based upon a cutoff score on some observable characteristic or a particular date of implementation. In short, the difference between the predicted outcomes on either side of the cutoff point becomes a way to estimate the size of the treatment effect. This section of the course explores various uses of these models.

- Joshua D. Angrist and Jörn-Steffen Pischke. "Regression Discontinuity Designs." In *Mastering 'Metrics: The Path from Cause to Effect*, chapter 4. Princeton: Princeton University Press. (this may be a refresher for you)
- Matias D. Cattaneo, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2019. "The Regression Discontinuity Design." Manuscript.
- David S. Lee. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142: 675–697.
- Analisa Packham and Brittany Street. 2018. "The Effects of Physical Education on Student Fitness, Achievement, and Behavior." Manuscript.
- Pei Zhu. 2019. "Using a Regression Discontinuity Design for Evaluation Studies." MDRC.

October 19: Mid-term Exam

Estimation Methods for Time-Series and Panel Data

October 21 & 26: Regression with Time-Series Data

The standard assumption is that our data are independent and identically distributed. When we have time-series data, such as the results of monthly polls on presidential approval or some other case of repeated observations of the same object, this assumption is violated. The stochastic component of one observation may be correlated with the one preceding it and the one that follows. We spend about two sessions on this topic.

- Aptech Data Analytics Blog. 2020. “Introduction to the Fundamentals of Time Series Data and Analysis.” https://tinyurl.com/vzud6yj.
- Jon C. Pevehouse and Jason D. Brozek. 2008. “Time-Series Analysis.” In *The Oxford Handbook of Political Methodology*, chapter 19.
- Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt, and Jon C. W. Pevehouse. 2014. “Dynamic Regression Models.” In *Time Series Analysis for the Social Sciences*, chap. 3. Cambridge University Press.

October 28: Multi-Level Models

When our data consist of individual cases that are embedded within higher-level units – such as school children inside classrooms, which are inside schools – we need methods that can estimate both individual and unit-level effects. In this section of the course, we explore the use of multilevel models for this purpose.

- Ana V. Diez-Roux. 2000. “Multilevel Analysis in Public Health Research.” *Annual Review of Public Health* 21: 171–192.
- Michael Freeman. “An Introduction to Hierarchical Modeling.” Visualization available at http://mfviz.com/hierarchical-models/.
- Tim Smith and Gerald Shively. 2019. “Multilevel analysis of individual, household, and community factors influencing child growth in Nepal.” *BMC Pediatrics* 19(91): 1–14.

November 2 & 4: Regression with Panel Data

We have a panel when our data contain repeated observations of a sample of objects, such as a set of individuals who are surveyed periodically or a time series of cross-national data. In this scenario, we need to think about the non-independence of our observations both across time and space.

- Aptech Data Analytics Blog. 2020. “Introduction to the Fundamentals of Panel Data.” https://tinyurl.com/y4gt52db. (short intro)
- Tae Ho Eom, Sock Hwan Lee, and Hua Xu. 2007. “Introduction to Panel Data Analysis. Concepts and Practices.” In *Handbook of Research Methods in Public Administration*, Gerald R. Miller and Kaifeng Yang, eds. CRC Press.
- Robert Kubinec. 2020. “What Panel Data is Really All About.” http://www.robertkubinec.com/post/fixed_effects/.
- Nathaniel Beck and Jonathan N. Katz. 2011. “Modeling Dynamics in Time-Series-Cross-Section Political Economy Data.” *Annual Review of Political Science* 14: 331–52.
- Elisabeta Jaba, Christiana Brigitte Balan, and Iaon-Bogan Robu. 2014. “The Relationship Between Life Expectancy at Birth and Health Expenditures Estimated by a Cross-Country and Time-Series Analysis.” *Procedia Economics and Finance* 15: 108–114.

November 9: Duration Models

Duration models, also known as survival models, deal with situations in which we model the amount of time that some phenomenon lasts as a function of independent variables.

- Bradford S. Jones and Regina P. Branton. 2005. "Beyond Logit and Probit: Cox Duration Models of Single, Repeating, and Competing Events for State Policy Adoption." *State Politics and Policy Quarterly* 5(4): 420–443.
- "Time-To-Event Data Analysis." Columbia University Mailman School of Public Health. Web resource.
- Online simulation by Shawna Metzger: https://mybinder.org/v2/gh/MetzgerSK/shinyElement/major?urlpath=shiny/whySurv/.
- Ted Aranki and Corrado Macchiarelli. 2013. "Employment Duration and Shifts into Retirement in the EU." Working Paper Series No. 1517, European Central Bank.

Causal Inference for Time-Series and Panel Data

November 11 & 16: Difference-in-Difference Models

Difference-in-difference models are useful when: 1) we have observations of each of the members of the sample for at least two periods in time; and, 2) one group within the sample experienced the treatment between these two periods, while the other did not. We can thus observe whether group-level differences in the outcome of interest changed from one period to the next, facilitating an estimate of the treatment effect.

- Joshua D. Angrist and Jörn-Steffen Pischke. "Differences in Differences." In *Mastering 'Metrics: The Path from Cause to Effect*, chapter 5. Princeton: Princeton University Press. (this may be a refresher for you)
- Andrew Baker. "Difference in Differences Methodology." https://andrewcbaker.netlify.com/2019/09/25/difference-in-differences-methodology/.
- Rafael Di Tella and Ernesto Schargrodsky. 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack." *The American Economic Review*, 94(1): 155–133.
- Janet Currie and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3(1): 65-90.

November 18 & 30: Interrupted Time Series

With cross-sectional data, we used regression discontinuity models to make a causal inference around a cutoff point on an observable characteristic. With time-series data, that cutoff point can be a point in time, such as the date of a policy implementation. Comparison cases facilitate the identification of shifts in levels and trends of the outcome variable.

- Youseop Shin. 2017. "Time Series Analysis as an Impact Analysis Method." In *Time Series Analysis in the Social Sciences: The Fundamentals*, chap. 7. University of California Press.
- Kelly Hallberg, Ryan Williams, Andrew Swanlund, and Jared Eno. 2018. "Short Comparative Interrupted Time Series using Aggregate School-level Data in Education Research." *Educational Researcher* 47(5): 295–306.
- Hinda Ruton et al. 2018. "The Impact of an mHealth Monitoring System on Health Care Utilization by Mothers and Children: An Evaluation Using Routine Health Information in Rwanda." *Health Policy and Planning* 33: 920–927.
- David K. Humphreys, Antonio Gasparrini, and Douglas J. Wiebe. 2017. "Evaluating the Impact of Florida's 'Stand Your Ground' Self-defense Law on Homicide and Suicide by Firearm: An Interrupted Time Series Study." *JAMA Internal Medicine* 177(1): 44–50.
- Catherine Hausman and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10(21): 1–20.

December 2 & 7: Synthetic Control Method

This method builds upon matching methods by creating a "synthetic" comparison case out of a pool of possible cases. Ideally, we would be able to compare a case that has undergone some treatment to the counterfactual scenario in which it had not undergone the treatment. Since this counterfactual scenario does not exist, we construct one using a weighted average of similar cases that did not undergo the treatment. We then can compare the observed outcomes against those predicted by under the counterfactual.

- Alberto Abadie. 2019. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." Manuscript prepared for the Journal of Economic Literature.
- Robert McClelland and Sarah Gault. 2017. "The Synthetic Control Method as a Tool to Understand State Policy." The Urban Institute.
- John Springford. 2018. "The Cost of Brexit to June 2018." Center for European Reform Insight, September 30, 2018.
- Timo Mitze, Reinhold Kosfeld, Joannes Rode and Klaus Wälde. 2020. "Face Masks Considerably Reduce COVID-19 Cases in Germany: A Synthetic Control Method Approach." IZA Discussion Paper Series.