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.

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

<table>
<thead>
<tr>
<th>Problem Set</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Set 1</td>
<td>September 21</td>
</tr>
<tr>
<td>Problem Set 2</td>
<td>October 5</td>
</tr>
<tr>
<td>Problem Set 3</td>
<td>October 12</td>
</tr>
<tr>
<td>Midterm Exam</td>
<td>October 19</td>
</tr>
<tr>
<td>Problem Set 4</td>
<td>November 2</td>
</tr>
<tr>
<td>Problem Set 5</td>
<td>November 16</td>
</tr>
<tr>
<td>Problem Set 6</td>
<td>December 7</td>
</tr>
<tr>
<td>Final Exam</td>
<td>December 17</td>
</tr>
</tbody>
</table>

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](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](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](https://cran.r-project.org). Students who use R are strongly encouraged to download the free RStudio Desktop compan-
ion 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:


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.

• 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.


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.


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.


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.


### 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.


#### 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.

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.


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.
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 multi-level models for this purpose.


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.

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.

- “Time-To-Event Data Analysis.” Columbia University Mailman School of Public Health. Web resource.

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.


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.
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.


