MGMT 737: Applied Empirical Methods

Spring 2021

Instructor:	Paul Goldsmith-Pinkham	**Time:**	TTh, 1-2:20pm
Email:	paul.goldsmith-pinkham@yale.edu	**Place:**	Online

Course Pages:

1. https://github.com/paulgp/applied-methods-phd

2. https://yale.instructure.com/courses/64286 [Yale students only]

Office Hours: After class, or by appointment.

Course Description: This course is primarily designed for graduate students interested in econometric methods used in empirical research. The goal of this class is to provide an overview of different empirical methods, with an emphasis on practical implementation. I will provide a set of lecture slides and notes. There are additional background papers that are largely optional.

More generally, this is a course where I focus on providing my understanding and intuition of empirical methods, as they are used by practicioners. This means that this is not a course where we will spend a lot of time on the formal details (beyond what is necessary), but instead focus on the intuitive framework that guides these papers. I'll also do my best to communciate how any of these topics fit together.

This is a course very much focused on communication and artisanship. By the end of the term, my hope is for three things:

1. You will have been exposed to a wide range of empirical methods, and have at least a passing familiarity with their pros and cons. Moreover, you will know where to go look if you decide to use these methods.

2. Much of the terminology and jargon that we use in econometric methods will be less intimidating to you. When someone says "I use semiparametric inference," now instead of intimidate you, it will bother you that they are not using clearer language.

3. You will approach research papers with the desire to disentangle the underlying framework and "experiment" that drives their causal inferences.

Assignments: There will be problem sets every week. These will involve both theoretical calculations and computer exercises in which you will be asked to analyze data sets. You can use any computer package you wish to use. Solutions will be handed out written in R. Since there will be a fair number of problem sets, and in order to allow me to post the solutions quickly on the webpage for the course, I will not accept late problem sets. If you anticipate difficulty meeting the deadline, you can ask me for the problem set earlier to give you additional time to work on it.

You can work together on the problem sets and discuss them with classmates, but you need to write up the results individually and hand them in separately. Grades will be based on the problem sets, divided evenly over the problem sets.

I expect these assignments to be coded from "scratch." I will specify when canned packages are appropriate. In other words, when estimating a regression, I am not looking for the results of `lm( y ~ x)`. Rather, I expect you to construct two matrices and calculate the estimates using this. I also expect you to attempt to maintain good coding practices while doing so – this will likely be challenging for those of you who are inexperienced at programming, so please plan accordingly – I will not be providing additional instruction on coding beyond what I cover in class. See the following resources in R for guidance (Many thanks to Max Kasy for organizing these materials):

- Introduction to Base R: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

- R for Data Science: https://r4ds.had.co.nz/

- Guidance on Data Visualization: https://socviz.co/

Main References: This is a partial list of various interesting and useful books that will be touched during the course.

- Joshua Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics*

- Scott Cunningham, *Causal Inference: The Mixtape*, https://mixtape.scunning.com/

- Peter M. Aronow and Benjamin T. Miller, *Foundations of Agnostic Statistics*

- Kieran Healy, *Data Visualization: A Practical Introduction*, https://socviz.co/

Prerequisites: ECON 550, ECON 551

Course Outline

1. Causality, Statistics, and Economics

   (a) Potential Outcomes and Directed Acyclic Graphs
     - Chapter 7, Aronow and Miller
     - Chapter 3, Cunningham
     - "Structural vs. Reduced Form" Language, Confusion and Models in Empirical Economics, Haile http://www.econ.yale.edu/~pah29/intro.pdf (or on course website)
     - "Statistics and Causal Inference", Holland 1986
     - "The Identification Zoo: Meanings of Identification in Econometrics", Lewbel, 2019

   (b) Randomization: Experiments, Non-experimental, and Fisher
     - "Causality and design-based inference", Bowers, J. & Leavitt, T. (2020)
     - "Instruments, Randomization, and Learning about Development" Deaton (2010)
     - "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)" Imbens (2010)
     - "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy"
     - "Let's take the Con out of Econometrics", Leamer (1983)
     - "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", Angrist and Pischke (2010)

   (c) Propensity Scores
     - "The central role of the propensity score in observational studies for causal effects" Rosenbaum and Rubin (1983)
     - "Matching As An Econometric Evaluation Estimator", Heckman, Ichimura, and Todd (1998)
     - "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", Hirano, Imbens and Rider (2003)
     - "Propensity Score-Matching Methods for Nonexperimental Causal Studies", Dehijia and Wahba (2002)
     - "Does matching overcome LaLonde's critique of nonexperimental estimators?", Smith and Todd (2005)

- "Practical propensity score matching: a reply to Smith and Todd", Dehijia (2005)
- "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review", Imbens (2004)

(d) Interference, Spillovers and Dynamics

- "Identification of Endogenous Social Effects: The Reflection Problem" Manski (1993)
- "Social Networks and the Identification of Peer Effects" Goldsmith-Pinkham and Imbens (2013)
- "Identification of treatment response with social interactions" Manski (2013)
- "Estimating average causal effects under general interference." Aronow, P. M. & Samii, C. (2017)
- "Exact p-Values for Network Interference", Athey, Eckles and Imbens (2018)
- "Estimating peer effects in networks with peer encouragement designs" Eckles, Kizilcec and Bakshy (2016)
- "Causal Inference under Temporal and Spatial Interference." Wang (2020) https://www.yewang-polisci.com/publications

2. <u>Linear Regression</u>

(a) Bootstrap, Randomization Inference + Clustering

- "Robust Standard Errors in Small Samples: Some Practical Advice." Imbens and Kolesár (2016)
- "GMM estimation with cross sectional dependence" Conley (1999)
- "The Standard Errors of Persistence" Kelly (2019)
- "Clustering, spatial correlations, and randomization inference." Barrios et al. (2012)
- "Sampling-based vs. Design-based Uncertainty in Regression Analysis." Abadie et al. (2019)
- Abadie et al. "When Should You Adjust Standard Errors for Clustering?" 2017
- Athey et al. "Using Wasserstein Generative Adversarial Networks for the Design of Monte Carlo Simulations"

(b) Binscatter, Figures, and Graphical design

- "On Binscatter" Cattaneo et al. (2019)
- "Validation of Visual Statistical Inference, Applied to Linear Models", Majumder, Hoffman and Cook (2014)
- "Visual Inference and Graphical Representation in Regression Discontinuity Designs," Kortin, Lieberman, Matsudaira, and Shen (2020)
- "Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks" Schwabish (2021)
- "Data Visualization: A Practical Introduction", https://socviz.co/, Healy

(c) Quantile Regression

- Koenker and Hallock. Quantile Regression. 2001
- Koenker. Quantile Regression: 40 Years On. 2017

(d) Regularized linear regression

- "Regression shrinkage and selection via the lasso" Tibshiriani (1996)
- "The adaptive lasso and its oracle properties", Zou (2006)
- "Sparse estimators and the oracle property, or the return of Hodges' estimator" Leeb and Potscher (2008)
- "Preconditioning the lasso for sign consistency" Jia and Rohe (2015)
- "Machine learning: an applied econometric approach" Mullainathan and Spiess (2017)

- "Double/debiased/neyman machine learning of treatment effects" Chernozhukov et al. (2017)
- "Double/debiased machine learning for treatment and structural parameters" Chernozhukov et al. (2018)
- "Generic machine learning inference on heterogenous treatment effects in randomized experiments" Chernozhukov et al. (2020)
- "High-Dimensional Methods and Inference on Structural and Treatment Effects", Belloni, Chernozhukov and Hansen (2014)
- "Inference on treatment effects after selection among high-dimensional controls" Belloni, Chernozhukov and Hansen (2014)
- "On model selection consistency of Lasso" Zhao and Yu (2006)
- "Omitted variable bias of Lasso-based inference methods: A finite sample analysis" Wuthrich and Zhu (2020)

3. <u>Likelihood Methods</u>

   (a) Discrete Choice Models I
      - "Discrete Choice Methods with Simulation" Train (2009) https://eml.berkeley.edu/books/choice2.html

   (b) Discrete Choice Models II
      - "Maximum score estimation of the stochastic utility model of choice", Manski (1975)
      - "Analysis of covariance with qualitative data" Chamberlain (1980)
      - "Binary Response Models for Panel Data: Identification and Information" Chamberlain (2010)

   (c) MLE + Duration models
      - "Econometric Methods for the Duration of Unemployment", Lancaster (1979)
      - "Generalised residuals and heterogeneous duration models: With applications to the Weilbull model", Lancaster (1985)
      - "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment", Kroft, Lange and Notowidigdo (2013)
      - "Economic duration data and hazard functions", Kiefer (1988)
      - "Duration Models: Specification, Identification and Multiple Durations", Van Den Berg (2001)

   (d) Hierarchical modeling + Shrinkage Estimation
      - "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates" Chetty and Hendren (2018)
      - "Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments" Meager (2019)
      - "Investing for the Long Run when Returns Are Predictable." Barberis (2000)

4. <u>Canonical Research Designs</u>

   (a) Difference-in-differences
      - "Difference-in-differences with variation in treatment timing", Goodman-Bacon (2018)
      - "Two-way fixed effects estimators with heterogeneous treatment effects" de Chaisemartin and d'Haultfoeuille (2020)
      - "Design-based analysis in difference-in-differences settings with staggered adoption" Athey and Imbens (2018)
      - "Difference-in-differences with multiple time periods", Callaway and Santa'Anna (2020)

- "Pre-event trends in the panel event-study design", Freyaldenhoven et al. (2019)
- "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data", Imai and Kim (2020)
- "Fuzzy differences-in-differences" de Chaisemartin and d'Haultfoeuille (2018)
- "Semiparametric difference-in-differences estimators" Abadie (2005)

(b) Synthetic control + difference-in-difference

- "Using synthetic controls: Feasibility, data requirements, and methodological aspects" Abadie (2019)
- "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program" Abadie, Diamond and Hainmueller (2010)
- "Synthetic Difference In Differences", Arkhangelsky et al. (2019)

(c) Instrumental Variables (Part I)

- "Identification and estimation of local average treatment effects" Imbens and Angrist (1994)
- "Identification of causal effects using instrumental variables" Angrist, Imbens and Rubin (1996)
- "Identification of Causal Effects Using Instrumental Variables: Comment", Heckman (1996)
- "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations" Heckman (1997)
- "Comment on James J. Heckman,'Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations"' Angrist and Imbens (1999)
- "Instrumental variables: response to Angrist and Imbens" Heckman (1999)

(d) Instrumental Variables (Part II)

- "On the structure of IV estimands" Andrews (2019)
- "Weak instruments in instrumental variables regression: Theory and practice" Andrews, Stock and Sun (2019)
- "Jackknife instrumental variables estimation" Angrist, Imbens and Krueger (1999)
- "Random effects estimators with many instrumental variables" Chamberlain and Imbens (2004)
- "Tolerating defiance? Local average treatment effects without monotonicity" de Chaisemartin

(e) Bartik + Simulated Instruments

- "Bartik Instruments: What, When, Why and How" Goldsmith-Pinkham, Sorkin and Swift (2020)
- "Quasi-experimental shift-share research designs" Borusyak, Hull and Jaravel (2020)
- "Shift-share designs: Theory and inference" Adao, Kolesar and Morales (2019)
- "Non-random exposure to exogenous shocks: Theory and applications" Borusyak and Hull (2021)
- "The Estimation of Treatment Effects in Simulated Instrument Designs", Aronow, Goldsmith-Pinkham and Sorkin (mimeo)

(f) Judge IV

- "Consumer bankruptcy and financial health" Dobbie, Goldsmith-Pinkham and Yang (2016)
- "The criminal and labor market impacts of incarceration." Mueller-Smith (2015)
- "Judging Judge Fixed Effects" Frandsen, Lefgren and Leslie (2020)
- "Racial bias in bail decisions" Arnold, Dobbie, and Yang (2018)
- "Family Welfare Cultures", Dahl, Kostol and Mogstad (2014)

(g) Regression Discontinuity and Regression Kink Design

- "Identification and estimation of treatment effects with a regression-discontinuity design" Hahn, Todd and Van Der Klaauw (2001)
- "Inference in Regression Discontinuity Designs with a Discrete Running Variable", Kolear and Rothe (2018)
- "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs", Gelman and Imbens (2018)
- "Inference on causal effects in a generalized regression kink design" Card et al. (2015)
- "Robust nonparametric confidence intervals for regression-discontinuity designs" Calanico et al. (2014)
- "Regression discontinuity designs using covariates" Calanico et al (2019)

(h) Bunching Estimators
- "On Bunching and Identification of the Taxable Income Elasticity" Blomquist et al. (2019)

5. <u>Machine Learning</u>

(a) Supervised Machine Learning
- "Machine Learning Methods Economists Should Know About."Athey S, Imbens G. (2019)
- "Predictably Unequal? The Impact of Machine Learning on Credit Markets" Fuster et al. (2020)
- "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests" Athey and Wager (2019)
- "Deep Neural Networks for Estimation and Inference" Ferrell, Liang and Misra (2020)

(b) Unsupervised ML
- "Text as Data", Kelly, Gentzkow and Taddy (2020)
- "Measuring Technological Innovation Over the Long Run", Kelly et al. (2020)

6. <u>Miscellaneous</u>

(a) Partial Identification
- "Nonparametric bounds on treatment effects" Manski. (1990)
- "Confidence intervals for partially identified parameters" Imbens and Manski (2004)
- "Inference on regressions with interval data on a regressor or outcome" Manski and Tamer (2002)
- "Estimation and Confidence Regions for Parameter Sets in Econometric Models" Chernozhukov, Hong and Tamer (2007)

Grading Policy: Grades will be based on the problem sets, divided evenly over the problem sets. There will be no mid-term or final exam

Class Policy:

- Regular attendance is not required, due to Covid circumstances, but is preferred.