MIT, Fall term 2022, Syllabus for:
14.385 Nonlinear Econometric Analysis, second half
(preliminary)

instructor	Maximilian Kasy
email	teachingmaxkasy@gmail.com
webpage	https://maxkasy.github.io/home/Nonlineareconometrics_MIT_2022/
canvas page	https://canvas.mit.edu/courses/15680
class time	Monday and Wednesday, 1:00-2:30
location	E51-361
office hours	Office 454, Wednesday, 3:30-5pm
TA	Jaume Vives, vives@mit.edu
recitation	Friday, 3:30-5:00, E51-376

1 Overview and Objectives

This is the syllabus for the second half of "Nonlinear Econometric Analysis." The first half is taught by Whitney Newey. Lectures for the second half start on October 31.

We will start by discussing **identification of causal effects**. After introducing basic concepts and talking briefly about historical origins, our focus will be on canonical settings and assumptions that allow us to recover causal relationships, including randomized experiments, conditional exogeneity, instrumental variable methods, difference in differences, and regression discontinuity.

Thereafter, we will consider a set of topics in **applied econometrics**, and related topics in **machine learning**. Each of these topics is interesting in its own right. In addition, by contrasting econometric and machine learning methods, we will be able to discuss commonalities and differences in the approaches of these two fields, and what economists can usefully learn from machine learning.

2 Outline of the course

Identification and causality

1. Basic concepts: Causality, structural objects, identification

   Historical origins: Linear systems of structural equations, selection models

2. Potential outcomes, randomization, and treatment effects

   Instrumental variables, local average treatment effects

3. Conditional independence, reweighting and regression with controls

   Difference in differences

   Regression discontinuity

Applied econometrics and machine learning mashup

1. Randomized experiments

   Multi-armed bandits

2. Dynamic discrete choice

   Reinforcement learning

3. Kernel regression

   Probably approximately correct learning theory

4. Matching

   Double/debiased treatment effect estimation

5. Synthetic controls

   Matrix completion

3 Assignments

Grading will be based on problem sets where you are asked to implement some simulations and estimators on the computer. You should implement these using R and R-Markdown (or Quarto).[1] We will not accept late problem sets. There will be 3 problemsets for the second half of 14.385. Solutions are to be submitted via Canvas, by the due-dates posted there (i.e., November 18, December 2, and December 16).

Students are allowed to collaborate in small groups (of no more than 5 students) for the assignments. Students in a group are allowed to share jointly written computer code and to work together to solve the assignments. However, each student must write up her or his answers completely independently. At the beginning of the code please write down the names of all the people who helped create it.

All lecture slides are available on the course webpage. The slides for the applied econometrics part (randomized experiments, dynamic discrete choice, kernel regression, and matching) are based on slides by Alberto Abadie.

[1] Links to useful resources for learning R can be found at `https://maxkasy.github.io/home/computationlinks/`.

4 Suggested readings

The following readings (except for mostly harmless econometrics) can be downloaded from the course webpage.

4.1 Causality and identification

Textbooks

Angrist, J. D. and Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press, chapters 2, 4, 5, and 6.

Manski, C. (2003). *Partial identification of probability distributions.* Springer Verlag, chapters 2 and 7

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press

Local average treatment effects

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455

Regression discontinuity

Hahn, J., Todd, P., and der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209

4.2 Applied econometrics and machine learning

Randomized experiments

Athey, S. and Imbens, G. W. (2017). The econometrics of randomized experiments. In *Handbook of Economic Field Experiments*, volume 1, pages 73–140. Elsevier.

Bandit problems

Bubeck, S. and Cesa-Bianchi, N. (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122.

Dynamic discrete choice

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38–67.

Reinforcement learning

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Kernel regression

Härdle, W. and Linton, O. (1994). Applied nonparametric methods. *Handbook of econometrics*, 4:2295–2339

Probably approximately correct learning theory

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press, chapters 2 to 6.

Matching

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Double/debiased machine learning

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Synthetic controls

Abadie, A., Diamond, A., and Hainmueller, J. (2010). 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.

Matrix completion

Agarwal, A., Dahleh, M., Shah, D., and Shen, D. (2021). Causal matrix completion. *arXiv preprint arXiv:2109.15154*.