

### **Overview of the Theme:**

Scholars and judges have long made arguments about laws and regulations and justified their arguments with theories about the effects of these legal rules. What empirical law and economics scholars like to do is to evaluate these claims with data. Now a particularly challenging dimension of studying the effects of rules and regulations is that many other aspects of society change at the same time, so what we need are natural experiments.

Why is this even an important endeavor? There are jurists, such as Judge Richard Posner, who argue that understanding the empirical consequences of judicial decisions is important so that judges can make better cost-benefit utilitarian analyses. There are also jurists, such as Justice Stephen Breyer, who argue that understanding the consequences of their decisions is important so judges can make decisions that accord with the democratic will of the people.

The dominant legal theory in US courts, at least, is law and economics, which articulates deterrence as the primary explanation for societal response to law. But a large body of work in psychology and sociology suggests that laws can shape behaviors simply by telling individuals what is the right thing to do. From abolition of slavery, to women's liberation, to environmentalism, the courts are speculated to play a key role in shaping values, yet little causal evidence exists to date. Whether courts shape values and in which direction is important to both arbitrate between competing theories about the effects of laws and as inputs into better judicial decisions.

### **Introduction: Causal Inference**

- (Required) Duflo, Esther. 2001. "[Empirical Methods](#)"
- (Required) Mullainathan, Sendhil, and Jann Spiess. 2017. "[Machine Learning: An Applied Econometric Approach](#)." *Journal of Economic Perspectives* 31 (2): 87-106.
- (Recommended) Esther Duflo, Rachel Glennerster, and Michael Kremer. "[Using randomization in development economics research: A toolkit](#)." In T. Paul Schultz and John A. Strauss, editors, **Handbook of Development Economics**, volume 4, chapter 61, pages 3895–3962. Elsevier, 2008

### **Lecture 1: Natural Experiments Using Random Assignment of Judges**

- (Required) Williams, Heidi, and Bhaven Sampat. [How Do Patents Affect Follow-on Innovation? Evidence from the Human Genome](#); Revise-and-resubmit, *American Economic Review*; [NBER working paper #21666](#)

- (Required) D. Autor, M. Mogstad, A. Kostol, and B. Setzler. “[Disability Receipt, Consumption Insurance, and Family Labor Supply](#)” Revise-and-resubmit, *American Economic Review*.
- (Recommended) Dobbie, W. and J. Song. [Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection](#), *American Economic Review*, 105(3), 2015.
- (Recommended) Dahl, G., M. Mogstad, and A. Kostol. “[Family Welfare Cultures](#)”, *Quarterly Journal of Economics*, 129(4): 1711-1752, 2014.

## Lecture 2: Text Analysis

- (Required) Gentzkow, Matt, Jesse Shapiro, and Matt Taddy. [Measuring Polarization in High-dimensional Data: Method and Application to Congressional Speech](#). Last updated: May 2017; *NBER Working Paper* <http://www.nber.org/papers/w22423>
- (Required) Ash, Elliott. [The political economy of tax laws in the U.S. states](#)
- (Recommended) Gentzkow, Matt and Jesse Shapiro. [What Drives Media Slant? Evidence from U.S. Daily Newspapers](#). *Econometrica*. 78 (1). January 2010.

## Lecture 3: Predictive Analytics

- (Required) Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovic, Jens Ludwig, and Sendhil Mullainathan. “[Human Decisions and Machine Predictions](#)”. Forthcoming. *Quarterly Journal of Economics* [w23180.pdf](#)
- (Required) Arnold, David, Will Dobbie, and Crystal Yang. [Racial Bias in Bail Decisions](#), Revise-and-resubmit, *Quarterly Journal of Economics*.
- (Recommended) Chernozhukov, Victor, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey. [Double/Debiased/Neyman Machine Learning of Treatment Effects](#) *American Economic Review Papers and Proceedings*, 107(5): 261-65. May 2017.
- (Recommended) Stefan Wager and Susan Athey. [Estimation and Inference of Heterogeneous Treatment Effects using Random Forests](#). *Journal of the American Statistical Association*, forthcoming, 2017.

For projects, there are a variety of legal datasets that comprise in total 4 terabytes. We have a readme to teach you how to access the appropriate cloud computing like AWS (Amazon Web Services). You may also use Microsoft Azure, where we received a grant.

The data are extremely expensive in terms of money, time, effort, and resources. More than 10 years have been invested in these datasets. The data is strictly confidential. Anyone who wants to access needs to sign a non-disclosure agreement in terms of safeguarding the confidentiality of the data.

The datasets involve random assignment of judges, their history of how they write and decide, and in some settings, how they speak.

You are encouraged to come up with your own ideas, but a codebase does exist. If you and your team find it more efficient to learn from sample machine learning code, we can discuss economics extensions.

1. Digitized universe of all 380,000 cases, 1 million judge votes, across 94 hand-labeled legal topics in U.S. Circuit Courts from 1891, engineered into 2 billion N-grams of length eight, and 5 million citation edges across cases. This is merged with the 268 judges who served during this time period, 250 biographical features, a 5% random sample with 400 hand-labeled features (330-paged [codebook](#)), and 6000 cases hand-coded for meaning in 25 polarized legal areas. This is merged with administrative data (date of key milestones, e.g., oral arguments, when was the last brief filed, etc.), publicly available U.S. Supreme Court datasets, U.S. District Court datasets, geocoded judge seats, (some) biographies of judicial clerks, and (some) oral arguments' audio files. The identities of randomly assigned judges sitting on 3-judge panels (who is authoring the opinions, writing dissents, or writing concurrences) render a random seating network among the judges. To study equal treatment before the law and equality based on recognition of difference and to study the transmission of legal thought.
2. Digitized universe of 1.5 million criminal sentencing decisions across 94 U.S. District Courts from 1991 FOIA-requested to link to the judge identity, then merged to the biographies of 1300 judges and available corpus of their opinions since 1923. Judges are randomly assigned. There is an 83-paged [codebook](#) from U.S. Sentencing Commission on features of the case and the criminal defendant. The data is also linked to local daily weather and sporting events.
3. Digitized speech patterns in U.S. Supreme Court oral arguments since 1955—longitudinal data on speech intonation (linguistic turns) are rare. Linked to hand-labeled oral advocates' biographies, (some) lawyers' faces, clipped identical introductory sentences, ratings of their traits, and publicly available U.S. Supreme Court databases containing dozens of additional features and preceding U.S. Circuit Court data. Actual analysis of speech patterns is statistically challenging, since speech is modified dynamically. A common measure for variation in speech patterns considers resonances of vowel sounds. In order to properly measure these, the starting locations of all distinct vowel sounds have been manually flagged. A machine then measured vowel resonances and assigned to each vowel sound a multidimensional continuous quantity. Therefore, the size of the oral data set is much larger than the size of the underlying text. Text is traditionally treated with discrete models. Speech measurements (for example resonances) by contrast are continuous. Recordings of U.S. Circuit Court oral arguments for a limited time period.
4. We also have the digitized universe of U.S. State Supreme Court cases from 1947-1994 (roughly 400,000 cases), identities of judges sitting on the panels, hand-labeled biographies, citation network, and original text. Some of these judges run for election.

5. We have the administrative data on the universe of 1 million refugee asylum and 15 million hearing sessions, their time of day, across 50 courthouses (with randomly assigned judges), since the founding of the U.S. immigration courts in 1983.
6. We have the universe of individuals in a federal prosecutor's office for 12 years comprising 430,000 charges, 145,000 defendants, and many stages of random assignment. The 594 page codebook includes detailed information regarding each individual offender and the corresponding prosecutor and judge, names, victims, witnesses, defense attorneys, police officers, etc. The novelty of this dataset is that it tracks individuals from the time of arrest to the final sentence if any. Other datasets are not linked.
7. We are developing a database on the European Court of Justice and have some administrative contacts. Please inquire if interested.