PLSC 438/536: Applied Quantitative Research Design

Instructor: Prof. Shiro Kuriwaki*

Fall 2023	Yale University
Professor	Shiro Kuriwaki
Teaching Fellows	Changwook Ju Sean Nossek Robin Wang
Lecture	twice a week, 75 minutes
Review Section	50 minutes

Note: This is a public version of the Fall 2023 course syllabus. It focuses on course organization and readings, while omitting details on university-specific information on dates, accommodations, and logistical notes. Future versions of the class will differ.

Course Description

Research designs are strategies to obtain empirical answers to theoretical questions. Research designs using quantitative data for social science questions are more important than ever. This class, intended for advanced students interested in social science research, trains students with best practices for implementing rigorous quantitative research. We cover techniques in causal inference, prediction, and missing data, such as fixed effects, time series, instrumental variables, survey weighting, and shrinkage. This is a hands-on, application-oriented class. Exercises involve programming and statistics used in exemplary articles in quantitative social science. The final project advances a research question built on a replication of a paper chosen in consultation with the instructor.

Prerequisite: Any statistics or data science course that teaches ordinary least squares regression (OLS) and p-values. Past or concurrent experience with a programming language such as R is strongly recommended.

*Assistant Professor of Political Science, Yale University. https://www.shirokuriwaki.com. This syllabus borrows from class material designed by Dan Levy, Teddy Svoronos, Matt Blackwell, and Kosuke Imai. I thank them for sharing some of their course material. If you would to see any course material including problem sets, please contact me.

Assessment

The course grade will be a weighted average of the following components:

- Ten weekly or semi-weekly problem sets: 35%
- In-class exam on concepts: 10%
- Final Paper: 40%
- Participation, including pre-class reading responses for papers: 15%

Most problem sets will have 3 parts covering 8-10 exercises. These cover visualization, table construction, replicating statistical analyses of the papers we discuss, and discussion questions from the papers.

Schedule

This course revolves around the close reading, discussion, and data analysis of the following seven academic papers and case studies:

Paper	Topics Covered
Banerjee, Abhijit, Esther Duflo, Dean Karlan, *et al.* "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." (2015) *Science*. [Dataverse]	RCT, Regression, non-compliance
Malhotra, Neil, Yotam Margalit, and Cecilia Mo, "Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact." (2013). *American Journal of Political Science*. [Dataverse]	Observational regression
Miguel, Edward, Shanker Satyanath, and Ernest Sergenti (2004). "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." (2004) *Journal of Political Economy.* [Replication]	Instrumental Variables
Scheve, Kenneth and David Stasavage, "Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation." (2012) *American Political Science Review*. [Replication]	Panel Data, Difference in differences
Alix-Garcia, Jennifer, Katherine Sims, *et al.*,"Payments for environmental services supported social capital while increasing land management." (2022) *American Political Science Review* [Replication]	Regression Discontinuity
Toeffel, Michael, Dan Levy, et al. "Improving Worker Safety in the Era of Machine Learning." (2018). HBS Case	Out of sample prediction
Kuriwaki, Shiro, *et al.*, "The Geography of Racially Polarized Voting: Calibrating Surveys at the District Level". [Dataverse] *American Political Science Review*	Survey weighting

We will also refer to the following textbooks:

- Our main text for the econometrics is *Mastering 'Metrics: The Path from Cause to Effect* by Josh Angrist and Jörn-Steffen Pischke. I will sometimes assign parts of their other book, *Mostly Harmless Econometrics*, which is more advanced.
- We will also draw from: Kosuke Imai and Nora Webb Williams, *Quantitative Social Science: An Introduction in tidyverse*, 2022.
- Most coding principles and examples will come from Hadley Wickham, Mine Cetikaya-Rundel, and Garrett Grolemund, *R for Data Science* (2nd edition), available at `https://r4ds.hadley.nz/`.
- The main methods we will implement are also provided as 3-5 minute screencasts that students can play at their convenience: `https://vimeo.com/shirokuriwaki`

The methods we will cover is organized in three major components: Causal inference, prediction, and uncertainty, in that order. Some of the other weeks focus on using R and writing academic papers. The detailed schedule below lists topics, readings, and assignments for each class. QSS refers to the Imai text and TCD refers to Bueno de Mesquita and Fowler. Lectures are numbered by week and number.

Class	Topic	Detailed Topic and Readings	Due
0.1	**Course overview**		Before class, install R and RStudio (R screencast)
0.2	**Workflow**	RStudio Projects, tidyverse, dplyr, pipes • R for Data Science 2e, chapters 4 and 6 "Workflow: code style" and "Workflow: scripts and projects"	
1.1	**Visualizing Data**	grammar of graphics with ggplot, visual perception • Rauser, "How Humans See Data" (talk)	Pset 1 (RStudio project, code style, dplyr, histograms)

Continued on the next page

2.1	**Randomized Control Trials**	Linear regression as difference in means, Estimating treatment effects with linear regression • *Mastering Metrics* 1.1 • QSS 2.1-2.4 "Causal Effects in the Counterfactual," "Randomized Controlled Trials"; 4.4 "Regression - Randomized Control Trials"
2.2	**Discuss Banerjee et al.**	"A multifaceted program causes lasting progress for the very poor: Evidence from six countries"
		Pset 2 (analyze Banerjee *et al.* with OLS, barplots, regression tables)
3.1	**Confounding and Omitted Variable Bias**	Using regression in observational data for causation • QSS 2.5.2 "Confounding Bias", 4.3.2 "Regression with multiple predictors" • *Mastering Metrics* Ch. 2, "Regression"
3.2	**Units**	Substantive interpretation of regression coefficients, percent vs. percentage point, implications for changing the units of the left-hand and right-hand side, Z-scores • QSS p.122-123; p.330-331 for Z-scores • *Mastering Metrics* p.93-94 for regressions with logs
		Pset 3 (replicate Malhotra *et al.*, shapefiles, recoding values, standardizing variables)
4.1	**Discuss Malhotra et al.**	"Economic Explanations for Opposition to Immigration: Distinguishing between Prevalence and Conditional Impact."
4.2	**Non-compliance**	Never-takers, compliers, Treatment on Treated vs. Intent to treat • *Mastering Metrics* 3.1 (end of that section) • Thinking Clearly with Data 11 "Randomized Experiments", p. 225-231 "Noncompliance and instrumental variables"

Continued on the next page

		Pset 4 (Review of OVB in Malhotra *et al.*, for loops with `purrr::map`, TOT vs. ITT)
5.1	**Instrumental variables**	Discuss Miguel, Satyanath, Sergenti, "Economic Shocks and Civil Conflict: An Instrumental Variables Approach."
5.2	**Difference in Differences**	Card and Krueger study, Long and wide data, reshaping with `pivot_longer` and `pivot_wider` • *Mastering Metrics* 5.1
		Pset 5 (replicating Miguel et al. IV; line plots of trend data) (R screencast)
6.1	**Fixed Effects**	Time-invariant and unit-invariant confounding, logic of fixed effects, implementation in R with `fixest` • *Mastering Metrics* 5.2 • QSS 2.5 "Observational studies"
6.2	**Discuss Scheve and Stasavage**	"Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation." Parallel trends, leads and lags, unit-specific time trends • *Mostly Harmless*, leads and lags text around equation 5.2.6
6.3	**Panel Data, Review**	Connection between TWFE and DID, standard errors • *Mastering Metrics* p.95-97; 205-208 on SEs in regression and DID. • Advanced: Goodman-Bacon decomposition
		Pset 6 (Replicating Scheve and Stasavage TWFE, leads and lags) R screencast

Continued on the next page

7.1	**Regression Discontinuity**	Bias and variance, RD identifying assumptions, interactions in regression coefficients • *Mastering Metrics* ch. 4 • QSS 7 on coefficient standard errors • P. M. Aronow, PLSC 503 lecture notes
7.2	**Discuss Alix-Garcia et al.**	"Payments for environmental services supported social capital while increasing land management." Pset 7 (RDD) (R screencast)
8.1	**Regression for Prediction**	Prediction with new data `predict`, MSE, RMSE, and R^2 • QSS 4 "Prediction", especially 4.1-4.2
8.2	**Shrinkage**	Out of sample prediction, Overfitting, LASSO regression, tuning parameters • Mullainathan and Spiess (2017), "Machine Learning: An Applied Econometric Approach" *Journal of Economic Perspectives* Pset 8 (OSHA case, pre-case prediction) (R screencast)
9.1	**Discuss HBS Case, Worker Safety at OSHA**	"Improving Worker Safety in the Era of Machine Learning" • Also view PBS documentary on OSHA, *A Dangerous Business, Revisited* 55 minutes
9.2	**Surveys and Standard Error**	Survey recruitment, sources of error, cross-tabulation, ratio estimator for unrepresentativeness • QSS 7.1: Estimation - Uncertainty, up to equation 7.16, CLT and confidence intervals

Continued on the next page

10.1	**Weighting and Outcome Modeling**	Inverse probability weighting, standard error, design effects, outcome models • "Trump Leads in 5 Critical States as Voters Blast Biden, Times/Siena Poll Finds". (2023, Nov. 5)
10.2	**Discuss Kuriwaki *et al.*, Synthetic Control**	The Geography of Racially Polarized Voting." Outcome modeling, random effects, bias-variance trade-off • Abadie et al., JASA article • Advanced: Imai short lecture Pset 9 (Replicating Kuriwaki et al.) R screencast
11.1 11.2	**Exam Review** **Writing**	Discuss anonymous peer writing samples. • McEnery, "The Craft of Writing Effectively" (video lecture) • King, "Publication, Publication" • Fiske, "Words to the Wise on Writing Scientific Papers"
12.1	**In-class Exam**	
12.2	**Github / Quarto**	Git, linking Github to RStudio Projects, Final paper • R4DS Quarto chapter (link) • Happy Git with R (link) Pset 10 (clustered SEs, course eval, course review)

Final Project

Assignment: Present original research findings, either by extending one of the papers we read in the class OR (with permission of the instructor) by writing a research paper on another topic. There are three prototypes of projects you can choose between (i.e., choose just one of these):

1. A replication of a suggested paper accompanied by an extension. The papers use the same techniques (instrumental variables) or data structure (country-year panel data) as the class papers, but

they are on different papers by different topics. Students may not replicate a class paper this year.

2. An analysis of a new dataset that asks a similar or related question as one of our course papers.

3. An analysis of an entirely different social science question using similar quantitative research designs, upon permission of the instructor. This is typically reserved for PhD students or students writing a senior essay.

Details about the paper requirements are provided in a separate handout.