

Economics 270D: Research Transparency Methods in the Social Sciences (Ph.D. course)
Professor Edward Miguel (emiguel@berkeley.edu)
SPRING 2015 SYLLABUS

Description: This course covers a range of approaches that aim to enhance the transparency and reproducibility of social science research. It is appropriate for Ph.D. students in social science disciplines and related fields.

Prerequisites: Graduate coursework in econometrics and/or statistical methods, and some experience with statistical software.

Lectures: Tuesday 2-4pm, Moffitt 145

Office hours: Please email Elisa Cascardi (ecascardi@berkeley.edu) for an appointment.

Extra discussion sections and office hours: Garret Christensen (garret@berkeley.edu) and Felipe Gonzalez (fgonzalez@econ.berkeley.edu) will lead occasional discussion sections related to class assignments, including on the statistical and software tools used in the course.

Additional materials: For further readings and presentation materials on related topics, refer to the BITSS 2014 Summer Institute page (<http://bitss.org/training/>), the slides from the BITSS Annual Meetings from 2014 (<https://osf.io/view/bitss2014/>), 2013 (<http://bitss.org/annual-meeting/2013-2/>), and 2012 (<http://bitss.org/annual-meeting/2012-2/>).

The following books may also be useful reference materials:

- Angrist, Joshua, and Jorn-Steffen Pischke. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Cooper, H., L. V. Hedges and J. C. Valentine. (eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation.
- Dunning, Thad. (2012). *Natural Experiments in the Social Sciences: A Design-based Approach*. Cambridge University Press.

Assignments and Grading:

(Note: assignments should be posted on bCourses by 2pm on the due date.)

(1) Referee reports on working papers (30 percent):

- Each referee report should be no more than 3 pages double-spaced. The report should start off with a one paragraph summary of the main argument of the article. You should describe your main 3-4 points in detail as if you were writing directly to the author. Conclude the report with more minor comments. A good referee report not only clearly states the shortcomings of the work, but also lays out constructive, detailed and realistic suggestions for improvement.
- These assignments will allow you to read and critique papers on the research frontier on research transparency topics.

-- Report 1 on Abel Brodeur, Mathias Le, Marc Sangnier, and Yanos Zylberberg, "Star Wars: The empirics strike back", due Tuesday 2/3 [<https://sites.google.com/site/abelbrodeur/papers>]

-- Report 2 on Ingvild Almas, Alexander Cappelen, and Bertil Tungodden, "Pre-analysis Plan: An economic experiment on social preferences with nationally representative populations: the United States versus Europe", due Tuesday 3/3 [<https://www.socialscienceregistry.org/trials/487>]

-- Report 3 on Eva Vivalt, "How much can we generalize from impact evaluations?", due Tuesday 3/31 [<http://evavivalt.com/wp-content/uploads/2014/12/Vivalt-JMP-latest.pdf>]

* Extra credit option for each referee report: please critically assess the presentation of data or other forms of information in each of the referee report papers. You can receive one point of extra credit for each report if you also include a detailed discussion of the shortcomings of one particular table / figure in the paper, and the concrete modifications you would make to improve the clarity of presentation. (This does not count toward the 3 page limit for the referee report itself.)

(2) Three problem sets (30 percent):

Problem sets require manipulation and analysis of data. I ask you to use either STATA or R, to make the solutions more comparable and to facilitate grading. I will ask you to apply some of the methods and approaches from the course to real data, and ask you to interpret the results. These assignments will provide good practice in applying statistical methods to data, something most of you will need to do in your own research. Leading sources for publicly available social science datasets are the Harvard Dataverse Network [<http://thedata.harvard.edu/dvn/>] and the University of Michigan Inter-university Consortium for Political and Social Research (ICPSR) [<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>], but there are many other active data repositories, and some journal websites also host data.

-- Problem set 1, due Tuesday 2/17: Mean effects estimation (based on Kling et al. 2007) and/or multiple hypothesis testing adjustments (based on Anderson 2008) using publicly available data from a published social science research paper.

-- Problem set 2, due Tuesday 3/17: Publication bias, P-curves (based on Simmons et al 2014), and/or meta-analysis (based on Gelman et al 2004) using publicly available data from an existing social science research literature.

-- Problem set 3, due Tuesday 4/14: Apply simple differential privacy techniques (based on Dwork and Smith 2010) to a real dataset (the Netflix Prize data), and identify implications of these techniques for the reliability of standard statistics in the dataset.

(3) One research proposal, 8-9 pages (30 percent):

The research proposal should briefly (3-4 pages) survey the existing literature on a topic related to research transparency and reproducibility, and then describe a planned original research

project (5-6 pages). Proposals should be in 12 point font, double-spaced with 1 inch margins. Proposals exceeding 9 pages in length will lose credit. This assignment will encourage you to generate, refine and receive feedback on a research idea that might form a research article and/or part of your dissertation.

-- The research proposal is due Friday 5/1 at 2pm (uploaded on bCourses).

(4) Class participation (10 percent): In borderline cases, attendance at lecture and the quality of classroom comments may be a factor in assigning grades. Active class participation enriches the course, benefiting yourself and other students (and me), and it is highly encouraged.

(5) There are no exams.

Special Accommodations: If you need disability-related accommodations in this class, if you have emergency medical information you wish to share with me, or if you need special arrangements in case the building must be evacuated, please inform me immediately by email or in office hours.

Honor Code: The student community at U.C. Berkeley has adopted the following Honor Code: *"As a member of the U.C. Berkeley community, I act with honesty, integrity, and respect for others."* The hope and expectation is that you will adhere to this code.

COURSE SYLLABUS

I. Introduction to issues of research transparency and reproducibility

Lecture 1: Understanding the problem (1/20)

Merton, Robert K. (1942). "Science and Technology in a Democratic Order", *Journal of Legal and Political Sociology*, 1, 115-126.

Anderson, Melissa S., Brian C. Martinson, and Raymond De Vries. (2007). "Normative dissonance in science: Results from a national survey of U.S. scientists", *Journal of Empirical Research on Human Research Ethics*, 3-14.

Carey, Benedict. (2011). "Fraud case seen as a red flag for psychology research", *New York Times*, Nov. 2, 2011.

Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, M. Van der Laan. (2014). "Promoting Transparency in Social Science Research." *Science*, 10.1126/science.1245317.

Desposato, Scott. (2014). "Ethical challenges and some solutions for field experiments", unpublished working paper, U.C. San Diego.

Lectures 2-3: Publication bias and the file-drawer problem (1/27, 2/3)

Rosenthal, Robert. (1979). "The file drawer problem and tolerance for null results". *Psychological Bulletin*, 86 (3): 638–641. doi:10.1037/0033-2909.86.3.638.

Card, David, and Alan Krueger. (1995). "Time-series minimum-wage studies: A Meta-analysis", *American Economic Review Papers and Proceedings*, 85(2), 238-243.

Ioannidis, John P.A. (2005). "Why most published research findings are false", *PLoS Medicine*, 2(8), e124.

Gerber, Alan, and Neil Malhotra. (2008). "Do statistical reporting standards affect what is published? Publication bias in two leading political science journals", *Quarterly Journal of Political Science*, 3, 313-326.

Mathieu, Sylvain, Isabelle Boutron, David Moher, Douglas G. Altman, and Philippe Ravaud. (2009). "Comparison of registered and published primary outcomes in randomized controlled trials", *JAMA*, Sept. 2, 2009, 302(9), 977-984.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. (2011). "False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant", *Psychological Science*, 22, 1359-1366.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. (2014). "Publication bias in the social sciences: Unlocking the file drawer", *Science*, 10.1126/science.1255484.

Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. (2015). "P-curve: A key to the file drawer", forthcoming *Journal of Experimental Psychology: General*.

II. Approaches to pre-registration

Lectures 4-5: Using pre-analysis plans (2/10, 2/17)

Leamer, Edward E. (1983). "Let's Take the Con out of Econometrics", *American Economic Review*, 73(1), 31-43.

Glennerster, Rachel, and Edward Miguel. (2009). "Community Driven Development in Sierra Leone: GoBifo Analysis Plan", posted on J-PAL Hypothesis Registry August 21, 2009. [http://www.povertyactionlab.org/sites/default/files/documents/gbf_analysis_plan_final_2009_08.21.pdf]

Casey, Katherine, Rachel Glennerster, and Edward Miguel. (2012). "Reshaping institutions: Evidence on aid impacts using a pre-analysis plan", *Quarterly Journal of Economics*, 127(4), 1755-1812.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph Newhouse, Heidi Allen, Katherine Baicker, and the Oregon Health Study Group. (2012). "The Oregon Health Insurance Experiment: Evidence from the First Year", *Quarterly Journal of Economics*, 127(3), 1057-1106.

Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter van der Windt. (2013). "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration", *Political Analysis*, 21, 1-20.

Additional readings:

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. (2007). "Experimental Analysis of Neighborhood Effects", *Econometrica*, 75(1): 83-119.

Anderson, Michael L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484), 1481-1495.

Lecture 6: Transparency in non-experimental research (2/24)

Neumark, David. (1999). "The Employment Effects of Recent Minimum Wage Increases: Evidence from a Pre-specified Research Design", NBER Working Paper #7171.

Neumark, David. (2001). "Evidence on Employment Effects of Recent Minimum Wage Increases from a Pre-Specified Research Design", *Industrial Relations*, 40(1): 121-144.

Dal-Re, Rafael, John P. Ioannidis, Michael B. Bracken, Patricia A. Buffler, An-Wen Chan, Eduardo L. Franco, Carlo La Vecchia, Elisabete Weiderpass. (2014). "Making prospective registration of observational research a reality", *Science Translational Medicine*, 6(224), 224cm1.

Moravcsik, Andrew. (2014). "Transparency: The Revolution in qualitative research", *PS: Political Science and Politics*, 47(1), 48-53

Lecture 7: Data Adaptive Pre-specification Approaches (3/3)

Petersen M, Schwab J, Geng E, van der Laan M. (2015). "Illustration of analytic methods based on the ADAPT-R Trial", in Moodie E and Kosorok M (eds.) *Dynamic Treatment Regimes in Practice: Planning Trials and Analyzing Data for Personalized Medicine*. Forthcoming, ASA-SIAM, Philadelphia, PA.

Petersen, M. (2015). ADAPT-R Trial Simulation, R code.

III. Building scientific knowledge

Lecture 8: Approaches to the replication of research (3/10)

Hamermesh, Daniel S. (2007). "Replication in Economics", *Canadian Journal of Economics*, 40(3), 715-733.

Albouy, David. (2012). "Colonial Origins of Comparative Development: An Empirical Investigation: Comment", *American Economic Review*, 102(6), 3059-3076.

Klein, R. A., et al. & Nosek, B. A. (2014). "Investigating variation in replicability: A "many labs" replication project", *Social Psychology*, 45, 142-152.

Additional readings:

Ball, Richard and Norm Medeiros. (2013). "Project TIER (Teaching Integrity in Empirical Research) Data Management Protocol". [<http://www.haverford.edu/TIER/files/protocol.pdf>]

Long, J. Scott. (2008). *The Workflow of Data Analysis Using Stata*. Stata Press.

Lectures 9-10: Meta-analysis techniques (3/17, 3/31)

Pritchett, Lant. (2002). "It pays to be ignorant: A simple political economy model of rigorous program evaluation", *Journal of Policy Reform*, 5(4), 251-269.

Hsiang, Solomon, Marshall Burke, and Edward Miguel. (2013). "Quantifying the Influence of Climate on Human Conflict", *Science*, 10.1126/science.1235367.

Allcott, Hunt. (2014). "Site Selection Bias in Program Evaluation", unpublished working paper. [https://files.nyu.edu/ha32/public/research/Allcott_SiteSelectionBias.pdf]

Additional readings:

Ashenfelter, Orley, Colm Harmon, and Hessel Oosterbeek. (1999). "A review of estimates of the schooling/earnings relationship, with tests for publication bias", *Labour economics*, 6(4), 453-470.

Cooper, H., L. V. Hedges and J. C. Valentine. (eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russel Sage Foundation, Part VI (Chapters 14-17).

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald R. Rubin. (2004). *Bayesian Data Analysis, Third Edition*, Chapman and Hall/CRC Texts in Statistical Science, Chapter 5.

[SPRING BREAK – 3/24]

IV. Open data and materials

Lecture 11: What does open data do? (4/7)

Simonsohn, Uri. (2013). "Just Post it: The Lesson from Two Cases of Fabricated Data Detected by Statistics Alone," *Psychological Science*, 24(10), 1875-1888.

Piwowar, Heather A., and Todd J. Vision. (2013). "Data reuse and the open data citation advantage", *PeerJ*, 1:e175. DOI 10.7717/peerj.175.

Dafoe, Allan. (2014). "Science Deserves Better: The Imperative to Share Complete Replication Files." *PS: Political Science & Politics*. 47(1), 60-66.

Vines, Timothy H., et al. (2014). "The Availability of Research Data Declines Rapidly with Article Age", *Current Biology*, 24, 94-97.

Goodman, Alyssa, et al. (2014). "Ten Simple Rules for the Care and Feeding of Scientific Data", *PLoS Computational Biology*, 10(4), e1003542.

Lecture 12: Differential privacy and the cost of openness (4/14)

Dwork, Cynthia, and Adam Smith. (2010). "Differential Privacy for Statistics: What We Know and What We Want to Learn." *Journal of Privacy and Confidentiality*, 1(2), 135– 54.

Sweeney, Latanya, Akua Abu, and Julia Winn. (2013). "Identifying Participants in the Personal Genome Project by Name." [<http://dataprivacylab.org/projects/pgp/1021-1.pdf>]

Heffetz, Ori, and Katrina Liggett. (2014). "Privacy and data-based research", *Journal of Economic Perspectives*, 28(2), 75-98.

V. Looking forward

Lecture 13: Presenting and visualizing data (4/21)

Tufte, Edward R. (1983). *The Visual Display of Quantitative Information*. Cheshire: Graphics Press.

Gelman, Andrew, Cristian Pasarica, and Rahul Dodhia. (2002). "Let's Practice what we preach: Turning tables into graphs", *The American Statistician*, 56(2), 121-130.

Hsiang, Solomon. (2013). "Visually-weighted regression", unpublished working paper.

Lecture 14: Next steps in changing scientific research practices (4/28) – Guest lecturer, Garret Christensen (U.C. Berkeley)

Kuhn, Thomas S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press, Chapters 1-3.

Angrist, Joshua D. and Jorn-Steffen Pischke. (2010). "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics", *Journal of Economic Perspectives*, 24(2), 3-30.

Christensen, Garret and Courtney Soderberg. (2014). "Manual of Best Practices in Transparent Research", unpublished working paper.