



## MGMT 737: Applied Empirical Methods (Quantitative Storytelling) Professor Olav Sorenson - Fall 2017

---

Tuesdays and Thursdays 2:40pm to 4:00pm (165 Whitney Ave, Room 1556)

### Course overview

This course has been designed as a PhD-level survey of empirical methods used in current research on organizations and in sociology. It has three main goals: (1) to introduce students to practical issues that arise in analyzing data; (2) to develop an intuition for why and how various estimation choices matter; and (3) to provide students with a better sense of ways in which one might provide a compelling set of analyses in support of some theoretical story. The course will also give particular attention to common methods in organizations and sociology that often receive less attention in courses on econometrics, such as the modeling of binary and count dependent variables and of survival rates.

The subtitle refers to “storytelling” for two reasons. First, in many cases, the nature of the question of interest or of the data available preclude the identification of a causal effect. Data analyses in these cases remain descriptive. They nevertheless often attempt to convince the reader of the plausibility of one particular interpretation of the relationships observed. Second, even in cases that allow for identification of a causal effect, that estimation often serves as a necessary but not a sufficient condition to providing empirical support for the underlying theoretical mechanism of interest. One should not read “storytelling” to mean that the treatment of methods in the course will not be serious.

Note that the teaching of the course will presume that students arrive with some familiarity with general linear models and causal inference, at least at a theoretical level (e.g., ECON 550a, PLSC 503b, SOCY 501b). In the spirit of a “flipped” class, most of the time spent together will be spent on analyzing data and interpreting the results. Although I am happy to answer any questions that may arise, I will generally not lecture on the content of the background readings in class. Because of the learning-by-doing nature of the course, the workload will likely end up being on the higher side, even relative to other PhD-level offerings. Plan accordingly.

Assignments and exercises for the course will use Stata 14 as the default but students may use other statistical programming languages with the understanding that the instructor may have limited ability to support them in those languages.

## Assessment

Evaluation for this course consists of three components: class participation, assignments, and a final paper.

Component	Weight
Participation	15%
Assignments	60%
Final paper	25%

**Class participation:** Each week will consist mostly of discussion and working through analyses. I expect that students will have read the assigned readings prior to the lecture and come prepared with any questions they may have.

**Assignments:** Much of the course involves learning-by-doing. Almost every week therefore has an assignment to put the ideas covered and discussed into practice.

**Final paper:** As a final paper, you should find a published paper for which you can find data (either because the authors have made the data available or because they use a data source to which you have access). First, try to replicate their results exactly. Second, explore the robustness of the results to alternative modelling choices. Third, propose at least one modification or extension to the story and explore that proposition empirically.

## Schedule

### 1. AUGUST 31: WHAT MAKES (QUANTITATIVE) ANALYSES COMPELLING?

- **Discussion paper:** Sorenson, Olav, and Pino G. Audia. “The social structure of entrepreneurial activity: Geographic concentration of footwear production in the United States.” *American Journal of Sociology*, 106: 424-462
- **Discussion paper:** Saperstein, Alyia, and Andrew M. Penner (2012). “Racial fluidity and inequality in the United States.” *American Journal of Sociology*, 118: 676-727
- **Discussion paper:** Young, Cristobal, Charles Varner, Ithai Z. Lurie, and Richard Prisinzano (2016). “Millionaire migration and taxation of the elite: Evidence from administrative data” *American Sociological Review*, 81: 421-446
- **Background:** Gentzkow, Matthew, and Jesse M. Shapiro (2014). *Code and Data for the Social Sciences: A Practitioner’s Guide*. Available at: <https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf>

- *Recommended:* Healy, Kieran (2017). *The Plain Person's Guide to Plain Text Social Science*. Available at: <https://kieranhealy.org/resources/>

## 2. SEPTEMBER 5: TELLING THE STORY WITH GRAPHICS I

- **Pre-class assignment:** Find at least one graph from a published paper that you find convincingly illustrates the central claim of the paper. Please e-mail a PDF of the article to [olav.sorenson@yale.edu](mailto:olav.sorenson@yale.edu) by noon on September 5.
- *Recommended:* Tufte, Edward R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press

## 3. SEPTEMBER 7: TELLING THE STORY WITH GRAPHICS II

## 4. SEPTEMBER 12: WHAT ALLOWS FOR A CAUSAL INTERPRETATION?

- **Pre-class assignment:** Draw graphs of the causal structure for two of the three discussion papers from August 31. What would you consider the primary threat(s) to a causal interpretation of the results.
- **Background:** Angrist, Joshua D., and Jörn-Steffen Pischke (2008). “The experimental ideal.” Chapter 2 of *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press
- **Background:** Morgan, Stephen L., and Christopher Winship (2007). “Causal graphs, identification, and models of causal exposure.” Chapter 3 of *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press

## 5. SEPTEMBER 19: MEDIATION, MODERATION, AND PLACEBO TESTS

- **Background:** Morgan, Stephen L., and Christopher Winship (2007). “Mechanisms and causal explanation.” Chapter 8 of *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press
- **Background:** Pearl, Judea (2012). “The mediation formula: A guide to the assessment of causal pathways in nonlinear models” Chapter 12 of *Causality: Statistical Perspectives and Applications*. West Sussex: John Wiley & Sons

## 6. SEPTEMBER 26: MATCHING METHODS I

- **Discussion paper:** Eggers, J.P., and Lin Song (2015). “Dealing with failure: Serial entrepreneurs and the costs of changing industries between ventures.” *Academy of Management Journal*, 58: 1785-1803
- **Discussion paper:** Zhang, Jianjun, Christopher Marquis, and Kunyuan Qiao (2016). “Do political connections buffer firms from or bind firms to the government? A study of corporate charitable donations of Chinese firms.” *Organization Science*, 27: 1307-1324

- **Background:** Morgan, Stephen L., and Christopher Winship (2007). “Matching estimators of causal effects.” Chapter 4 of *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press
- **Background:** Iacus, Stefano M., Gary King, and Giuseppe Porro (2012). “Causal inference without balance checking: Coarsened exact matching.” *Political Analysis*, 20: 1-24
- *Recommended:* Rosenbaum, Paul R., and Donald B. Rubin (1983). “The central role of the propensity score in observational studies for causal effects.” *Biometrika*, 70: 41-55
- *Recommended:* Angrist, Joshua D., and Jörn-Steffen Pischke (2008). “Heterogeneity and nonlinearity.” Chapter 3.3 of *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press

#### 7. SEPTEMBER 28: MATCHING METHODS II

#### 8. OCTOBER 3: FIXED EFFECTS

- **Discussion paper:** Zhang, Letian (2017). “A fair game? Racial bias and repeated interaction between NBA coaches and players.” *Administrative Science Quarterly*, 62: in press
- **Background:** Angrist, Joshua D., and Jörn-Steffen Pischke (2008). “Individual fixed effects.” Chapter 5.1 of *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press

#### 9. OCTOBER 5: DIFFERENCES-IN-DIFFERENCES

- **Discussion paper:** Azoulay, Pierre, Joshua S. Graff, and Jialin Wang (2010). “Superstar extinction.” *Quarterly Journal of Economics*, 125: 549-589
- **Discussion paper:** McDonnell, Mary-Hunter, and Brayden King (2013). “Keeping up appearances: Reputational threat and impression management after social movement boycotts.” *Administrative Science Quarterly*, 58: 387-419
- **Background:** Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). “How much should we trust differences-in-differences estimates?” *Quarterly Journal of Economics*, 119: 249-276
- *Extended reference:* Athey, Susan, and Guido Imbens (2006). “Identification and inference in nonlinear difference-in-differences models.” *Econometrica*, 74: 431-497
- *Extended reference:* Abadie, Alberto, Alexis Diamond, and Jens Hainmueller (2010). “Synthetic control methods for comparative case studies: Estimating

the effect of California’s tobacco control program.” *Journal of the American Statistical Association*, 105: 493-505

#### 10. OCTOBER 10: INSTRUMENTAL VARIABLES I

- **Discussion paper:** Tabakovic, Haris, and Thomas Wollmann (2016). “The impact of money on science: Evidence from unexpected NCAA football outcomes.” Working paper, Harvard Business School
- **Discussion paper:** Kang, Jingoo, and Andy Y. Han Kim (2017). “The relationship between CEO media appearances and compensation.” *Organization Science*, 28: 379-394
- **Background:** Angrist, Joshua D., and Alan B. Krueger. “Instrumental variables and the search for identification: From supply and demand to natural experiments.” *Journal of Economic Perspectives*, 15: 69-85
- **Background:** Murray, Michael P. (2006). “Avoiding invalid instruments and coping with weak instruments.” *Journal of Economic Perspectives*, 20: 111-132
- *Recommended:* Morgan, Stephen L., and Christopher Winship (2007). “Instrumental variable estimators of causal effects.” Chapter 7 of *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press
- *Extended reference:* Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi “Plausibly exogenous.” *Review of Economics and Statistics*, 94: 260-272

#### 11. OCTOBER 12: INSTRUMENTAL VARIABLES II

#### 12. OCTOBER 19: REGRESSION DISCONTINUITY

- **Discussion paper:** Kerr, William R., Josh Lerner, and Antoinette Shoar (2014). “The consequences of entrepreneurial finance: Evidence from angel financings.” *Review of Financial Studies*, 27: 20-55
- **Discussion paper:** Ferguson, John-Paul (2015). “The control of managerial discretion: Evidence from unionization’s impact on employment segregation.” *American Journal of Sociology*, 121: 675-721
- **Background:** Angrist, Joshua D., and Jörn-Steffen Pischke (2008). “Getting a little jumpy: Regression discontinuity designs.” Chapter 6 of *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press
- *Recommended:* Imbens, Guido, and Thomas Lemieux (2008). “Regression discontinuity designs: A guide to practice.” *Journal of Econometrics*, 142: 615-635

13. OCTOBER 24: BINARY OUTCOMES I (LPMS, LOGIT, PROBIT)
- **Background:** Mood, Carina (2010). “Logistic regression: Why we cannot do what we think we can do, and what we can do about it.” *European Sociological Review*, 26: 67-82
  - **Background:** Angrist, Joshua D., and Jörn-Steffen Pischke (2008). “Limited dependent variables and marginal effects.” Chapter 3.4.2 of *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press
  - **Background:** Ai, Chunrong, and Edward C. Norton (2003). “Interaction terms in logit and probit models.” *Economics Letters*, 80: 123-129
  - *Extended reference:* Muris, Chris (2017). “Estimation in the fixed-effects ordered logit model.” *Review of Economics and Statistics*, 99: 465-477
14. OCTOBER 26: BINARY OUTCOMES II (LPMS, LOGIT, PROBIT)
15. OCTOBER 31: SURVIVAL ANALYSES I
- **Background:** Pötter, Ulrich, and Götz Rohwer (2007). Introduction to Event History Analysis. Ruhr University
  - *Recommended:* Cleves, Mario, William W. Gould, and Yulia V. Marchenko (2016). *An Introduction to Survival Analysis Using Stata, Revised Third Edition*. Stata Press
  - *Extended reference:* Strang, David, and Nancy Brandon Tuma (1993). “Spatial and temporal heterogeneity in diffusion.” *American Journal of Sociology*, 99: 614-639
16. NOVEMBER 2: SURVIVAL ANALYSES II
17. NOVEMBER 7: SURVIVAL ANALYSES III
18. NOVEMBER 9: COUNT MODELS I
- **Background:** King, Gary (1989). “Variance specification in event count models: From restrictive assumptions to a generalized estimator.” *American Journal of Political Science*, 33: 762-784
  - **Background:** Allison, Paul D., and Richard P. Waterman (2002). “Fixed effects negative binomial regression models.” *Sociological Methodology*, 32: 247-265
  - *Recommended:* Ver Hoef, Jay M., and Peter L. Boveng (2007). “Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data?” *Ecology*, 88: 2766-2772

- *Extended reference:* Mullahy, John (1997). “Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior.” *Review of Economics and Statistics*, 77: 586-593

19. NOVEMBER 14: COUNT MODELS II

20. NOVEMBER 28: MULTI-LEVEL MODELS I

- **Background:** Gelman, Andrew, and Jennifer Hill (2007). “Multilevel structures.” Chapter 11 of *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press
- **Background:** Gelman, Andrew, and Jennifer Hill (2007). “Multilevel linear models: the basics.” Chapter 12 of *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press
- **Background:** Gelman, Andrew, and Jennifer Hill (2007). “Multilevel linear models: varying slopes, non-nested models, and other complexities.” Chapter 13 of *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press

21. NOVEMBER 28: MULTI-LEVEL MODELS II

22. DECEMBER 5: MODEL UNCERTAINTY

- **Background:** Young, Cristobal, and Katherine Holsteen (2015). “Model uncertainty and robustness: A computational framework for multimodal analysis.” *Sociological Methods & Research*, 46: 3-40