

LONGITUDINAL DATA ANALYSIS

Oxford University, Nuffield College
Department of Politics and International Relations
Trinity Term 2019

Instructor: Andreas Wiedemann

Email: andreas.wiedemann@nuffield.ox.ac.uk

Lecture: Tuesday 2:30-4:30pm, Weeks 1-4, Nuffield College, Clay Room

Office hours: Monday 11:30am-12:30pm, Nuffield College, K1.

Course Description

This course offers an introduction to advanced methods of analyzing longitudinal and grouped data. We will cover advanced methods for panel data, including fixed and random effects models, non-parametric estimation techniques, difference-in-differences designs, and synthetic control methods. We also learn to understand grouped data structures such as multilevel and clustered data and estimation techniques such as mixed effect models and hierarchical models. The goal of this course is to understand, apply, and evaluate these methods that are essential for political science and public policy research. We will analyze the strengths and weaknesses of these methods using applications from political science and economics.

Prerequisites

- Basic level of calculus and linear algebra.
- Knowledge of probability theory, statistics, and linear regression covered in an introductory statistics or causal inference class.
- Familiarity with at least one statistical software. I will use R in this course (see below).

Statistical Programming

The course will be thought in R, an open-source statistical computing environment that is widely used in statistics and political science. You can download it for free from www.r-project.org. I also recommend using R Studio, an integrated development environment for R. You can download it [here](#).

The web provides many great tutorials and resources to learn R. A list of these is provided [here](#). A great way to start you off are the two video tutorials provided by Dan Goldstein [here](#) and also [here](#). Other good resources are the set of tutorials provided by [DataCamp](#) and, especially for data analyses, [UCLA idre](#).

If you are very familiar with another statistical software package you may use that for the course at your own risk. Sample code and support will be provided in R.

Assessment

Students taking the course for credit will submit an essay of no more than 2500 words, due at noon on Friday of 6th week (6th June) via email to the instructor. You can choose one of the following three options:

- A critical review of 2-4 papers using methods discussed in the course (possibly including replication)
- An in-depth replication and extension of a single paper using methods discussed in the course
- An original analysis using methods from the course

If you want to suggest another approach please contact me. A one-paragraph proposal of your final project is due in class on Tuesday, May 14th.

Readings

We will read chapters from the following books. I strongly recommend that you purchase the ones marked with (*). Other books cover particular sections of the course more in depth and are recommended for your reference, particularly if the sections are directly relevant for your final project or future research.

- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge; New York (*)
- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton (*)
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Mass
- Box-Steffensmeier, J. M., Freeman, J. R., Hitt, M. P., and Pevehouse, J. (2014). *Time Series Analysis for the Social Sciences*
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, Cambridge; New York

There are also a few papers that summarize some of the methods we will learn in this course (and many more). They are sometimes very detailed and technical, so you may not understand all of the details the first time. If you plan on applying some of these methods in the future, these papers might come in handy.

- Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86
- Beck, N. (2001). Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years? *Annual Review of Political Science*, 4(1):271–293
- Beck, N. and Katz, J. N. (2011). Modeling Dynamics in Time-Series–Cross-Section Political Economy Data. *Annual Review of Political Science*, 14(1):331–352

Topics and Tentative Course Schedule

The following is a preliminary schedule of course topics. Required readings are marked with a (*).

WEEK 1 | PANEL METHODS: FROM POOLED DATA TO FIXED EFFECTS

- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, Chapter 5.1 (*)
- La Ferrara, E., Chong, A., and Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, 4(4):1–31 (*)
- Boef, S. D. and Keele, L. (2008). Taking Time Seriously. *American Journal of Political Science*, 52(1):184–200
- Ladd, J. M. and Lenz, G. S. (2009). Exploiting a Rare Communication Shift to Document the Persuasive Power of the News Media. *American Journal of Political Science*, 53(2):394–410
- Acemoglu, D., Johnson, S., Robinson, J. A., and Yared, P. (2008). Income and Democracy. *American Economic Review*, 98(3):808–842
- Berrebi, C. and Klor, E. F. (2008). Are Voters Sensitive to Terrorism? Direct Evidence from the Israeli Electorate. *American Political Science Review*, 102(03):279–301

WEEK 2 | DIFFERENCE-IN-DIFFERENCES

- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, Chapter 5.2 (*)
- Card, D. and Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4):772–793 (*)
- Lyall, J. (2009). Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya. *Journal of Conflict Resolution*, 53(3):331–362

- Sances, M. W. (2015). The Distributional Impact of Greater Responsiveness: Evidence from New York Towns. *The Journal of Politics*, 78(1):105–119
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, 119(1):249–275
- Bai, Y. and Jia, R. (2016). Elite Recruitment and Political Stability: The Impact of the Abolition of China’s Civil Service Exam. *Econometrica*, 84(2):677–733

WEEK 3 | SYNTHETIC CONTROL METHODS

- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510 (*)
- Bohn, S., Lofstrom, M., and Raphael, S. (2013). Did the 2007 Legal Arizona Workers Act Reduce the State’s Unauthorized Immigrant Population? *The Review of Economics and Statistics*, 96(2):258–269 (*)
- 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
- Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The Value of Connections in Turbulent Times: Evidence from the United States. *Journal of Financial Economics*, 121(2):368–391
- Xu, Y. (2017). Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models. *Political Analysis*, 25(1):57–76

—Final Paper Proposal Due in class Tuesday, May 14th—

WEEK 4 | PANEL METHODS EXTENSIONS AND INTRODUCTION TO MULTILEVEL MODELS

- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge; New York, Chapters 11-13 (*)
- Clark, T. S. and Linzer, D. A. (2015). Should I Use Fixed or Random Effects? *Political Science Research and Methods*, 3(2):399–408 (*)
- Wawro, G. (2002). Estimating Dynamic Panel Data Models in Political Science. *Political Analysis*, 10(1):25–48
- Steenbergen, M. R. and Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1):218–237
- Caughey, D. and Warshaw, C. (2018). Public Opinion in Subnational Politics. *The Journal of Politics*, 81(1):352–363