

PLSC 504: “Topics in Political Methodology”

Fall 2024

Professor Christopher Zorn
Department of Political Science
Pennsylvania State University
E-mail: zorn@psu.edu
Wednesdays, 9:00 a.m. - 12:00 p.m.
Burrowes Building, Room 025

Course Description

This is the third (full) course in quantitative methods in Penn State’s political science Ph.D. program. The course is the “proseminar” for the methods field in the department; this means that it is designed to introduce students to topics that they then can learn about in greater depth in other courses. It begins with a brief review of regression-like statistical methods, mostly generalized linear models. We will also discuss models for time-series data, survival (time-to-event) data, and panel and time-series cross-sectional data. We will conclude by examining multivariate models for data reduction and the measurement of latent concepts, methods for making causal inferences with observational data, tools for network analysis, and statistical approaches for analyzing text.

The models discussed in this course are among the most widely used in the social sciences today. It is, effectively, not possible to function as an empirical social scientist without at least a passing familiarity with these models; moreover, given the rapid and increasing rate at which more advanced models are being adopted in these fields, these techniques increasingly represent a minimal level of statistical competence necessary to do publishable-quality quantitative work. In other words: knowing these models, and using them appropriately and well, can increase your odds of writing a strong (quantitative) dissertation, landing a job, publishing books and articles, being granted tenure, and generally leading a happier and more fulfilling professional life.

Much of the material in this course is fairly technical. While I have chosen readings that present the models as clearly and with as little jargon as possible, most of the material will still require several readings to fully comprehend. A solid understanding of calculus and linear algebra is required for this class, and the course assumes familiarity with linear and generalized linear regression models at the level of PLSC 503 (that is, at the level of Weisberg’s *Applied Linear Regression*, Greene’s *Econometric Analysis*, or the like). Students are also expected to have at least a nodding acquaintance with probability theory, statistical inference, and simple descriptive statistics, as well as with data management.

This syllabus is designed to provide an overview to the course. Clickable links are printed in [Penn State blue](#).

Course Readings

Recommended Text/Materials

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

AND/OR

Faraway, Julian J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Non-parametric Regression*. London: Chapman & Hall. (preferred)

AND

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. New York: Cambridge University Press.

AND

Everitt, Brian, and Thorsen Hothorn. 2011. *An Introduction to Applied Multivariate Analysis with R*. New York: Springer

The course will also draw on additional readings as necessary, all of which will be available on the course [github repository](#) and/or through JSTOR. For the required texts, you should feel free to order them from whatever sources you deem best.

A Few Other Potentially Useful Readings

... can be found in a list, [here](#).

Some Other Useful Resources

The **Political Methodology Section** of the American Political Science Association was created to provide APSA members with an interest in political methodology with a forum in which to meet and discuss ideas. The section publishes blog-ish thing (*The Political Methodologist*), a quarterly journal on political methodology (*Political Analysis*), and conducts a [discussion list](#) on topics relating to political methodology. They also support the [International Methods Colloquium](#), as well as a host of other activities.

Also, the **Inter-University Consortium for Political and Social Research** (ICPSR), at the University of Michigan, maintains an extensive archive of data in the social and behavioral sciences. Much of it is accessible via their homepage.

Grading

Grading will be based on ten(ish) more-or-less weekly homework exercises (50 points each) and a final paper/poster presentation (500 points). In most instances, exercises will be due one week from being assigned.

Homework exercises will generally involve both simulation-based work and estimation and interpretation of models on real/existing data, using statistical computer software (see below). Feel free to work on the assignments in groups of two or three, but you must write up all assignments individually. Details for the homework assignments and the final project will be announced in class.

Also, note that homework exercises and the final paper/poster should be submitted as electronic files, **in Adobe PDF format**. If you do not know how to create a PDF file, please go learn.

Software, Statistical and Otherwise

You are welcome to make use of whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions reached are transparent. However, due to the limits of instructor and TA time and patience, we will support only two software packages. Both are available to PSU students; contact CLA IT for details.

R

R is a statistical environment and high-level programming language for data analysis and display. It is effectively the GNU version of the **S** language; as such, it is free (both as in speech and as in beer) and open source. The current (late August 2024) version of **R** is 4.4.1. **R** is an *object-oriented* language; unlike **Stata** (and most other statistical software), it operates by assigning values to objects in the workspace. In the notes, handouts, etc., **R** commands will be preceded by a caret (“>”):

```
> my.results<-lm(Y~X)
```

The [Comprehensive R Archive Network](#) (CRAN) is the go-to spot for all things **R**-related. I cannot begin to list all the **R**-related resources available on the web; for newbies, however, it might be useful to check out the [Introduction to R](#), [this page](#) in getting data into **R**, and the various **R** “cheat sheets” [here](#), [here](#), and [here](#). **Stata** users who are interested in learning **R** should check out the [Moving from Stata to R](#) page at the **R** Project’s [wiki](#).

All in-class examples, code, graphics, and so forth will use R.

Stata

Once upon a time (not that long ago), **Stata** was the most widely-used statistical package in the social sciences. (It remains the most widely used statistical software in the discipline of economics, which... is a thing.) It is a powerful tool for data management, analysis, and display, and boasts some of the best manuals and on-line help of any existing software package. **Stata** is commercial software; the current version of **Stata** is 18.0, but previous versions (back to v. 14, at least) can also be used for the class. In the exceedingly rare instance when they appear in the class notes, handouts, etc., **Stata** commands will appear in a fixed-width font and will be preceded by a period (“.”):

```
. regress Y X
```

Stata newbies may want to check out:

Accock, Alan. 2023. *A Gentle Introduction to Stata*, Revised Sixth Edition. College Station, TX: Stata Press.

and/or Stata's dedicated "new users" page:

<https://www.stata.com/links/resources-for-learning-stata/>.

Beyond this, the [Stata](#) homepage is a valuable resource for questions about the Stata statistical software. There are a number of useful Stata references on the web, including [Scott Long's page](#) at IU and an excellent Stata "help page" sponsored by UCLA.

Other Software Considerations

In no particular order:

- For many years, your instructor did not have a formally-stated preference for either Stata or R. That time has passed: If you want to do professional-level data analysis in political science, and/or be competitive for jobs in industry, government, and the non-profit sector, learn R, not Stata. Be aware that Stata has a far flatter learning curve than R, which means people tend to gravitate toward it given a choice. But R is far more flexible and powerful, and will likely be more valuable and useful to you in the long run.
- Learn to use \LaTeX , now, while you have the time. You will be glad you did. If you're struggling with \LaTeX and would like some example documents, let me know.
- If you insist on using Microsoft Word (or any other WYSIWYG editing program) for writing assignments, papers, etc., **do not under any circumstances cut and paste graphs from Stata and R into those programs**. Save whatever figures you want to use as .pdf, .png, .tif, or .jpg files, and import them into the software.
- Similarly: Document creation and editing software (\LaTeX , Word, etc.) are not suitable for distribution of work. Unless the person / people reading them will also be editing them, convert all documents to PDFs before sending / submitting / etc. This includes homework assignments and class papers in PLSC 504.

Obligatory Statement on Academic Integrity

Academic integrity is the pursuit of scholarly activity in an open, honest and responsible manner. Academic integrity is a basic guiding principle for all academic activity at The Pennsylvania State University, and all members of the University community are expected to act in accordance with this principle. Consistent with this expectation, the University's Code of Conduct states that all students should act with personal integrity, respect other students' dignity, rights and property, and help create and maintain an environment in which all can succeed through the fruits of their efforts.

Academic integrity includes a commitment by all members of the University community not to engage in or tolerate acts of falsification, misrepresentation or deception. Such acts of dishonesty violate the fundamental

ethical principles of the University community and compromise the worth of work completed by others.

In cases of any violation of academic integrity it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts. More information on academic integrity and procedures followed for violation can be found [here](#).

Obligatory Statement on Accommodations for Disabilities

Penn State welcomes students with disabilities into the University's educational programs. Every Penn State campus has an office for students with disabilities. Student Disability Resources (SDR) website provides contact information for every Penn State campus (<http://equity.psu.edu/sdr/disability-coordinator>). For further information, please visit the Student Disability Resources website (<http://equity.psu.edu/sdr/>).

In order to receive consideration for reasonable accommodations, you must contact the appropriate disability services office at the campus where you are officially enrolled, participate in an intake interview, and provide documentation: See documentation guidelines at <http://equity.psu.edu/sdr/guidelines>. If the documentation supports your request for reasonable accommodations, your campus disability services office will provide you with an accommodation letter. Please share this letter with your instructors and discuss the accommodations with them as early as possible. You must follow this process for every semester that you request accommodations.

Obligatory Statement on Counseling and Psychological Services

Many students at Penn State face personal challenges or have psychological needs that may interfere with their academic progress, social development, or emotional wellbeing. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients' cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

Counseling and Psychological Services at University Park (CAPS)
(<http://studentaffairs.psu.edu/counseling/>): 814-863-0395

Counseling and Psychological Services at Commonwealth Campuses
(<http://senate.psu.edu/faculty/counseling-services-at-commonwealth-campuses/>)

Penn State Crisis Line (24 hours / 7 days/week): 877-229-6400. Crisis Text Line (24 hours / 7 days/week): Text LIONS to 741741.

Obligatory Statement on Educational Equity and Reporting Bias

Penn State takes great pride to foster a diverse and inclusive environment for students, faculty, and staff. Consistent with University Policy AD29, students who believe they have experienced or observed a hate

crime, an act of intolerance, discrimination, or harassment that occurs at Penn State are urged to report these incidents as outlined on the University's Report Bias webpage (<http://equity.psu.edu/reportbias/>).

Obligatory Statement on Religious Observances

The [Religious and Spiritual Observances Calendar](#) is compiled by the Center for Spiritual and Ethical Development in consultation with campus and community religious leaders. It specifies those holy days of the major world religions for which observance may require students to depart from their normal routine at the University. Please note that only those holy days which occur when Penn State classes are in session are listed. This is not, therefore, an exhaustive list of all major holy days in each religious tradition.

Non-Obligatory Statement on Generative AI, Large Language Models, etc.

You're undoubtedly well aware of the existence of large language models (LLMs) – e.g., [ChatGPT](#) – and other artificial intelligence (AI) tools for language / image creation. Having been described as everything from [making everyone their own version of Tony Stark](#) to a [Lovecraftian shoggoth](#) (and [most things in between](#)), LLMs are currently creating a sometimes-depressing, sometimes-hilarious panic among faculty in legacy academic disciplines and programs. Most of that panic revolves around the use of LLMs to “cheat,” in the traditional sense: to create work that deceptively gives the impression that the student knows something they do not. Beyond its intrinsically duplicitous nature, such use in a conventional classroom setting gives rise to concerns about equity and (potentially) devalues the experience / credential for other class members.

The other side of the LLM equation is that they are powerful tools for augmenting learning and creating new knowledge. Experience suggests that, in line with other technological advances (the printing press, personal computers, search engines, etc.), it is wiser to adapt to LLMs than to attempt to limit or ban their use. This is especially true in a course like this one, where (a) LLMs are particularly useful tools for learning technical skills (e.g., the R programming language) and (b) the long-term, repeated nature of graduate school creates disincentives for “cheating” in a conventional sense. Accordingly, enrollees in PLSC 502 are welcome to use generative AI tools, such as ChatGPT, to assist them with their work in the course. In doing so, it is important to remember that such AI tools are capable of making errors, and that it is each student's responsibility to verify the information they receive from the such a tool. In addition, any information obtained from a generative AI source must be noted/cited in the student's work, just as they would cite any other source.

In addition, LLMs are among the most useful tools for quickly and efficiently writing computer code. Consider familiarizing yourself with one of the good LLM-based code helpers (e.g., [CoPilot](#), [SuperMaven](#), etc.) or (when you're stuck) just ask a general-use LLM (like [ChatGPT](#)) to help you out. If/when you do so, bear a couple things in mind:

- Many of these tools are better at more orthodox languages from computer / information science (e.g., Python, Java, etc.) than they are at writing R code.
- Relatedly: Such tools are usually good either for very basic things, or for getting a “starter” block of code that you can then fix, modify, etc. In my experience, it's pretty rare for (say) ChatGPT to give you the ideal code for your task at hand the very first time you ask for it.

Course Schedule

Readings should be completed prior to coming to class on the assigned day. Note that we will not, in general, hew closely (or at all) to the readings themselves, other than topically. Links are generally to DOIs or to stable PDFs at JSTOR.

As a rule, we will cover one broad topic per week. Readings will be assigned weekly. It will generally be a good idea to have completed the assigned readings prior to Monday's class each week.

Part One: Regression-Like Models and Causal Inference

August 28: Introduction + Overview / Review: Maximum Likelihood

- Readings

- Required:

- Long, Chapter 2 (esp. pp. 25-33), pp. 52-61, and Chapter 4.
 - Faraway, pp. 279-285 (review).
 - Buse, A. 1982. [“The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note.”](#) *The American Statistician* 36(3):153-57.
 - Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. [“What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory.”](#) *American Sociological Review* 86:532-565.

- Suggested:

- Eliason (1993), pp. 1-28.
 - Greene (2003), §17.4.
 - Breusch, T. S. 1979. [“Conflict Among Criteria for Testing Hypotheses: Extensions and Comments.”](#) *Econometrica* 47(1):203-07.
 - Greene (2003), pp. 484-496.
 - Meeker, William Q. and Luis A. Escobar. 1995. [“Teaching About Approximate Confidence Regions Based on Maximum Likelihood Estimation.”](#) *The American Statistician* 49(1):48-53.

September 4: Regression Models for Binary Outcomes

- Readings

- Required:

- Long, pp. 34-112.
 - Faraway, pp. 25-38 (review).
 - King, Gary and Langche Zeng. 2001. [“Logistic Regression in Rare Events Data.”](#) *Political Analysis* 9(2):137-63.

- Puhr, Rainer, Georg Heinze, Mariana Nold, Lara Lusa, and Angelika Geroldinger. 2017. “Firth’s Logistic Regression With Rare Events: Accurate Effect Estimates and Predictions?” *Statistics in Medicine* 36:2302–2317.
- Zorn, Christopher. 2005. “A Solution to Separation in Binary Response Models.” *Political Analysis* 13(2):157-70.
- *Suggested:*
 - Alvarez, R. Michael and John Brehm. 1995. “American Ambivalence Toward Abortion Policy: A Heteroskedastic Probit Method for Assessing Conflicting Values.” *American Journal of Political Science* 39(4):1055-82.
 - Berry, William D., Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. “Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?” *American Journal of Political Science* 54(January): 248-66.
 - Freedman, David A., and Jasjeet S. Sekhon. 2010. “Endogeneity in Probit Response Models.” *Political Analysis* 18(2):138-50.
 - Hagle, Timothy M., and Glenn E. Mitchell. 1992. “Goodness of Fit Measures for Probit and Logit.” *American Journal of Political Science* 36(3):762-84.
 - Herron, Michael C. 2000. “Postestimation Uncertainty in Limited Dependent Variable Models.” *Political Analysis* 8(1):83-98.
 - King, Gary and Langche Zeng. 2001. “Explaining Rare Events in International Relations.” *International Organization* 55(3):693-715.
 - Kosmidis, Ioannis, and David Firth. 2009. “Bias Reduction in Exponential Family Nonlinear Models.” *Biometrika* 96:793-804.

● **Exercise One: Estimate and interpret some nominal-response models.**

September 11: Regression Models for Nominal Responses (+ C-SODA talk)

● **Readings**

- *Required:*
 - Long, pp. 148-185.
- *Suggested:*
 - Alvarez, R. Michael, and Jonathan Nagler. 1998. “When Politics and Models Collide: Estimating Models of Multiparty Elections.” *American Journal of Political Science* 42(1):55-97.
 - Ashford, J. R. and R. R. Sowden. 1970. “Multi-variate Probit Analysis.” *Biometrics* 26(3):535-46.
 - Dow, Jay K., and James W. Endersby. 2004. “Multinomial Probit and Multinomial Logit: A Comparison of Choice Models for Voting Research.” *Electoral Studies* 23(1):107-22.
 - Fry, Tim R., and Mark N. Harris. 1998. “Testing for Independence of Irrelevant Alternatives: Some Empirical Results.” *Sociological Methods and Research* 26(3):401-23.

- Glasgow, Garrett. 2001. “Mixed Logit Models for Multiparty Elections.” *Political Analysis* 9(2):116-36.
- Lu, Xiao. 2020. “Discrete Choice Data with Unobserved Heterogeneity: A Conditional Binary Quantile Model.” *Political Analysis* 28:147-167.
- Patty, John W., and Elizabeth M. Penn. 2019. “A Defense of Arrow’s Independence of Irrelevant Alternatives.” *Public Choice* 179:145-164.
- Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1999. “Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models.” *American Journal of Political Science* 43(4):1231-47.
- Seshadri, Arjun, and Johan Ugander. 2020. “Fundamental Limits of Testing the Independence of Irrelevant Alternatives in Discrete Choice.” Working paper, arXiv:2001.07042.

• **Exercise Two: Fit and compare some models for nominal outcomes.**

September 18: Sample Selection + Introduction to Potential Outcomes

• **Readings**

◦ *Required:*

- Berk, R. A. 1983. “An Introduction to Sample Selection Bias in Sociological Data.” *American Sociological Review* 48(June):386-398.
- Little, Roderick J. A., and Donald B. Rubin. 2000. “Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytical Approaches.” *Annual Review of Public Health* 21:121-145.
- Stolzenberg, Ross M., and Daniel A. Relles. 1997. “Tools for Intuition about Sample Selection Bias and Its Correction.” *American Sociological Review* 62:494-507.

◦ *Suggested:*

- Heckman, James J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47(January):153-161.
- Leeman, Lucas. 2014. “Strategy and Sample Selection: A Strategic Selection Estimator.” *Political Analysis* 22(3):374-397.
- Pearl, Judea. 2000. *Causality*. New York: Cambridge University Press.
- Sartori, Anne E. 2003. “An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions.” *Political Analysis* 11(2):111-138.
- Sekhon, Jasjeet S. 2004. “Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals.” *Perspectives on Politics* 2(2):281-293.
- Sigelman, Lee, and Langche Zeng. 1999. “Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models.” *Political Analysis* 8(2):167-82.
- Vella, Francis. 1998. “Estimating Models with Sample Selection Bias: A Survey.” *The Journal of Human Resources* 33:127-169.

September 25: Causal Inference with Observational Data

• Readings

◦ Required:

- Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59:391-425.
- Heinze, Georg, and Peter Juni. 2011. "An Overview Of The Objectives Of And The Approaches To Propensity Score Analyses." *European Heart Journal* 32:1704-1708.
- Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15:199-236.
- Keele, Luke, Randolph T. Stevenson, and Felix Elwert. 2020. "The Causal Interpretation Of Estimated Associations In Regression Models." *Political Science Research and Methods* 8:1-13.
- Titunik, Rocio, and Jasjeet Sekhon. 2016. "Understanding Regression Discontinuity Designs as Observational Studies." *Observational Studies* 2:174-182.

◦ Suggested:

- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables (with discussion)." *Journal of the American Statistical Association* 91(434):444-472.
- Braumoeller, Bear F., Giampiero Marra, Rosalba Radice, and Aisha E. Bradshaw. 2018. "Flexible Causal Inference for Political Science." *Political Analysis* 26(1):54-71.
- de la Cuesta, Brandon, and Kosuke Imai. 2016. "Misunderstandings About the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19(1):375-396.
- Eggers, Andrew C., Guadalupe Tuñón, and Allan Dafoe. 2023. "Placebo Tests for Causal Inference." *American Journal of Political Science*: forthcoming.
- Hartman, Erin. 2021. "Equivalence Testing for Regression Discontinuity Designs." *Political Analysis* 29:505-521.
- Imbens, Guido. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *The Review of Economics and Statistics* 86(1):4-29.
- Marshall, John. 2022. "Can Close Election Regression Discontinuity Designs Identify Effects of Winning Politician Characteristics?" *American Journal of Political Science* 66: forthcoming.
- Rosenbaum, Paul and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.
- Rubin, Donald B. 1973. "Matching to Remove Bias in Observational Studies." *Biometrics* 29(1):159-183.
- Rubin, Donald B., and Neal Thomas. 1996. "Matching Using Estimated Propensity Scores: Relating Theory to Practice." *Biometrics* 52(1):249-264.
- Sekhon, Jasjeet S. 2007. "Matching: Multivariate and Propensity Score Matching with Automated Balance Search." *Journal of Statistical Software*.

- Sekhon, Jasjeet S. 2009. “[Opiates for the Matches: Matching Methods for Causal Inference.](#)” *Annual Review of Political Science* 12:487-508.
- Thistlethwaite, D., and D. Campbell. 1960. “Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment.” *Journal of Educational Psychology* 51:309-317.

- **Exercise Three: Estimate causal effects on observational data, using matching techniques.**

Part Two: Longitudinal Data

October 2: A Brief Introduction to Time Series Analysis

- **Readings**

- *Required:*

- Pickup, Mark. 2015. *Introduction to Time Series Analysis*. Sage Publications. Chapters 1-3.
- Keele, Luke and Nathan Kelly. 2006. “Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables.” *Political Analysis* 14:186-205.
- Murray, Michael P. 1994. “A Drunk and Her Dog: An Illustration of Cointegration and Error Correction.” *The American Statistician* 48(February):37-39.
- Clayton Webb, Suzanna Linn, and Matthew J. Lebo. 2020. “[Beyond the Unit Root Question: Uncertainty and Inference.](#)” *American Journal of Political Science* 64:275-292.

- *Suggested:*

- Beck, Neal. 1991. “Comparing Dynamic Specifications: The Case of Presidential Approval.” *Political Analysis* 3:51-87.
- De Boef, Suzanna and Luke Keele. 2008. “Taking Time Seriously.” *American Journal of Political Science* 52:184-200.
- Durr, Robert. 1993. “An Essay on Cointegration and Error Correction Models.” *Political Analysis* 5:185-228, and discussion by Williams, Beck, and Smith.
- Freeman, James R. 1983. “Granger Causality and the Time Series Analysis of Political Relationships.” *American Journal of Political Science* 27:327-58.
- Grant, Taylor, and Matthew J. Lebo. 2016. “Error Correction Methods with Political Time Series.” *Political Analysis* 24:3-30.
- Keele, Luke, Suzanna Linn, and Clayton M. Webb. 2016. “Treating Time with All Due Seriousness.” *Political Analysis* 24:31-41.
- Philips, Andrew Q. 2017. “Have Your Cake and Eat It Too? Cointegration and Dynamic Inference from Autoregressive Distributed Lag Models.” *American Journal of Political Science* 61:1-15
- Stroe-Kunold, Esther, and Joachim Werner. 2009. “A Drunk and her Dog: A Spurious Relation? Cointegration Tests as Instruments to Detect Spurious Correlations Between Integrated Time Series.” *Quality and Quantity* 43:913-940.

- **Exercise Four: Fit and interpret some time series models.**

October 9: Panel/TSCS: Unit Effects and Dynamics

• Readings

◦ Required:

- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (And Not To Do) With Time-Series Cross-Section Data." *American Political Science Review* 89(September): 634-647.
- Clark, Tom S. and Drew A. Linzer. 2015. "Should I Use Fixed Or Random Effects?" *Political Science Research and Methods* 3(2):399-408.
- Imai, Kosuke, and In Song Kim. 2021. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29:405-415.
- Keele, Luke, and Nathan J. Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14(2):186-205.
- Kropko, Jonathan, and Robert Kubinec. 2020. "Interpretation and Identification of Within-Unit and Cross-Sectional Variation in Panel Data Models." *PLoS ONE* 15(4): e0231349. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231349>
- Mummolo, Jonathan, and Erik Peterson. 2018. "Improving the Interpretation of Fixed Effects Regression Results." *Political Science Research and Methods* 6:829-835.

◦ Suggested:

- Bartels, Larry M. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40(August):905-42.
- Beck, Nathaniel, and Jonathan Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14:331-52.
- Blackwell, Matthew, and Adam N. Glynn. 2018. "How to Make Causal Inferences with Time-Series Cross-Sectional Data under Selection on Observables." *American Political Science Review* 112(4):1067-1082.
- Collischon, Matthias, and Andreas Eberl. 2020. "Let's Talk About Fixed Effects: Let's Talk About All the Good Things and the Bad Things." *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 72:289-299.
- Cook, Scott J., Jude C. Hays, and Robert J. Franzese. 2020. "Fixed Effects In Rare Events Data: A Penalized Maximum Likelihood Solution." *Political Science Research and Methods* 8:92-105.
- Gassen, Joachim. 2020. "Using ExPanD for Panel Data Exploration." R vignette: https://cran.r-project.org/web/packages/ExPanDaR/vignettes/use_ExPanD.html (accessed 31 May 2022).
- Hsiao, Cheng. 2003. *Analysis of Panel Data*. Chapters 1 and 3.
- Liu, Licheng, Ye Wang, Yiqing Xu. 2020. "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data." Working paper: Stanford University.
- Neuhaus, J. M., and J. D. Kalbfleisch. 1998. "Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data." *Biometrics* 54:638-45.
- Nuamah, Nicholas N. N. 1986. "Pooling Cross Section and Time Series Data." *The Statistician* 35:345-51.

- Pickup, Mark, and Vincent Hopkins. 2020. “Transformed-Likelihood Estimators for Dynamic Panel Models with a Very Small T .” *Political Science Research Methods* 7:forthcoming.
- Plumper, Thomas, and Vera E. Troeger. 2007. “Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects.” *Political Analysis* 15(2):124-139.
- Plümper, Thomas, and Vera E. Troeger. 2019. “Not So Harmless After All: The Fixed-Effects Model.” *Political Analysis* 27:21-45.
- Stimson, James. 1985. “Regression in Space and Time: A Statistical Essay.” *American Journal of Political Science* 29:914-47.
- “Symposium on Fixed-Effects Vector Decomposition.” 2011. *Political Analysis* 19(2).
- Wawro, Gregory. 2002. “Estimating Dynamic Panel Data Models in Political Science.” *Political Analysis* 10(Winter):25-48.
- Wilson, Sven E., and Daniel M. Butler. 2007. “A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications.” *Political Analysis* 15(2):101-123.
- Xu, Yiqing. 2017. “Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models.” *Political Analysis* 25:57-76.
- Zorn, Christopher. 2001. “Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes.” *International Interactions* 27(4):433-45.

October 16: Panel / TSCS Models for Binary, Count, and Other Odd-Looking Responses

• Readings

◦ Required:

- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science* 42(October):1260-88.
- Cameron, A. Colin, and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press. Chapter 9.
- Hsiao, Cheng. 2003. *Analysis of Panel Data*. Chapter 7, §7.1-7.3 and Chapter 8 (or Croissant and Milla, *Panel Data Econometrics with R*, Chapter 9).
- Neuhaus, J. M., J. D. Kalbfleisch, and W. W. Hauck. 1991. “A Comparison of Cluster-Specific and Population-Averaged Approaches for Analyzing Correlated Binary Data.” *International Statistical Review* 59(1):25-35.
- Zorn, Christopher. 2001. “Generalized Estimating Equation Models for Correlated Data: A Review with Applications.” *American Journal of Political Science* 45(April):470-90.

◦ Recommended:

- Ballinger, Gary A. 2004. “Using Generalized Estimating Equations for Longitudinal Data Analysis.” *Organizational Research Methods* 7:127-50.

- Cook, Scott J., Jude C. Hays, and Robert J. Franzese. 2020. “Fixed Effects In Rare Events Data: A Penalized Maximum Likelihood Solution.” *Political Science Research and Methods* 8:92-105.
- Crisman-Cox, Casey. 2021. “Estimating Substantive Effects in Binary Outcome Panel Models: A Comparison.” *The Journal of Politics* 83:532-546.
- Green, Donald P., Soo Yeon Kim, and David Yoon. 2001. “Dirty Pool.” *International Organization* 55:441-68 (and commentary by Beck & Katz, Oneal & Russett, and King).
- Katz, Ethan. 2001. “Bias in Conditional and Unconditional Fixed Effects Logit Estimation.” *Political Analysis* 9(Autumn):379-84 (and also see Coup’e, Tom (2005) “Bias in Conditional and Unconditional Fixed Effects Logit Estimation: A Correction.” *Political Analysis* 13(Summer):292-95).
- Korre, A.K., and V.G.S. Vasdekis. 2018. “Goodness of Fit Tests for Random Effect Models with Binary Responses.” *Statistics in Medicine*. Forthcoming.
- Li, Quan, and Drew Schaub. 2004. “Economic Globalization and Transnational Terrorism: A Pooled Time-Series Analysis.” *Journal of Conflict Resolution* 48:230-258.
- Martin, Andrew D. 2003. “Bayesian Inference for Heterogeneous Event Counts.” *Sociological Methods and Research* 32:30-63.
- McGrath, Liam F. 2015. “Estimating Onsets of Binary Events in Panel Data.” *Political Analysis* 23:534-549.
- Wawro, Gregory. 2001. “A Panel Probit Analysis of Campaign Contributions and Roll Call Votes.” *American Journal of Political Science* 45(July):563-579.
- Whitford, Andrew B., Jeff Yates, and Holona L. Ochs. 2006. “Ideological Extremism and Public Participation.” *Social Science Quarterly* 87(1):36-54.
- Wooldridge, Jeffrey. 1999. “Distribution-Free Estimation of Some Nonlinear Panel Data Models.” *Journal of Econometrics* 90(May):77-97.

• **Exercise Five: Fit and discuss some panel / TSCS data models.**

October 23: Survival Analysis

• **Readings**

○ *Required:*

- Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*, Chapters 3-5 and 8-11.

○ *Suggested:*

- Alt, James E., Gary King and Curtis S. Signorino. 2001. “Aggregation Among Binary, Count and Duration Models: Estimating the Same Quantities from Different Levels of Data.” *Political Analysis* 9(Winter):21-44.
- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.” *American Journal of Political Science* 42(October):1260-88 (and [erratum](#)).

- Box-Steffensmeier, Janet M., Suzanna L. De Boef and Kyle A. Joyce. 2007. "Event Dependence and Heterogeneity in Duration Models: The Conditional Frailty Model." *Political Analysis* 15(3):237-256.
- Box-Steffensmeier, Janet M., and Christopher Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(October):951-67.
- Box-Steffensmeier, Janet M., and Christopher Zorn. 2002. "Duration Models for Repeated Events." *Journal of Politics* 46(November):1069-94.
- Cox, David Roxbee. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Series B* 34(2):187-220.
- Signorino, Curt, and David Carter. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis* 18(3):271-292. Also read response by Beck and rejoinder by Signorino & Carter.
- Desmarais, Bruce A., and Jeffrey J. Harden. 2012. "Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model." *Political Analysis* 20(1):113-135. DOI:10.1093/pan/mpr042
- Licht, Amanda A. 2011. "Change Comes with Time: Substantive Interpretation of Nonproportional Hazards in Event History Analysis." *Political Analysis* 19(2):227-243.
- Metzger, Shawna K., and Benjamin T. Jones. 2016. "Surviving Phases: Introducing Multi-state Survival Models." *Political Analysis* 24(4):457-477.
- Pintilie, Melania. 2007. "Analysing and Interpreting Competing Risk Data." *Statistics in Medicine* 26:1360-67.
- Tsodikov, A. 1998. "A Proportional Hazards Model Taking Account of Long Term Survivors." *Biometrics* 54:1508-15.
- Wolbers, Marcel, et al. 2014. "Competing Risks Analyses: Objectives and Approaches." *European Heart Journal* 35:2936-2941.
- Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8(Autumn): 367-380.

- **Exercise Six: Fit and interpret some survival models.**

Part Three: Measurement

October 30: Principal Components, Factor Analysis, and Clustering

- **Readings**

- *Required:*

- Ahlquist, John, and Christian Breunig. 2012. "Model-based Clustering and Typologies in the Social Sciences." *Political Analysis* 20:92-112.
- Everitt and Hothorn 2011. Chapters 3, 5, and 6. (Also scan chapter 7, if you have the time.)
- Greenacre, Michael. 2012. "Biplots: The Joy of Singular Value Decomposition." *Wiley Interdisciplinary Reviews: Computational Statistics* 4:399-406.

- Henson, Robin K., and J. Kyle Roberts. 2006. "Use of Exploratory Factor Analysis in Published Research: Common Errors and Some Comment on Improved Practice." *Educational and Psychological Measurement* 66:393-416.
- *Suggested:*
 - Flora, David B., and Jessica K. Flake. 2017. "The Purpose and Practice of Exploratory and Confirmatory Factor Analysis in Psychological Research: Decisions for Scale Development and Validation." *Canadian Journal of Behavioural Science* 49:78-88.
 - Gabriel, K.R. 1971. "The Biplot Graphic Display of Matrices with Application to Principal Components Analysis." *Biometrics* 58:453-467.
 - Greenacre, Michael J., and Patrick J. F. Groenen. 2016. "Weighted Euclidean Biplots." *Journal of Classification* 33:442-459.
 - Jakulin, Alecs, W. Buntine, T. Pira, and H. Brasher. 2009. "Analyzing the U.S. Senate in 2003: Similarities, Clusters, and Blocs." *Political Analysis* 17:291-310.
 - MacCallum, R.C. 1974. "Relations Between Factor Analysis and Multidimensional Scaling." *Psychological Bulletin* 81: 505-516.
 - Ristei Gugiu, M., and M. Centellas. 2013. "The Democracy Cluster Classification Index." *Political Analysis* 21:334-349.

● **Exercise Seven: Do a little bit of multivariate statistics.**

November 6: Scaling and Item Response Theory

● **Readings**

- *Required:*
 - Hambleton et al. (1991), pp. 7-46, 53-88, 109-122.
 - Hout, Michael C., Megan H. Papesh, and Stephen D. Goldinger. 2013. "Multidimensional Scaling." *WIREs Cognitive Science* 4: 93-103.
 - Mair, Patrick, and Jan De Leeuw. 2015. "Unidimensional Scaling." In *Wiley StatsRef: Statistics Reference Online*. New York: Wiley. pp. 1-3.
 - Sijtsma, Klaas. 2009. "On the Use, Misuse, and the Very Limited Usefulness of Cronbach's Alpha." *Psychometrika* 74:107-120.
- *Suggested:*
 - Bakker, Ryan, and Keith Poole. 2013. "Bayesian Metric Multidimensional Scaling." *Political Analysis* 21:125-140.
 - Coombs, Clyde H. 1950. "Psychological Scaling Without a Unit of Measurement." *Psychological Review* 57: 145-158.
 - De Leeuw, J., and P. Mair. 2009. "Multidimensional Scaling Using Majorization: SMA-COF in R." *Journal of Statistical Software* 31:1-30.
 - De Santis, Gustavo, Mauro Maltagliati, and Silvana Salvini. 2016. "A Measure of the Cultural Distance Between Countries." *Social Indicators Research* 126:1065-1087.

- Lord, Frederic M. 1983. “Unbiased Estimates of Ability Parameters, of Their Variance, and of Their Parallel Forms Reliability.” *Psychometrika* 48:477-82.
- Martin, Andrew D., Kevin M. Quinn, and Jong Hee Park. “MCMCpack: Markov Chain Monte Carlo Package.”
- Poole, Keith, and Howard Rosenthal. 1985. “A Spatial Model of Legislative Roll Call Analysis.” *American Journal of Political Science* 29(2):357-384
- Poole, Keith T. 1984. “Least Squares Metric, Unidimensional Unfolding.” *Psychometrika* 49: 311-323.
- Poole, Keith. 2005. *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press.
- Rasch, Georg. 1961. “On General Laws and the Meaning of Measurement in Psychology.” *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability* 4:321-333.
- Rizopoulos, Dimitris. 2006. “ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses.” *Journal of Statistical Software* 17(5).
- Young, Forrest W. 1984. “Scaling.” *Annual Review of Psychology* 35: 55-81.

- **Exercise Eight: Fit and discuss some item response models.**

November 13: Network Analysis

- **Readings**

- *Required:*

- Barabási, Albert-László, and Reka Albert. 1999. “Emergence Of Scaling In Random Networks.” *Science* 286:509-512.
- Leenders, Roger Th.A.J. 2002. “Modeling Social Influence Through Network Autocorrelation: Constructing The Weight Matrix.” *Social Networks* 24:21-47.
- Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. “Homophily And Contagion Are Generically Confounded In Observational Social Network Studies.” *Sociological Methods and Research* 40:211-239.
- Wasserman, Stanley, and Philippa Pattison. 1996. “Logit Models And Logistic Regressions For Social Networks: I. An Introduction To Markov Graphs And p^* .” *Psychometrika* 61:401-425.

- *Suggested:*

- Bisbee, James, and Jennifer M. Larson. 2017. “Testing Social Science Network Theories With Online Network Data: An Evaluation Of External Validity.” *American Political Science Review* 111:502-521.
- Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos. 2013. “Reasoning About Interference Between Units: A General Framework.” *Political Analysis* 21:97-124.
- Desmarais, Bruce A., and Skyler J. Cranmer. 2017. “Statistical Inference In Political Networks Research.” In *The Oxford Handbook of Political Networks*. New York: Oxford University Press.

- Garlaschelli, Diego, and Maria I. Loffredo. 2004. "Patterns Of Link Reciprocity In Directed Networks." *Physical Review Letters* 93:268701.
- Gross, Justin H., Joshua M. Jansa. 2017. "Relational Concepts, Measurement, And Data Collection." In *The Oxford Handbook of Political Networks*. New York: Oxford University Press.
- Lazer, David. "Networks In Political Science: Back To The Future." 2011. *PS: Political Science & Politics* 44:61-68.
- Pfeffer, Jurgen. 2017. "Visualizing Political Networks." In *The Oxford Handbook of Political Networks*. New York: Oxford University Press.

- **Exercise Nine: Fit and discuss some network models.**

November 20: Bayesian Statistics

- **Readings**

- *Required:*

- Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, Martin Modrák. 2020. "[Bayesian Workflow](#)." (Also soon to be a [book](#).)
- Gelman, Andrew, and Cosma Rohilla Shalizi. 2012. "[Philosophy and the Practice of Bayesian Statistics](#)." *British Journal of Mathematical and Statistical Psychology* 66: 8-38.

- *Suggested:*

- Gelman, Andrew, John B. Carlin, Hal S. Stern, David Dunson, Aki Vehtari, and Donald B. Rubin. 2020. [Bayesian Data Analysis, Third Edition](#). New York: Wiley.
- Gill, Jeff. 2014. [Bayesian Methods: A Social and Behavioral Sciences Approach, Third Edition](#). New York: Chapman and Hall/CRC (and [R package](#)).
- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. New York: Wiley.
- Savage, Leonard J. 1972. *The Foundations of Statistics*, Second Revised Edition. New York: Dover.
- Many, many more...

November 27: No Class: Thanksgiving Break

December 4: Text Analysis: Introduction and Overview

- **Readings**

- *Required:*

- Denny, Matthew J., and Arthur Spirling. 2018. “Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It.” *Political Analysis* 26:168-189.
- Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21:267-297.
- Miner, Gary, John Elder IV, Thomas Hill, Robert Nisbet, Dursun Delen, and Andrew Fast. 2012. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, 1st Ed. Cambridge, MA: Academic Press. Chapters 2 and 3. Available at <https://nlp.stanford.edu/IR-book/>.
- O’Connor, Brendan, David Bamman, and Noah A. Smith. 2011. “Computational Text Analysis for Social Science: Model Assumptions and Complexity.” NIPS Conference.
- *Suggested* :
 - Monroe, Burt and Phillip Schrodt. 2008. “Introduction to the Special Issue: The Statistical Analysis of Political Text.” *Political Analysis* 16:351-355.
 - Pathak, Manas A. 2014. *Beginning Data Science with R*. New York: Springer. Chapter 8.
 - Roberts, Margaret E. 2017. “Introduction to the Virtual Issue: Recent Innovations in Text Analysis for Social Science.” *Political Analysis* Virtual Issue.

December 11: Text Analysis: Sentiment, Topics, Scaling, LLMs...

• Readings

- *Required (choose 3-4)*:
 - Benoit, Kenneth, Kevin Munger, and Arthur Spirling. 2019. “Measuring and Explaining Political Sophistication Through Textual Complexity.” *American Journal of Political Science* 63:491-508.
 - Blei, David. 2012. “Probabilistic Topic Models.” *Communications of the ACM* 55:77-84.
 - Laver, Michael, Kenneth Benoit, and John Garry. 2003. “Extracting Policy Positions from Political Texts Using Words as Data.” *American Political Science Review* 97:311-331.
 - Lowe, Will. 2008. “Understanding Wordscores.” *Political Analysis* 16:356-371.
 - Pang, Bo, and Lillian Lee. 2008. “Opinion Mining and Sentiment Analysis.” *Foundations and Trends in Information Retrieval* 2:1-135. (read quickly)
 - Slapin, Jonathan and Sven-Oliver Proksch. 2008. “A Scaling Model for Estimating Time-Series Party Positions from Texts.” *American Journal of Political Science* 52:705-722.
- *Suggested*:
 - Blei, David, Andrew Ng, and Michael Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning* 3:993-1022.
 - Dodds, Peter and Christopher Danforth. 2009. “Measuring the Happiness of Large- Scale Written Expression: Songs, Blogs, and Presidents.” *Journal of Happiness Studies* 11:441-456.
 - Grimmer, Justin. 2010. “A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases.” *Political Analysis* 18:1-35.

- Kim, In Song, John Londregan, and Marc Ratkovic. 2018. “Estimating Spatial Preferences from Votes and Text.” *Political Analysis* 26:210-229.
- Lauderdale, Benjamin, and Alexander Herzog, “Measuring Political Positions from Legislative Speech.” *Political Analysis* 24:374-394.
- Lowe, Will. 2016 (etc.) “Scaling Things We Can Count.” Manuscript: Princeton University. <http://dl.conjugateprior.org/preprints/all-on-the-line.pdf>
- Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev. 2010. “How to Analyze Political Attention with Minimal Assumptions and Costs.” *American Journal of Political Science* 54:209-228.
- Rice, Douglas R., and Christopher Zorn. 2018. “Corpus-Based Dictionaries for Sentiment Analysis of Specialized Vocabularies.” *Political Science Research and Methods* 6:online.
- Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2018 “stm: R Package for Structural Topic Models.” *Journal of Statistical Software*, forthcoming.
- Soroka, Stuart, Lori Young, and Meital Balmas. 2015. “Bad News or Mad News? Sentiment Scoring of Negativity, Fear, and Anger in News Content.” *The Annals of the American Academy of Political and Social Science* 659:108-121.

- **Exercise Ten: Analyze some text.**

December 18: Final papers/projects are due.