

Department of Political Science
University of Toronto

POL 2519 H1F: Quantitative Methods and Data Analysis Winter 2025

Course Information

- **Class Time:** Friday, 10AM–12PM
- **Class Location:** Sidney Smith Hall, Room SS561
- **Instructor:** Nicola Nones
Email: nicola.nones@utoronto.ca
- **Office Hours:** By appointment.
Office Location: TBA / Zoom.

Course Description

This course provides graduate students with advanced training in quantitative methods focusing on two families of models commonly used in political science:

1. Models for categorical dependent variables, and
2. Models for panel data analysis.

The course builds on materials from POL 2504 and POL 2507, designed for PhD students advancing their methodological skills and use statistical methods in their own research. It may also suit MA students with a strong quantitative background and interested in an advanced course going beyond linear regression. The course comprises lectures presenting the theory behind each statistical model, discussions of concrete examples based on published articles, as well as interactive sessions using R.

Course Format

The course is conducted in person at Sidney Smith Hall, where students can use lab computers or their own laptops (the latter option is strongly suggested). Classes include:

- Advanced lectures introducing statistical concepts (first half)
- Interactive R exercises using real-world datasets (second half)

While the course requires prior knowledge in statistics (see requirements below), the pedagogical approach is tailored to students who may not have had an extended training in mathematics as undergraduate students, as is often the case in the social sciences.

Requirements

Normally, PhD students will register for POL 2519 after having taken the quantitative methods sequence POL 2504 and POL 2507. However, MA and PhD students who already have an equivalent background may also register for the course. To maximize the benefits of taking this course, students should have a good understanding of basic statistics and the linear regression model.

Software

The course uses R and RStudio, which are open-source and free. These tools are available in the computer lab or for download:

- <https://www.r-project.org/>
- <https://www.rstudio.com/products/rstudio/download/>

Marking Scheme

- **Written Assignment #1:** 30%, Due: tentatively, early March.
- **Written Assignment #2:** 30%, Due: tentatively, early April.
- **Term Paper:** 30%, Due: tentatively, mid-April.
- **Participation:** 10%

Readings

Students will be provided with lecture notes covering the course materials. One or more required reading that supplement these lecture notes is associated with some week of the class. Reference textbooks include:

- J. Scott Long. *Regression Models for Categorical and Limited Dependent Variables*. Sage Publications, 1997.
- D.W. Hosmer, S. Lemeshow, R.X. Sturdivant. *Applied Logistic Regression*. Wiley, 2013.
- John Fox. *Applied Regression Analysis and Generalized Linear Models*. 3rd Ed., Sage, 2016.

Other Useful References

- Edward W. Frees. *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge University Press, 2004.
- P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall.
- Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt and Jon C. W. Pevehouse. *Time Series Analysis for the Social Sciences*. Cambridge University Press, 2014.
- Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, Chapters 10, 12–20.
- William H. Greene. *Econometric Analysis*. 7th Edition, Pearson Education, Chapters 11–12, 14, 17–20.
- Peter Kennedy. *A Guide to Econometrics*. 6th Edition. Wiley-Blackwell.
- John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. 2nd Edition. Sage.

Evaluations

Written Assignments

The two written assignments are set of problems designed to assess students' ability to put the methods learned into practice. They may involve running models using a statistical package and answering short factual questions about these models and the results. There is no better way to improve one's skills than by practicing. Therefore, these assignments are not only useful as evaluations, they serve as a valuable exercise helping students to gain hands-on expertise with the subject matter. These assignments are done individually and sent to Quercus on the due date.

The first assignment will revolve around categorical dependent variables in the cross-sectional case. The second assignment will be on linear panel data.

Term Paper

An empirical section of a research paper (between 3,000 and 5,000 words maximum), involving any course-taught models. Students may collaborate and must provide replication materials.

Students may opt to work on a dissertation chapter or use this opportunity to write a stand-alone paper intended for publication. Students can decide to work as a team for the term paper.

The term paper must begin with a concise introduction that outlines the research question, provides a theoretical framework, and includes testable hypotheses. Although this section is not formally evaluated, the clarity and logical consistency of the theory and hypotheses are essential, as they will directly influence the empirical analysis.

The core of the term paper should present the research design and undertake a complete empirical analysis using one or more models covered in the course. Students are required to submit replication materials to accompany their work.

Please notice that the term paper is due at around the same time as the second assignment. Don't worry! As I will explain in class, you will have short homework to be done almost every week. These homework will not be graded nor are mandatory. Nevertheless, it is hardly suggested that you do the homework on your own dataset. By doing so, you should have the term partly finished by the end of class and all you will need to do is to write the final report.

Class Schedule: Summary

Date	Topic	Assignments
January 10	Introduction	
January 17	Maximum Likelihood Estimation	
January 24	Models for Binary Dependent Variables I	
January 31	Models for Binary Dependent Variables II	
February 7	Models for Ordered Dependent Variables	
February 14	Models for Multinomial Dependent Variables	
February 21	Reading Week	
February 28	Models for Count Dependent Variables	Assignment 1 Due
March 7	Concepts in Time-Series and Panel Data	
March 14	Panel I: Random, Between, and Fixed Effects	
March 21	Panel II: Two-Way Fixed Effects, the Problem	
March 28	Panel III: Two-Way Fixed Effects, the Solution(s)	
April 4	Panel IV: Dynamic models? Synthetic Control?	Assignment 2 Due
April 11 (?)	No Class/End of Semester	Term Paper Due

Note: Topics and dates mentioned on this syllabus may be adjusted slightly due to unforeseen circumstances, students' interests in specific models, or the total number of registered students. If there is something that interests you, please let me know!

Class Schedule: Detailed

January 10: Introduction

- Presentations.
- Expectations.
- Syllabus reading.
- Inputs from you?

January 17: Maximum Likelihood Estimation

- Refresher on notation and least squares estimation.
- Principles of optimization.
- Maximum likelihood estimation (MLE).
- Numerical implementation of MLE.

Reading: Long, Chapter 2.

Recommended Reading: Long, Chapter 1; Fox, Appendix A-B-C-D (except D.7).

January 24: Models for Binary Dependent Variables I

- Logit and probit models.
- Interpretation.
- Maximum likelihood estimation and inference in R.

Reading: Long, Chapter 3; Fox, Chapter 14.1

Recommended Reading: Hosmer et al., Chapters 2 and 3.

January 31: Models for Binary Dependent Variables II

- Classification and goodness-of-fit statistics.
- Predicted probabilities and marginal effects.
- Separation and other limitations of logit and probit models.
- Other models for binary dependent variables.

Reading: Long, Chapter 4.

Recommended Reading: Hosmer et al., Chapters 5.

February 7: Models for Ordered Dependent Variables

- Ordered logit and probit models.
- Estimation and interpretation.
- Post-estimation techniques.

Reading: Long, Chapter 5; Fox, Chapter 14.2

Recommended Reading: Hosmer et al., Chapters 8.

February 14: Models for Multinomial Dependent Variables I

- Multinomial logit model.
- Alternative-specific variables (conditional logit).
- Interpretation and post-estimation analysis.
- Independence of irrelevant alternatives (IIA) assumption.

Reading: Long, Chapter 6.

Recommended Reading: Hosmer et al., Chapters 8.

February 21: Reading Week

February 28: Models for Count Dependent Variables

- Poisson and negative binomial models.
- Overdispersion.
- Estimation and interpretation.
- The Generalized Linear Model.

Reading: Long, Chapter 8; Fox, Chapter 15.2

March 7: Introduction to Panel Data

- Time-series vs. cross-sectional regressions.
- Autocorrelation and spurious regression.
- Dynamic regression.
- The problem of panel heterogeneity.

Reading: Fox, Chapter 16; Stimson, James. 1985. “Regression in Space and Time: A Statistical Essay.”

Recommended Reading: Lebo, Matthew J., and Christopher Weber. 2015. “An Effective Approach to the Repeated Cross- Sectional Design.”

March 14: Random, Between and Fixed Effects

- Introduction to panel data analysis.
- White/HAC and "panel-corrected" standard errors.
- Fixed effects estimator.
- Between effects estimator.
- Random effects estimator.

Reading: Beck, Nathaniel, and Jonathan N. Katz. 1995. "What To Do (And Not to Do) With Time-Series Cross-Section Data."; Green, Donald P., Soo Yeon Kim, and David Yoon. 2001. "Dirty Pool."; Clark, Tom S., and Drew A. Linzer. 2014. "Should I Use Fixed or Random Effects?"

Recommended Reading: Zorn, Christopher. 2001. "Estimating Between- and Within-Cluster Covariate Effects, with an Application to Models of International Disputes."

March 21: Two-Way Fixed Effects, the Problem(s)

- TWFE estimator.
- Problems of interpretation.
- Bias and negative weights.

Reading: Kropko, J., and Kubinec, R. 2020. "Interpretation and identification of within-unit and cross-sectional variation in panel data models"; Imai, Kosuke, and In Song Kim. 2021, "On the use of two-way fixed effects regression models for causal inference with panel data."

Recommended Reading: Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing."

March 28: Two-Way Fixed Effects, the Solution(s)

- More recent estimators in the economics literature.
- More recent estimators in the political science literature.

Reading: Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature."; Liu, Licheng, Ye Wang, and Yiqing Xu. 2022. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data.";

Recommended Reading: Imai, Kosuke, In Song Kim, and Erik H. Wang. 2023 "Matching methods for causal inference with time-series cross-sectional data."

April 4: Additional Topics. NO Synchronous Class (MPSA) - I may record it or we reschedule.

- Arellano-Bond (and similar) estimators.
- Models for panel data with discrete dependent variables.
- Synthetic Control Method.

Reading: Wawro, Gregory. 2002. “Estimating dynamic panel data models in political science.”; Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. “Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable.”; Xu, Yiqing. 2017. “Generalized synthetic control method: Causal inference with interactive fixed effects models.”

April 11: End of class (no class?)

References and Examples

This section contains a list of applications in the literature and additional references on the methods.

Models for Binary Dependent Variables

Carrubba, Cliff, Barry Friedman, Andrew D. Martin and Georg Vanberg. 2012. “Who Controls the Content of Supreme Court Opinions?” *American Journal of Political Science*, 56(2): 400–412.

Dion, Michelle L., Jane Lawrence Sumner and Sara McLaughlin Mitchell. 2016. “Gendered Citation Patterns across Political Science and Social Science Methodology Fields.” *Political Analysis*, 26(3): 312–327.

O’Brien, Diana Z. and Rickne, Johanna. 2016. “Gender Quotas and Women’s Political Leadership.” *American Political Science Review*, 110(1): 112–126.

Rubenzer, Trevor. 2011. “Campaign Contributions and U.S. Foreign Policy Outcomes: An Analysis of Cuban American Interests.” *American Journal of Political Science*, 55(1): 105–116.

Zorn, Christopher. 2002. “U.S. Government Litigation Strategies in the Federal Appellate Courts.” *Political Research Quarterly*, 55(1): 145–166.

Zorn, Christopher. 2005. “A Solution to Separation in Binary Response Models.” *Political Analysis*, 13(2): 157–170.

Goodness-of-Fit and Predicted Probabilities

Hagle, Timothy M. and Glenn E. Mitchell II. 1992. "Goodness-of-Fit Measures for Probit and Logit." *American Journal of Political Science*, 36(3): 762–784.

Hanmer, Michael J. and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science*, 57(1): 263–277.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Prediction." *American Journal of Political Science*, 44: 347–361.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis*, 8(1): 83–98.

Heteroskedastic Probit

Alvarez, R. Michael, and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science*, 39: 1055–1082.

Rare Events Logit

King, Gary, and Langsche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis*, 9(2): 137–163.

Interaction Effects

Berry, William D., Jacqueline H.R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction Effects in Binary Logit and Probit Models: Is the Product Term Essential?" *American Journal of Political Science*, 54(1): 248–266.

Berry, William D., Matt Golder, and Daniel Milton. 2012. "Improving Tests of Theories Positing Interaction." *Journal of Politics*, 74(August): 653–671.

Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis*, 14: 63–82.

Models for Ordered Dependent Variables

Alvarez, R. Michael, and John Brehm. 1998. "Speaking in Two Voices: American Equivocation about the Internal Revenue Service." *American Journal of Political Science*, 42(2): 418–52.

Franklin, Charles H. and Liane C. Kosaki. 1989. "Republican Schoolmaster: The Supreme Court, Public Opinion and Abortion." *American Political Science Review*, 83(3): 751–771.

Gelpi, Christopher. 1997. "Crime and Punishment: The Role of Norms in Crisis Bargaining." *American Political Science Review*, 91(2): 339–360.

Sanders, Mitchell S. 2001. "Uncertainty and Turnout." *Political Analysis*, 9(1): 45–57.

Multi-Class Goodness-of-Fit

Hand, David J. and Robert J. Till. 2001. "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems." *Machine Learning*, 54(2): 171–186.

Sokolova, Marina and Guy Lapalme. 2009. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing and Management*, 45: 427–437.

Models for Nominal (Unordered) Dependent Variables

Brownstone, David and Kenneth Train. 1999. "Forecasting New Product Penetration with Flexible Substitution Patterns." *Journal of Econometrics*, 89: 109–129.

Gidengil, Elisabeth, Neil Nevitte, André Blais, Joanna Everitt and Patrick Fournier. 2012. *Dominance and Decline: Making Sense of Recent Canadian Elections*. Toronto: University of Toronto Press.

Glasgow, Garrett. 2001. "Mixed Logit Models for Multiparty Elections." *Political Analysis*, 9(2): 116–136.

Iyengar, Shanto and Kyu S. Hahn. 2009. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication*, 59: 19–39.

Maltzman, Forrest, and Paul J. Wahlbeck. 1996. "May it Please the Chief? Opinion Assignments in the Rehnquist Court." *American Journal of Political Science*, 40(2): 421–43.

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): 1231–1247.

Rudolph, Thomas J. 2003. "Who's Responsible for the Economy? The Formation and Consequences of Responsibility Attributions." *American Journal of Political Science*, 47(4): 698–713.

Swait, Joffre and Jordan Louviere. 1993. "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Marketing Research*, 30(3): 305–314.

IIA Assumption

Alvarez, R. Michael and Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science*, 42(1): 55–96.

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): 107–122.

Models for Count Dependent Variables

King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science*, 32(3): 838–863.

King, Gary. 1989. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science*, 33(3): 762–784.

King, Gary. 1989. "Event Count Models for International Relations: Generalizations and Applications." *International Studies Quarterly*, 33: 123–47.

Gowa, Joanne. 1998. "Politics at the Water's Edge: Parties, Voters and the Use of Force Abroad." *International Organization*, 52(2): 307–24.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2012. "Institutional Foundations of Legislative Speech." *American Journal of Political Science*, 56(3): 520–537.

Models for Time Series Analysis

Methodological Literature

Beck, Nathaniel. 1993. "The Methodology of Cointegration." *Political Analysis*, 4(1): 237–248.

Box-Steffensmeier, Janet M., John R. Freeman, Matthew P. Hitt, and Jon C.W. Pevehouse. 2014. *Time Series Analysis for the Social Science*. Cambridge: Cambridge University Press.

Cowpertwait, Paul S. P. and Andrew V. Metcalfe. 2009. *Introductory Time Series with R*. Berlin: Springer-Verlag.

De Boef, Suzanna and Luke Keele. 2008. "Taking Time Seriously." *American Journal of Political Science*, 52(1): 184–200.

Engle, Robert F. and Clive W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, 55(2): 251–276.

Granger, Clive W. J. and Paul Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics*, 2: 111–120.

Keele, Luke and Nathan Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis*, 14: 186–205.

Pesaran, M. Hashem and Yongcheol Shin. 1999. "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis." In Steinar Strom, Ed., *Econometrics and Economic Theory in the 20th Century*. Cambridge: Cambridge University Press, pp. 371–413.

Applications in Political Science

Box-Steffensmeier, Janet M, Suzanna de Boef, and Tse-Min Lin. 2004. "The Dynamics of the Partisan Gender Gap." *American Political Science Review*, 98(3): 515–528.

Brandt, Patrick T. and John R. Freeman. 2009. "Modeling Macro-Political Dynamics." *Political Analysis*, 17(2): 113–142.

Carter, David B. and Curtis S. Signorino. 2010. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis*, 18(3): 271–292.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1989. "Macropartisanship." *American Political Science Review*, 83(4): 1125–1142.

Models for Panel Data Analysis

Methodological Literature

Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58: 277–297.

Arellano, Manuel. 1987. "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics*, 49(4): 431–434.

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(3): 634–647.

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(4): 1260–1288.

Beck, Nathaniel and Jonathan N. Katz. 2007. "Random Coefficient Models for Time-Series-Cross-Section Data." *Political Analysis*, 15(2): 182–195.

Beck, Nathaniel and Jonathan M. Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science*, 14: 331–352.

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics*, 87: 115–143.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge: Cambridge University Press.

Hausman, Jerry and William Taylor. 1981. "Panel Data and Unobservable Individual Effects." *Econometrica*, 49: 1377–1398.

Hood III, M.V., Quentin Kidd, and Irwin L. Morris. 2008. "Two Sides of the Same Coin? Employing Granger Causality Tests in a Time-Series-Cross-Section Framework." *Political Analysis*, 16(3): 324–344.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46(1): 69–85.

White, Halbert. 1980. *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

White, Halbert. 1984. "A Heteroskedasticity–Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica*, 48: 817–838.

Applications in Political Science

Ban, Pamela, Alexander Fourinaies, Andrew B. Hall, and James M. Snyder. 2018. "How Newspapers Reveal Political Power." *Political Science Research and Methods*, 7(4): 661–678.

Boix, Charles. 2011. "Democracy, Development, and the International System." *American Political Science Review*, 105(4): 809–828.

Pickering, Jeffrey and Emizet F. Kisangani. 2010. "Diversionary Despots? Comparing Autocracies' Propensities to Use and to Benefit from Military Force." *American Journal of Political Science*, 54(2): 477–493.

Stimson, James. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science*, 29: 914–947.

Zahariadis, Nikolaos. 1997. "Why State Subsidies? Evidence from European Community Countries 1981–1986." *International Studies Quarterly*, 41(2): 341–354.

Zorn, Christopher J.W. 2001. "Estimating Between-and Within-Cluster Covariate Effects, with an Application to Models of International Disputes." *International Interactions*, 27(4): 433–445.