Department of Political Science
University of Toronto

POL 2519 H1F
Quantitative Methods and Data Analysis

Course Outline

Fall 2017
(Section L0101)

Class Time: **Tuesdays, 10AM–12PM**
Class Location: **RW 109 (Ramsay Wright Computer Lab 109)**

Instructor: Ludovic Rheault
Email: ludovic.rheault@utoronto.ca
Office Hours: Tuesdays, 1–3PM.
Office Location: Sidney Smith 3005

Course Description

This course provides graduate students with an advanced training in quantitative methods focusing on the two families of models most commonly used in political science: 1) models for categorical dependent variables, and 2) models for panel data analysis. The course starts where POL 2504 ends, and represents a natural continuation for PhD students looking to develop their methodological skills and use statistical methods in their own research. POL 2519 is also suitable for MA students who received training in quantitative methods at the undergraduate level and interested in an advanced course going beyond linear regression. The course takes place in a computer lab. It comprises lectures presenting the theory behind each statistical model, discussions of concrete examples based on published articles, as well as interactive sessions with statistical packages, in particular the R programming language.

Course Format

The course takes place in the Ramsay Wright computer lab RW 109 (the building next to Sidney Smith). Students can use the computers available in the lab or bring their own laptop.

A typical class will combine an advanced lecture on statistical theory introducing new concepts, followed with interactive exercises using the R language and real-world datasets.

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). However, the course should help students acquire both an understanding of the models and practical experience with data analysis.

Requirements

Normally, PhD students will register for POL 2519 after having taken POL 2504. 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 at least a good understanding

of basic statistics and the linear regression model. In doubt, students may show up and discuss these requirements with the instructor.

Software

In line with POL 2504, the course will mainly rely upon the R programming language for teaching and illustrations. R is an open-source language available on all operating systems (that is, it is free to use). Students are invited to download RStudio, a free text editor to use the R language, which I will use for in-class examples. Some examples of equivalent commands using the proprietary software package Stata will also be provided during class.

Marking Scheme

Written Assignment #1	25%	Due: October 10, 2017
Written Assignment #2	25%	Due: October 31, 2017
Oral Presentation	10%	Last two weeks.
Term Paper	30%	Due: December 8, 2017
Participation	10%	

Readings

Students will be provided with the lecture notes used during the course. It is recommended to supplement these lecture notes with texts from other authors addressing the same material. The recommended readings for this course are:

- **Models for Categorical Dependent Variables**
  1. J. Scott Long. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks: Sage Publications.
     - This is an older reference, but a classic that covers all of the core models we are studying during the first module of the course.
  2. David W. Hosmer, Stanley Lemeshow, Rodney X. Sturdivant. 2013. *Applied Logistic Regression.* Hoboken: John Wiley & Sons.
     - Another classic with a recent edition that can be used as a substitute to Long. This book is available electronically in its entirety from the UofT Library. It covers most of the first module of the course in depth, with a few exceptions.
- **Models for Panel Data**
  1. Badi H. Baltagi. 2008. *Econometric Analysis of Panel Data.* 4th Edition. Hoboken: John Wiley & Sons.
     - For the second module of the course. Available at the Robarts Library.

- **Other Useful References**
  1. Kosuke Imai. 2017. *Quantitative Social Science: An Introduction.* Princeton: Princeton University Press.
     - A good refresher covering many concepts and models addressed in undergraduate courses or introductory graduate courses.
  2. Jeffrey M. Wooldridge. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press, Chapters 10, 12–20.
     - A general textbook covering most of the material studied in this course.
  3. William H. Greene. 2011. *Econometric Analysis.* 7th Edition. Upper Saddle River: Pearson Education, Chapters 11–12, 14, 17–20.
     - A general textbook covering the theoretical aspects of most of the models studied in this course.
  4. Peter Kennedy. 2008. *A Guide to Econometrics.* 6th Edition. Wiley-Blackwell.
     - For those who prefer English explanations to mathematics, Kennedy's book could be a useful acquisition. There is a good chapter on panel data and some chapters on regression models for limited dependent variables.
  5. John Fox and Sanford Weisberg. 2011. *An R Companion to Applied Regression.* 2nd Edition. Thousand Oaks: Sage.
     - A useful handbook to implement the various models using R, for those who may need it.

Evaluations

Written Assignments

The two short written assignments are problem sets designed to evaluate 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 practice. 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 handed in during class at the due date.

Oral Presentation

The oral presentation consists of presenting the research design for the term paper (and optionally some preliminary results, if available). Each presentation should take about 10 minutes, followed by feedback from the audience. Students' ability to invoke the concepts studied during class will be evaluated. This is also an opportunity to get useful feedback for the term paper from the other participants.

Term Paper

The term paper takes the form of the empirical section of a research paper on a topic of the graduate student's choosing, and involving any of the models discussed during the course.

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 in teams for the term paper and oral presentation.

The term paper will include a brief introduction stating the research question, an outline of the theory and some testable propositions (hypotheses). This section is not part of the evaluation per se, but the theory and hypotheses should be clearly stated and logically consistent, as this will inevitably affect the empirical analysis.

The main part of the term paper (roughly 4,000 to 6,000 words) consists of an empirical section introducing the empirical research design and performing all the stages of a state-of-the-art empirical analysis, again using any of the tools and methods acquired during the course. Students should make sure to provide the replication material for their study.

Class Schedule: Summary

Date	Topic	Evaluation
September 12	Maximum Likelihood Estimation	
September 19	Models for Binary Dependent Variables I	
September 26	Models for Binary Dependent Variables II	
October 3	Models for Ordered Dependent Variables	
October 10	Models for Nominal Dependent Variables I	Assignment 1 Due
October 17	Models for Nominal Dependent Variables II	
October 24	Models for Count Dependent Variables	
October 31	Generalized Linear Models and Bayesian Inference	Assignment 2 Due
November 7	**Reading Week**	
November 14	Concepts in Time-Series and Panel Data	
November 21	Panel Data: Random, Between and Fixed Effects	
November 28	Dynamic Panel Data	Presentations
December 5	Advanced Topics in Panel Data	Presentations
December 8	**No Class**	Term Paper Due

Note: Topics and dates may be adjusted due to unforeseen circumstances or student interest in specific themes.

Class Schedule: Detailed

1 Models for Categorical Dependent Variables

1.1 September 12: Maximum Likelihood Estimation

Structure of the Class:

1. Refresher on notation and least squares estimation.
2. Principles of optimization.
3. Maximum likelihood estimation (MLE).
4. Numerical implementation of MLE.

Recommended Readings:

- Long, Chapter 2; Hosmer et al., Chapter 1.

1.2 September 19: Models for Binary Dependent Variables I

Structure of the Class:

1. Logit and probit models.
2. Interpretation.
3. Maximum likelihood estimation and inference in R.

Recommended Readings:

- Long, Chapter 3; Hosmer et al., Chapters 2–3.

1.3 September 26: Models for Binary Dependent Variables II

Structure of the Class:

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

Recommended Readings:

- Long, Chapter 4; Hosmer et al., Chapter 5.

Written Assignment #1 Handed Out.

1.4 October 3: Models for Ordered Dependent Variables

Structure of the Class:

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

Recommended Readings:

- Long, Chapter 5; Hosmer et al., Chapter 8.

1.5 October 10: Models for Nominal Dependent Variables I

Structure of the Class:

1. Multinomial logit model (softmax).
2. Alternative-specific variables (conditional logit).
3. Interpretation and post-estimation analysis.

Recommended Readings:

- Long, Chapter 6; Hosmer et al., Chapter 8.

Written Assignment #1 Due.

1.6 October 17: Models for Nominal Dependent Variables II

Structure of the Class:

1. Independence of irrelevant alternatives (IIA) assumption.
2. Multinomial probit model and mixed logit model.
3. Other models for categorical dependent variables.

Recommended Readings:

- Long, Chapter 6; Hosmer et al., Chapter 8.

Written Assignment #2 Handed Out.

1.7 October 24: Models for Count Dependent Variables

Structure of the Class:

1. Poisson and negative binomial models.
2. Overdispersion.
3. Estimation and interpretation.

Recommended Readings:

- Long, Chapter 8.

1.8 October 31: Generalized Linear Models and Bayesian Inference

Structure of the Class:

1. A common framework for categorical dependent variables: GLMs.
2. Bayesian inference and logistic regression.
3. Sampling and interpretation of Bayesian models in R.

Recommended Readings:

- Long, Chapter 9.2; Hosmer et al., Chapter 10.6.

Written Assignment #2 Due.

November 7: Reading Week

No class.

2 Panel Data Analysis

2.1 November 14: Concepts in Time-Series and Panel Data

Structure of the Class:

1. Time-series v. cross-sectional regressions.
2. Autocorrelation and spurious regression.
3. ARMA model and dynamic regression.
4. Granger causality tests.
5. Panel heterogeneity.

Recommended Readings:

- Baltagi, Chapter 1.

2.2 November 21: Random, Between and Fixed Effects

Structure of the Class:

1. Introduction to panel data analysis.
2. Random effects estimator.
3. Between effects estimator.
4. Fixed effects estimator.
5. Estimation in R.

Recommended Readings:

- Baltagi, Chapters 2–3.

2.3 November 28: Dynamic Panel Data Estimators

Structure of the Class:

1. Autoregressive distributed lag (ADL) model.
2. Arellano-Bond estimator.
3. (Student presentations).

Recommended Readings:

- Baltagi, Chapter 8.

2.4 December 5: Advanced Topics in Panel Data

Structure of the Class:

1. White/HAC and “panel-corrected" standard errors.
2. Panel unit roots and cointegration.
3. Models for panel data with discrete dependent variables.
4. (Student presentations).

Recommended Readings:

- Baltagi, Chapter 11.

References

This section contains a list of example applications from the literature, and additional references on the methods. We will also look at datasets from published papers during the course.

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.

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): 145–66.

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): 8398.

Hetroskedastic 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):339–60.

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): 171–86.

Sokolova, Marina and Guy Lapalme. 2009. "A Systematic Analysis of Performance Measures for Classification Tasks." *Information Processing and Management* 45: 427–37.

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):116–36.

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 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 Stnadard 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 Times-Series-Cross-Section Data." *American Political Science Review* 89(3): 634-647.

Beck, Nathaniel and 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, Bond Stephen. 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 Multilevel/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. *Asymtotic 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

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.