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 builds upon the materials covered in POL 2504, 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 a strong training in quantitative methods at the undergraduate level and are 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 takes place in the Ramsey Wright computer lab RW 109 (the building next to Sidney Smith). Students who cannot travel to Canada or are unable to attend classes in person due to the pandemic will be able to join synchronously on Zoom. Students watching remotely will be able to see the shared screen for both presentations and practical exercises on R. If a second wave of the virus requires us to cancel in-person activities during the fall, we will transfer all activities on Zoom and continue from there as seamlessly as possible.

Students who attend in person can use the computers available in the lab or bring their own laptop. Students watching remotely will be able to perform exercises on their own computers.

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

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 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 a good understanding of basic statistics and the linear regression model.

Software

In line with POL 2504 and other courses offered in the department, POL 2519 relies on 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.

Marking Scheme

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Percentage</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written Assignment #1</td>
<td>25%</td>
<td>October 13, 2020</td>
</tr>
<tr>
<td>Written Assignment #2</td>
<td>25%</td>
<td>November 3, 2020</td>
</tr>
<tr>
<td>Oral Presentation</td>
<td>10%</td>
<td>Last two weeks.</td>
</tr>
<tr>
<td>Term Paper</td>
<td>30%</td>
<td>December 10, 2020</td>
</tr>
<tr>
<td>Participation</td>
<td>10%</td>
<td></td>
</tr>
</tbody>
</table>

Readings

Students will be provided with lecture notes covering the course materials. At least one reading that supplements these lecture notes is associated to each week of the class. Given the restrictions that students may be facing due to the pandemic, all readings will be available in electronic format on Quercus, to ensure everyone has access to them.

• Reference Textbooks
  The readings are chapters taken from the following textbooks:

• Other useful References
  Other textbooks that can be useful for students looking to explore the topics further.
    * A classic textbook covering the models under study during the first module of this course.
* A useful text for students interested in time series analysis, at a good level for political scientists.
  - A general textbook covering most of the materials studied in this course.
  - A comprehensive textbook in statistics covering the theoretical aspects of the models studied in this course, and more.
  - 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.
  - A useful handbook with many examples of fitting models using R.

**Evaluations**

**Written Assignments**

The two 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. The students’ ability to invoke 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 sections introducing the research design and performing all the stages of an empirical analysis, using one (or more) of the models studied in the course. Students should make sure to provide replication materials for their study.

Class Schedule: Summary

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 15</td>
<td>Maximum Likelihood Estimation</td>
<td></td>
</tr>
<tr>
<td>September 22</td>
<td>Models for Binary Dependent Variables I</td>
<td></td>
</tr>
<tr>
<td>September 29</td>
<td>Models for Binary Dependent Variables II</td>
<td></td>
</tr>
<tr>
<td>October 6</td>
<td>Models for Ordered Dependent Variables</td>
<td></td>
</tr>
<tr>
<td>October 13</td>
<td>Models for Multinomial Dependent Variables I</td>
<td>Assignment 1 Due</td>
</tr>
<tr>
<td>October 20</td>
<td>Models for Multinomial Dependent Variables II</td>
<td>Assignment 2 Due</td>
</tr>
<tr>
<td>October 27</td>
<td>Models for Count Dependent Variables</td>
<td></td>
</tr>
<tr>
<td>November 3</td>
<td>Review of Generalized Linear Models, Bayesian Inference</td>
<td>Assignment 2 Due</td>
</tr>
<tr>
<td>November 10</td>
<td><strong>Reading Week</strong></td>
<td></td>
</tr>
<tr>
<td>November 17</td>
<td>Concepts in Time-Series and Panel Data</td>
<td></td>
</tr>
<tr>
<td>November 24</td>
<td>Panel Data I: Random, Between and Fixed Effects</td>
<td></td>
</tr>
<tr>
<td>December 1</td>
<td>Panel Data II</td>
<td>Presentations</td>
</tr>
<tr>
<td>December 8</td>
<td>Dynamic Panel Data</td>
<td>Presentations</td>
</tr>
<tr>
<td>December 10</td>
<td><strong>No Class/End of Semester</strong></td>
<td>Term Paper Due</td>
</tr>
</tbody>
</table>

**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.
1 Models for Categorical Dependent Variables

1.1 September 15: Maximum Likelihood Estimation

Structure of the Class:
1. Refresher on notation and least squares estimation.
4. Numerical implementation of MLE.

Reading:
• Long, Chapter 2.

1.2 September 22: 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.

Reading:
• Long, Chapter 3; Hosmer et al., Chapters 2–3.

1.3 September 29: Models for Binary Dependent Variables II

Structure of the Class:
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 6: 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 13: Models for Multinomial Dependent Variables I

Structure of the Class:
1. Multinomial logit model (softmax).
3. Interpretation and post-estimation analysis.

Recommended Readings:
• Long, Chapter 6; Hosmer et al., Chapter 8.

Written Assignment #1 Due.

1.6 October 20: Models for Multinomial Dependent Variables II

Structure of the Class:
1. Independence of irrelevant alternatives (IIA) assumption.
2. Multinomial probit 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 27: Models for Count Dependent Variables

Structure of the Class:
1. Poisson and negative binomial models.
2. Overdispersion.
3. Estimation and interpretation.
1.8 November 3: Generalized Linear Models and Bayesian Inference

Structure of the Class:
The contents of that week will be adjusted based on progress.
1. A common framework for categorical dependent variables: GLMs.
2. Using Bayesian inference instead of MLE.
3. Sampling and interpretation of Bayesian models in R.

Recommended Readings:
• Fox, Chapter 15.

Written Assignment #2 Due.

November 10: Reading Week

2 Panel Data Analysis

2.1 November 17: Concepts in Time-Series and Panel Data

Structure of the Class:
1. Time-series v. cross-sectional regressions.
2. Autocorrelation and spurious regression.
3. Dynamic regression.
4. The problem of panel heterogeneity.

Recommended Readings:
• Fox, Chapter 16.

2.2 November 24: Random, Between and Fixed Effects

Structure of the Class:
1. Introduction to panel data analysis.
2. Fixed effects estimator.
3. Between effects estimator.
4. Random effects estimator.
5. Estimation in R.

Recommended Readings:
2.3 December 1: Panel Data and Multi-Level/Hierarchical Models

Structure of the Class:
1. Random effects estimator (continued).
2. Correspondence with multi-level/hierarchical models.
3. White/HAC and "panel-corrected" standard errors.
4. Student presentations.

Recommended Readings:
• Frees, Chapters 5–6.

2.4 December 8: Introduction to Dynamic Panel Data Estimators

Structure of the Class:
1. Autoregressive distributed lag (ADL) model.
2. Arellano-Bond estimator [if time permits].
3. Models for panel data with discrete dependent variables [if time permits].
4. Student presentations.

Recommended Readings:
• Frees, Chapter 8–9.
References and Examples

This section contains a list of applications in 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


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.


Goodness-of-Fit and Predicted Probabilities


Heteroskedastic Probit

**Rare Events Logit**


**Interaction Effects**


**Models for Ordered Dependent Variables**


**Multi-Class Goodness-of-Fit**


**Models for Nominal (Unordered) Dependent Variables**


**IIA Assumption**


**Models for Count Dependent Variables**


**Models for Time Series Analysis**

**Methodological Literature**


### Applications in Political Science


### Models for Panel Data Analysis

#### Methodological Literature


Applications in Political Science


