Categorical and Limited Dependent Variables

Public Affairs 56:824:708:01
Public Administration 56:834:652:01
Fall Semester 2013, Armitage 222, Thursdays 6-8:40pm

September 5, 2013

Professor

Paul A. Jargowsky, Ph.D.
856-225-2729; 321 Cooper St.; paul.jargowsky@rutgers.edu
Office Hours: TBA

Teaching Assistant

Zach Wood, Ashley Nickels (?)

Course Description

The estimation of empirical models is essential to public policy analysis and social science research. Ordinary Least Squares (OLS) regression analysis is the most frequently used empirical model, and is appropriate for analyzing continuous dependent variables that meet certain distributional assumptions. This course examines several types of advanced regression models for dependent variables that violate one or more of the assumptions of the OLS regression model. For example, some dependent variables may be categorical, such as pregnant or not, employed or not, etc. Other dependent variables may be truncated or censored, such as contributions to an individual retirement account that are limited by law to certain dollar amounts. Still others may be counts of things, like the number of children born to a given woman or the number of traffic accidents on a given day. The principal models examined in the course are binary logit and probit, multinomial logit, ordinal logit and probit, tobit, and the family of Poisson regression models. The Heckman correction for selection and Event History Analysis are also addressed. Finally, we study event history analysis, aka survival analysis or hazard analysis. All these models are estimated using maximum likelihood estimation (MLE). The course focuses primarily on the application and interpretation of the models, rather than statistical theory.

Course Pre-requisites

56:824:709 Quantitative Methods II or the equivalent is required. In other words, you should have a strong grounding in Ordinary Least Squares (OLS) regression at the level of Damodar Gujarati, Basic Econometrics, chapters 1-9, or Stock and Watson, Introduction to Econometrics, chapters 1-7. A familiarity with the basics of calculus and matrix algebra is helpful, although not required.
Student Learning Objectives/Outcomes

- Students will learn the theory and practice of regression models for limited and categorical dependent variables, including logit, probit, ordinal logit, ordinal probit, multinomial logit, Poisson regression, Tobit and related models, and event history analysis.
- Students will learn how to interpret these models by reviewing published papers drawn from social science literature.
- Students will develop proficiency in applying and interpreting these models using data provided by the instructor and/or data from their own research and employment.
- Students will demonstrate mastery of the material by writing an empirical paper using one or more of the models discussed in class and presenting their analysis and findings in class.

Required Textbooks and Materials


Recommended Course Materials

Long, S. J. and Freese, J. 2001. Regression Models for Categorical Dependent Variables Using Stata. College Station, TX: Stata Press. This book is a companion to the Long book, and contains more “nuts and bolts” on estimating models, and describes customized tools for helping to interpret the results of the models.

Course Software.

The software for the course is Stata version 13, although version 12 or any recent version will probably work just fine. Version 13 will be released in summer 2013. There is no need to upgrade from version 12 unless you want to have the latest and greatest. (However, you should make sure you loaded all the patches and bug-fixes. Type “update query” at the command line while connected to the internet to check.) You do not need to buy anything, because Stata software is available in the computing lab and online at http://apps.rutgers.edu. However, if you choose to buy Stata, you will get a substantial discount via the “Grad Plan” that has been set up for Rutgers. You want, at a minimum, the Intercooled version ($98 for a one-year license, $179 for a perpetual license). Please contact StataCorp directly: 800-782-8272 (Monday through Friday 8:00 to 5:00 Central Time) or online: http://www.stata.com/order/new/edu/gradplans/gp-direct.html.
If ordering online, be sure to use your "@rutgers.edu" or “...@camden.rutgers.edu” email address when ordering to qualify for the Rutgers-Camden discount. Typically, orders are shipped within 1 or 2 business days after the order is placed.

Requirements

1. Students are required to take a midterm and a final examination in class. The exams are open-note, open-book.

2. There are several short problem sets due at the beginning of class on the dates indicated on the schedule. The lowest problem set score will be dropped.

3. Students must complete an empirical paper on an approved topic using one or more of the techniques covered in this course. You are encouraged to think about topics and potential datasets early in the semester. A typical paper will be 15 to 20 double-spaced pages.

4. On the indicated dates, students will be asked to turn in a proposed paper topic and a first draft of the empirical paper. These items will not be graded.

Grading Policy

The grading in the course is based on the problem sets, examinations, and the empirical paper. The weights assigned to each are as follows:

- Problem Sets (lowest dropped) 40%
- Midterm Examination 30%
- Empirical Paper 30%

After computing the student semester average on a 100-point numeric scale and rounding to the nearest tenth (0.1), letter grades will be assigned as follows:

<table>
<thead>
<tr>
<th>Numeric Range</th>
<th>Letter Grade</th>
<th>Grade Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>min</td>
<td>max</td>
<td></td>
</tr>
<tr>
<td>90.0</td>
<td>100.0</td>
<td>A</td>
</tr>
<tr>
<td>85.0</td>
<td>89.9</td>
<td>B+</td>
</tr>
<tr>
<td>80.0</td>
<td>84.9</td>
<td>B</td>
</tr>
<tr>
<td>75.0</td>
<td>79.9</td>
<td>C+</td>
</tr>
<tr>
<td>70.0</td>
<td>74.9</td>
<td>C</td>
</tr>
<tr>
<td>0</td>
<td>69.9</td>
<td>F</td>
</tr>
</tbody>
</table>
| 1    | Sep. 5 | **Introduction to Course**  
|      |       | **Review of OLS and Its Limitations**  
|      |       | Long, Sections 2.1-2.5  
|      |       | * Gujarati, Chapters 2-9, especially 7-8  
|      |       | * Stock and Watson, Chapters 4-7, especially 6-7  
|      |       | **Background Information (review if/as needed):**  
|      |       | **1. Probability and Statistics**  
|      |       | * Paul Jargowsky and Rebecca Yang, “Descriptive and Inferential Statistics”  
|      |       | * Gujarati, Appendix A  
|      |       | * Stock and Watson, Chapters 2 and 3  
|      |       | **2. Introduction to Stata Programming**  
|      |       | * Long and Freese, Chapter 2  
| 2    | Sep. 12 | **Principles of Maximum Likelihood Estimation**  
|      |       | * Long, Section 2.6 (Technical)  
|      |       | **Binary Dependent Variables: Logit and Probit**  
|      |       | Long, Sections 3.1-3.4, 3.7-3.9  
|      |       | Explore “Logit Function of XiB” and “Comparison of Probit and Logit”  
| 3    | Sep. 19 | **Interpretation and Hypothesis Testing in Logit and Probit Models**  
|      |       | Long, Sections 4.1, 4.3  
|      |       | * Long, Section 4.2  
|      |       | * Long and Freese, Chapter 4  
|      |       |  
|      |       | **Problem Set 1 Due** |
## Schedule, Readings and Assignments

For the latest schedule, readings, and assignments, [consult the most recent schedule by going to the course website.](#)

* = Optional

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Readings</th>
</tr>
</thead>
</table>
| 4    | Sep. 26 | Ordinal Dependent Variables: Ordinal Logit and Ordinal Probit | Long, Sections 5.1-5.4  
* Long and Freese, Chapter 5  
Hughes and Waite (2002). “Health in Household Context: Living Arrangements and Health in Late Middle Age,” *Journal of Health and Social Behavior* 43:1-21. (Focus on Table 2 results.)  
|------|------|-------|----------|
| 5    | Oct. 3 | Nominal Dependent Variables: Multinomial Logit | Long, Sections 6.1-6.2, 6.4-6.6  
* Long and Freese, Sections 6.1-6.7  
|------|------|-------|----------|
| 6    | Oct. 10 | Nominal Dependent Variables: Conditional Logit | Long, 6.7-6.10  
* Long and Freese, Sections 7.1, 7.2  
|------|------|-------|----------|
| 7    | Oct. 17 | Nominal Dependent Variables: Multinomial Probit | * Long and Freese, Chapter 7.3  

*Problem Set 2 Due* |  
*Problem Set 3 Due* |  
*Problem Set 4 Due*
<table>
<thead>
<tr>
<th>Date</th>
<th>Reading/Assignment</th>
</tr>
</thead>
</table>
| 8 Oct. 24 | **Censored and Truncated Dependent Variables: Tobit**  
Long, Chapter 7  
| 9 Oct. 31 | **Censored and Truncated Dependent Variables: Extensions**  
Jargowsky, “Using Stata’s ML Utility.”  
| 10 Nov. 7 | **Midterm Exam**  
Open book, open note  
*Bring a calculator* |
| 11 Nov. 14 | **Count Dependent Variables: Poisson Regression and Related Models**  
Long, Chapter 8  
Hughes & Waite (2002). “Health in Household Context: Living Arrangements and Health in Late Middle Age,” *Journal of Health and Social Behavior* 43: 1-21 (Yes, it’s the same article as before. This time look at the Poisson results.)  
| 12 Nov. 21 | **Event History Analysis: Theory**  
Cleves, Chapters 1-4  
* Allison, Chapters 1-3 |

*First Draft of Empirical Paper Due*
### Schedule, Readings and Assignments

For the latest schedule, readings, and assignments, **consult the most recent schedule by going to the course website.**

* = Optional

| 13  | Nov. 26 | Event History Analysis: Interpretation and Implementation  
Cleves, Chapters 5-7, 9.1, 12.1, 13.1-13.3  
* Allison, Chapter 4  
Problem Set 6 Due |
| --- | --- | --- |
| 14  | Dec. 5 | Student Presentations  
Empirical Paper Due |

### Course & Instructor Policies

**Late Work.** Problem sets will not be accepted late, because the answers are discussed in class on the day they are due. Due to a medical emergency or other valid reason, you may be excused from turning in a problem set. In such cases, the grade will be computed based on the remaining problem sets. Consult me *in advance of the due date*, if at all possible, if such a contingency should arise. Likewise, I cannot give early or late examinations. Arrange your schedule now to avoid potential conflicts.

**Calculator.** A calculator is a virtual necessity for this class. However, any basic scientific calculator will do. The following functions are necessary: square root, $y^x$, $e^x$, and $\ln(x)$. Such calculators can often be obtained for under $10. You will not need nor use graphing capability or programmability.

**Attendance.** Attendance is entirely optional. However, be advised that you are responsible for any material covered in class, whether or not it was in the readings or lecture notes. You are also responsible for any announcements made in class. For most students, attendance is simply essential to learning the material. If you do need to miss a class, be sure to consult with a fellow student to learn what transpired.