EDUC 799 LAB NOTES
Categorical and limited dependent variables

Xiaoyang Ye
University of Michigan

December 25, 2016

Contents

EDUC 799 LAB #1
Non-Bayesian Empirical-minded `Stata` Runner

Xiaoyang Ye
University of Michigan

September 7, 2016

OBJECTIVES OF LAB 1

1. Introduction to lab
2. Getting started with `Stata`
3. Review of OLS regression using `Stata` (complete on your own if we don't finish)
4. Configuring AFS space and installing `SPOST`

1 Introduction to lab (welcome!)

- **Purpose:** Our weekly lab session is meant to supplement the lectures and readings, with an emphasis on applying the technical skills necessary to use a specific modeling technique and ensuring you can correctly interpret and communicate findings. We'll spend much of our time each week coding in `Stata`. Typically, the lab session will be spent trying the techniques discussed in lecture to prepare you for the coming week's assignment. When time allows, I will highlight `Stata` commands and workflow practices that may be helpful to you as you work on your final project (or/and other high quality quantitative projects).
- **What we expect of you:** Steve and I share a teaching philosophy of high expectations with high support. We expect you to come to lecture and lab having completed the readings and spent some time pondering their implications. We expect you to spend time struggling and muddling through material in an attempt to make sense of it before coming to us with questions; this is something you'll have to do as independent researchers. That said, do not suffer endlessly. We are here to help.
- **What you can expect from me:** You can expect careful preparation (e.g. handouts, lab files) and thoughtful instruction in lab. You can expect prompt and constructive responses to your emails. In addition, you can expect copious feedback on assignments. Whenever you have questions and comments, please let me know.
- **What you should expect of one another:** I expect we cultivate an environment of mutual respect and support. I hope that you will help one other through the rough patches, encourage one another throughout the course of your work, and be open to learning from one another. I aim to teach the lab "at the mean," so if you are a `Stata` pro and find yourself with free time in the lab, please help someone who is struggling.

- **What we will do in Lab:** Lab meets weekly on Wednesdays from 10 to 12 am. By 10:10 am, I expect you are in your seats, have loaded Stata, and navigated to the appropriate working directory. We will start promptly at 10:10 am. Please note that lab content is sequentially dependent: If you miss the first step in the process, you'll likely struggle to catch up with the rest of lab. Please be respectful of your peers and minimize tardiness.

- **You are very welcome to my office hours or email me:** My office hours are Mondays from 11 am to 1:30 pm (register at *Google Calendar*), but my schedule is flexible and I can meet with you at other times (Mon - Wed preferred; my office is 5204 Weill Hall) as well. Email me at *yxy@umich.edu* with subject starting with "799 - ."

- **Additional note (what I have learned from John DiNardo):** You may keep a self-created "notebook" for the class. The idea is that ultimately this notebook might prove useful to you when you are called to teach a class such as this, or as a reference for your own research. This notebook should contain, *inter alia*:

  - A Table of Contents and a bibliography
  - Summaries of lecture notes
  - All problem sets and problem set answers.
  - Very short summaries of assigned readings with critical commentary
  - Documented `Stata` (or other software) code and output when appropriate.
  - etc.

  You may find my notebook for this class at my personal website (not updated yet).

2 *A prior*: Non-Bayesian empirical-minded

(1) This class is about estimation of non-linear models, where the dependent variable is categorical or limited. We will focus on the conceptual and practical issues.

Below is a list of books are useful as theoretical (and empirical) reference.

- Wooldridge (2010), *Econometric Analysis of Cross Section and Panel Data*
- *Econometrics* by Bruce Hansen, you can download the 2016 version here.
- Hayashi (2000), *Econometrics*

Here is also a list of more empirical books.

- Angrist and Pischke (2008), *Most Harmless Econometrics*
- Murnane and Willett (2010), *Methods Matter*
- Cameron and Trivedi (2009), *Microeconometrics Using Stata*

(2) This class is not about causal inference and identification. But we will keep that in mind.

- Manski (2009), *Identification for Prediction and Decision*

- Imbens and Rubin (2015), *Causal Inference in Statistics, Social, and Biomedical Sciences*

(3) This class is not about *ex ante* evaluation ("structural").

- DiNardo and Lee (2011), *Program evaluation and research designs*
- Heckman and Vytlacil (2007a), *Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation*
- Heckman and Vytlacil (2007b), *Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments*
- Abbring and Heckman (2007), *Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation*

(4) This class is not about Bayesian econometrics. The key difference between Bayesians and non-Bayesians (frequenists) is about what "probability" is.

- Poirier (1995), *Intermediate Statistics and Econometrics*
- Greenberg (2012), *Introduction to Bayesian Econometrics*.

3 Getting started with `Stata`

Among the non-Bayesian empirical-minded regression runners, we primarily use `Stata`.

3.1 Understanding how `Stata` works

- `Stata` works through estimation command, which has two nice things:
  - `Stata` estimation commands have the same syntax (stored in .ado files)
    - `describe`
    - `regress`
    - `logit`
  - `Stata` estimation commands work with the postestimation commands
    - `test, lrtest`
    - `predict`
    - `margins`
- `Stata` runners "speak `Stata` " using syntax, which has four elements
  1. command name
  2. list of variable name (dependent variable is first)
  3. a comma
  4. an option

- Here is an outline of the tasks performed by a `Stata` estimation command.

  1. Parse the input to the command
  2. Compute results
  3. Store results in **e()** or **r()**, which makes the estimation-postestimation framework work
  4. Display output

- To make research producible, we use **.do** file that calls a number of `Stata` commands to perform a particular task on a specific dataset

  – Do-file programming: to write **.do** file programs to implement statistical methods
  – Ado-file programs are just automatic do-files

3.2 Workflow of data analysis using Stata

- There are two excellent references:

  – Mathew Gentzkow & Jesse Shapiro (2014), *Code and Data for the Social Sciences: A Practitioner's Guide*, you can download the handbook here
  – Long et al. (2009), *The workflow of data analysis using Stata*

- Steps in workflow (Long et al., 2009)

  1. Have a good idea for a project
  2. Prepare and clean the data for analysis
     – General goals and analysis plan
     – An organized director structure
     – Uniform formats for .do files
     – Look for the variables of interest
     – Variables must be carefully named, labeled, and cleaned
       * Categorical vs. continuous (e.g., code education)
       * Missing data
       * Truncated data
       * Drop cases
       * Random number generator with seed
  3. Conduct analysis
     – Estimate models
     – Postestimation
     – Create graphs
  4. Present results
     – Formats of tables and figures
  5. Protecting files
     – Replication is impossible without your data and .do files
     – Never rewrite original data files
     – Version control of .do files (+date and documentation)

3.3 A template .do file structure in “Lab 1 do" file

Stata codes

```
* ==============================================================================
* EDUC 799 FALL 2016
* LAB 1
* Sep 06, 2016
* PRACTICE DO FILE

* This do file is to help you get familiar with its basic commands

* CREATED BY:  XIAOYANG YE, AUG 26, 2016
* UPDATED BY:  XIAOYANG YE, SEP 06, 2016
* ==============================================================================


********************************************************************************
***** STANDARD HEADERS                                               *****
********************************************************************************

clear all
version 13
set more off, permanently
set matsize 5000


********************************************************************************
***** (1) SETUP WORKING DIRECTORY                                         *****
********************************************************************************

* Main work directory
cd  "/Users/yxy/Box Sync/Educ 799 - fall 2016/Lab/Lab 1/"

global do "do/"
global log "log/"
global data  "data/"
global table "table/"
global figure "figure/"



********************************************************************************
***** START A LOG FILE                                               *****
********************************************************************************

* Start a log file
set linesize 80
log using name.log, replace
```

```
*******************************************************************************
***** OPEN A DATA FILE                                                   *****
*******************************************************************************

* Open the "Iowa Data Files (ss-master).dta" file
use  "${educ799}Data/Iowa/Iowa Data File (ss-master.dta).dta", clear


*******************************************************************************
***** (1) EXPLORE THE DATA                                                *****
*******************************************************************************

*--------------------------------*
* (1.1) Describe basic statistics*
*--------------------------------*

* Overview of the data
describe

* Explore selected variables: (1) ACT Composite Score

lookfor act

codebook actcomp

*-------------------------*
* (1.2) Basic calculations*
*-------------------------*

* What's the percent of students in the data applied for Iowa?
lookfor apply
lookfor applied

tabulate app  /* 24,902/61,811; 40.29% */


*******************************************************************************
***** (2) REGRESSIONS                                                *****
*******************************************************************************

*----------------------------------------------------*
* (2.1) OLS: High School GAP and ACT Composite Score  *
*----------------------------------------------------*

*  Research question: Do female students have higher high school GPA, conditional
*  on ACT composite score?


*  Regress high school GPA on ACT composite score and gender
```

```
//check the variables first

* HS GPA
lookfor gpa
codebook hsgpa

cap drop m_hsgpa
gen m_hsgpa=(hsgpa==0)

replace hsgpa=. if m_hsgpa==1

* ACT composite score
codebook actcomp

* Female
codebook female

// OLS
regress hsgpa actcomp female

* Postestimation: Predict yhat and residuals

cap drop hsgpa_hat
predict hsgpa_hat, xb


********************************************************************************
***** END OF DO FILE                                                       *****
********************************************************************************

log close

/* convert log file to PDF */
translate log.log log.pdf

* Putting the exit command at the end of a dofile is a good idea.
exit, clear
```

4 Configuring AFS space

4.1 A "long" tradition of Educ 799: AFS space

- To minimize time spent in lab downloading files and configuring workspaces, use AFS space for lab (and assignments if you'd like).
- On UMich SITES or VirtualSites machines, your AFS workspace should be M:\ drive.

- On your own computer, you can access your AFS space at: http://mfile.umich.edu
- Creating lab folders
  - On your lab computer, double-click on the **My AFS Space** icon
  - Double-click on Private
  - Create a new folder **EDUC 799**
  - Make a directory structure
    - \EDUC 799\Labs\Lab1 (create "Lab 1 do" file in this folder)
    - \EDUC 799\Data (download Iowa data file to this folder)

4.2 In Xiaoyang's lab

- It is OK if you want to create these folders on your personal computer rather than on the AFS space
  - This reduces transition cost when you are not working in the computing lab
  - In both ways, we should keep the same directory structure from .../EDUC 799/...
- It is OK that you do not buy and install Stata on your personal computer
  - You can use PuTTY to connect to UMich's server, see steps here
  - You can transfer files to your AFS Space using WinSCP (windows) or FileZilla (Mac), see steps here

5 Additional section

Among the 13 respondents of the pre-class survey, nine reported that their learning goal of this class is to be able to write a high quality paper. To better help you gain this goal, I will provide you additional materials each week. Please let me know if you like this section, or you do not like it.

5.1 Additional reading

- DesJardins, Ahlburg, and McCall (2006), An integrated model of application, admission, enrollment, and financial aid. *Journal of Higher Education*, 77(3), 381-429.
  - A model paper of the applications of non-linear models
  - Same Iowa data
- Bettinger, Evans, and Pope (2013), Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy*, 5(2), 26-52.
  - A brilliant paper on the relationship between ACT and college performance

5.2 Paper review

This subsection is borrowed from the doctoral guide "How to summarize a paper" by Prof. Lee Sproull and Prof. Natalia Levina at NYU (01/09/2004).

- Theoretical/methodological articles

  1. What is the research question of interest?
  2. What theoretical basis is the article drawing from?
  3. What are the assumptions about social world and technology made by the authors?
  4. What is the proposed theory/framework/model?
  5. What are the key concepts?
  6. How the key concept relate to each other?
  7. what level of analysis is the theory targeted at?
  8. What is the contribution of the paper?
  9. What are the implications of the paper?
  10. What are the limitations?
  11. How can the model/theory/framework be applied to and/or tested in practice?

- Empirical articles

  1. What is the research question of interest?
     - Theoretical question
     - Empirical question
  2. What is the methodology used to conduct research?
  3. What are the key variables (independent and dependent, if appropriate)?
  4. What are the key findings (explain figures and tables)?
  5. What is the contribution of the paper?
  6. What are the implications of the work?
  7. What are the limitations of the work?
  8. Have these limitations been addressed by recent studies? And how?

5.3 `Stata` tip #1: Send an email using `Stata`

- `mail`: Module to send emails from Mac/Linux/Unix
- `emailme`: Module to send emails from Windows systems
- `email`: Module to send emails via Python

Stata codes

```
* Send an email using Mac's underlying sendmail program to send you emails

// Syntax
// mail From: x@X; To: y@Y; Subject #; # [using filename] [, option]

// Example
mail From: yxy@umich.edu; To: yxy@umich.edu; Subject: 799 Lab 1; ///
Lab 1 Do File has finished.

* It is useful when you run long do files remotely
```

EDUC 799 LAB #2
What Are We Regressing For?: OLS and LPM

Xiaoyang Ye
University of Michigan

September 14, 2016

OBJECTIVES OF LAB 2

1. Checklist
2. Review of OLS regression using `Stata`
3. Linear probability model (LPM)

1 Checklist

- Questions about week 1 lecture and lab?
- Stata programming questions in "lab 1" .do file?
- Notebook?
- Final project ideas or proposals?
- Assignment 1 due on Sep 20 (by the start of lecture)

2 Review of OLS regression using `Stata`

2.1 Replication Example

- Bettinger, Evans, and Pope (2013), Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy*, 5(2), 26-52.

2.1.1 Research question

- Using the ACT composite score, college admissions offices implicitly give equal weight to each of its four individual subtests
- Using four subjects in ACT test generates noise in identifying students at risk of underperforming and dropping out.
- Only two of the four subtests of the ACT, Englis and Mathematics, can effectively predict outcomes in college.

2.1.2 Policy implication

- A simple and low-cost change in the way colleges use the ACT exam in their admission decisions

2.1.3 Empirical strategy

- **Model:** Test for differences in predictive power among the four subscores using OLS
- **Possible data:** College student data including demographic information, college performance measures (GPA, drop-out), and ACT scores
  - *[Our Iowa data]* do not have college performance information
  - We cannot replicate Bettinger, Evans, and Pope (2013)
  - But we have information on both who applied and who were admitted

2.1.4 Extension: Predictive power of ACT subscores on admission using *our Iowa data*

- Sample: Applicants
- Outcome: Being admitted (=1, dummy variable)
  - **Question:** If we know nothing about the students in this dataset beyond their app behavior, how would we go about making a prediction about whether each student would be admitted or not (E(Yi=1)) ?
- Explanatory variables: ACT scores
- Controls (find and clean variables):
  - Cohort fixed effects
  - High school GPA fixed effects (why?)
  - Race and gender
  - High school fixed effects
  - (Campus fixed effects & college major fixed effects?)
- Baseline estimation model (linear probability model, sample model with subscript i):

$$Outcome_i = \beta_0 + \beta_1 ACTmath_i + +\beta_2 ACTEnglish_i + \beta_3 ACTreading_i + \beta_4 ACTscience_i + \epsilon_i \tag{1}$$

  - Outcome is dichotomous: admitted=1
  - Additional specifications: adding controls, more flexible functional forms (polynomials, interactions)
  - Hypothesis testing: $\beta_1 = \beta_2 = \beta_3 = \beta_4$

2.1.5 Results: Tables and figures

- Compare with Tables 1 in Bettinger, Evans, and Pope (2013)
- Caption, format, note

Table 1: Summary statistics on applicants (N=24,902)

Variables	(1) mean	(2) sd
Admission (=1)	0.888	0.315
Application (=1)	1.000	0.000
ACT composite score	24.180	3.840
ACT math score	24.027	4.447
ACT English score	23.776	4.493
ACT science score	23.857	3.952
ACT reading score	24.523	5.302
HS ranking percentile	72.086	19.087
Female (=1)	0.569	0.495
Black (=1)	0.026	0.161
American Indian (=1)	0.003	0.054
White (=1)	0.878	0.327
Asian (=1)	0.036	0.186
Hispanic (=1)	0.021	0.143
Other race (=1)	0.036	0.186

Notes: This table displays summary statistics of estimation sample, which consists students who applied for University of Iowa during 1999-2001.

- Compare with Tables 4 in Bettinger, Evans, and Pope (2013), see regression results in .do file
- Interpret results

Table 4—The Predictive Power of ACT Subscores on First-Year Dropout Indicator

Dependent variable: First-year dropout indicator	(1)	(2)	(3)	(4)	(5)	(6)
ACT composite score	−0.0088**					
	(0.0005)					
ACT Math score		−0.0065**	−0.0053**	−0.0031**	−0.0030**	−0.0025**
		(0.0006)	(0.0006)	(0.0006)	(0.0007)	(0.0007)
ACT English score		−0.0046**	−0.0035**	−0.0022**	−0.0021**	−0.0027**
		(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)
ACT Reading score		0.0009	0.0011*	0.0014**	0.0014**	0.0014**
		(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
ACT Science score		0.0010	0.0010	0.0013	0.0014	0.0012
		(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)
Campus fixed effects			X	X	X	X
High school GPA fixed effects				X	X	X
Race and gender fixed effects					X	X
College major fixed effects						X
Pseudo R^2	0.029	0.036	0.054	0.063	0.064	0.083
Observations	25,645	25,645	25,645	24,551	24,551	22,111

Notes: Marginal effects and robust standard errors are presented from logit regressions of a first-year dropout indicator on the composite ACT score (column 1) and each of the ACT subscores (columns 2–6). Campus fixed effects

2.1.6 Simulation: Changes using ACT math and English scores

- Ranking based on ACT composite, or ACT math and English, which will harm the disadvantaged?
- Demographics change using new approach
- In our data, college outcomes are unknown

3 Linear Probability Model

- In OLS, when the outcome variable is an indicator, we are regression a linear probability model
- All the interpretations of the LPM results are the same with general OLS results
- **Exception:** percentage point change in the outcome

4 Additional section

4.1 Additional reading

- Jacob, McCall, and Stange (2016), College as Country Club: Do Colleges Cater to Students' Preferences for Consumption?. *Journal of Labor Economics*, forthcoming

- Application of random utility maximization in college choice

4.2 Data sources for final project

You are free to use whatever data are available to you and/or the public for your final project.

- Common data sources for education come from the National Center for Education Statistics (e.g. HS&B, NELS, ELS, NPSAS, BPS, B&B, IPEDS). The federal government consolidates federally funded datasets online at data.gov.
- You might also consider reaching out to professional associations and non-profits.
- Finally, the University of Michigan is host to the Inter-university Consortium for Political and Social Research; ICPSR cleans, consolidates, and archives a vast range of data, including the Current Population Survey and the Panel Study of Income Dynamics.

4.3 `Stata` tip #2: Manage missing values using `missings`

- `missings`: A set of utility commands of managing missing values (., .a-.z in numeric variables, "" in string variables)
- Stata allows us to code different types of numeric missing values. It has 27 numeric missing categories. ".a" to ".z" and "."

Stata codes

```
* (1) Install "missings" package
ssc install missings

* (2) Example data set from Stata
webuse nlswork, clear

* (3) "missngs report" issues a report on the number of missing values
missings report
    // report for all variables
missings report year
    // report for year variable
missings report, min(1000)
    // report for variables with at least 1000 missing cases

* (4) "missings list" lists observations with missing values
missings list, min(5)
    // list observations with at least 5 missing variables

* (5) "missings table" tabulates observations by the number of missing values
missings table
by race: missings table
    // by race groups
```

```
* (6) "missings tag" generates a variable containing # of missing values
missings tag, generate(nmissing)

* (7) "missings dropvars" drops any variables which are missing on all values
gen frog = .
gen toad = .a
gen newt = ""

missings dropvars frog toad newt, force sysmiss
missings dropvars toad, force sysmiss
    // sysmiss applies to . missing only (not .a-.z in numeric and "" in string)

* (8) "missings dropobs" drops any observations which are missing on all values
set obs 30000
missings dropobs, drop
```

EDUC 799 LAB #3
The World is (Often) Non-linear

Xiaoyang Ye
University of Michigan

September 21, 2016

OBJECTIVES OF LAB 3

1. Checklist
2. Assignment 1 feedback
3. From LPM to non-linear models
4. Logit model for a binary world

1 Checklist

- Questions about lecture and lab?
- Final project 2-paragraph description due on Sep 27 (by the start of lecture)
- Assignment 2 due on Sep 27 (by the start of lecture)

2 Assignment 1 feedback: Elegant `Stata` Runners!

2.1 Analysis and writing

- Tell a story. Do not answer a question.
  - Treat the assignment as a mini paper: title, research question, data, methods, results; with possible theoretical discussion and implications
  - For example, "*In this paper, I estimate....*"
- Interpret coefficients
  - Sample? Comparison? Coefficient (unit change, but not UNIT change)? Significance? On average/holding all else equal?
  - Example #1: For those who applied for U of Iowa [sample], each one point increase [unit change] in ACT math score is statistically significant (p<0.001) [significance] associated [association, not causality] with 1.2percentage points increase in the probability of being admitted, holding all else constant [on average].

- Example #2: For those who applied for U of Iowa, black students on average are 0.5 percentage point less likely to be admitted (or, black students on average have 0.5 pp lower probability of admission); however, this correlation is not statistically significant [from 0].

- **Describe your data**: data source, sample, sample size; describe variables using numbers (mean, sd, %)
  - **Be careful about the sample you are using: full sample vs. sub-sample?** Explain why.
- Describe your methods: "Regression results...", what regression?
- Writing Tips for Ph.D. Students by John Cochrane that I read yesterday (09/20/2016)

2.2 Format

- **Table:** title/caption, row and column items, notes
- **Blind review**: Do not include your name in the text or title. Do files and log files are OK as we review your writing and data analysis separately.
  - PDF file name: UMID_Assignment_#.pdf
- Submit the do file and log file with relevant results
  - For a few journals, authors are required to submit replication data and do files
  - In the do file, indicate which part produces which table/figure

2.3 `Stata` programming

- `capture` command
- Output regression results using `outreg2`
  - I will share with you other ways in the coming weeks
- `predict` command: it will predict for the full sample even you have estimated a sub-sample
  - In analysis, it is called "out of sample prediction"
  - In Assignment 1, you should only predict the admission probabilities for those applicants

3 From LPM to non-linear models

3.1 Latent variable

- A latent or unobserved variable $y*$ ranging from $-\infty$ to ∞ that is related to the observed independent variables by the structural model:

$$y* = \mathbf{X}_i\beta + \epsilon \tag{2}$$

- **A link function** (cumulative density function, CDF) to the binary observed variable y:

$$y = \begin{cases} 1 & if \quad y^* > 0 \\ 0 & if \quad y^* \leq 0 \end{cases} \tag{3}$$

- From LPM to non-linear function: The probability of the event conditional on X is the cumulative density function of ϵ evaluated at $X\beta$:

$$Pr(y = 1|X) = Pr(y* > 0|X) = Pr(\varepsilon > -X\beta) = \mathbf{F}(X\beta) \tag{4}$$

  - If the $\mathbf{F}(.)$ function is linear, this is a linear probability model (LPM)
  - The $\mathbf{F}(.)$ function can be logit or probit
    - ∗ Logistic equation

$$Pr(y = 1|X) = \frac{e^{X\beta}}{1 + e^{X\beta}} \tag{5}$$

    - ∗ Why we need this transformation?
      - · Odds equation

$$\frac{p}{1-p} = e^{X\beta} \tag{6}$$

      - · Logit model (linearity in parameters like LPM)

$$\ln(\frac{p}{1-p}) = X\beta \tag{7}$$

    - ∗ Probit model

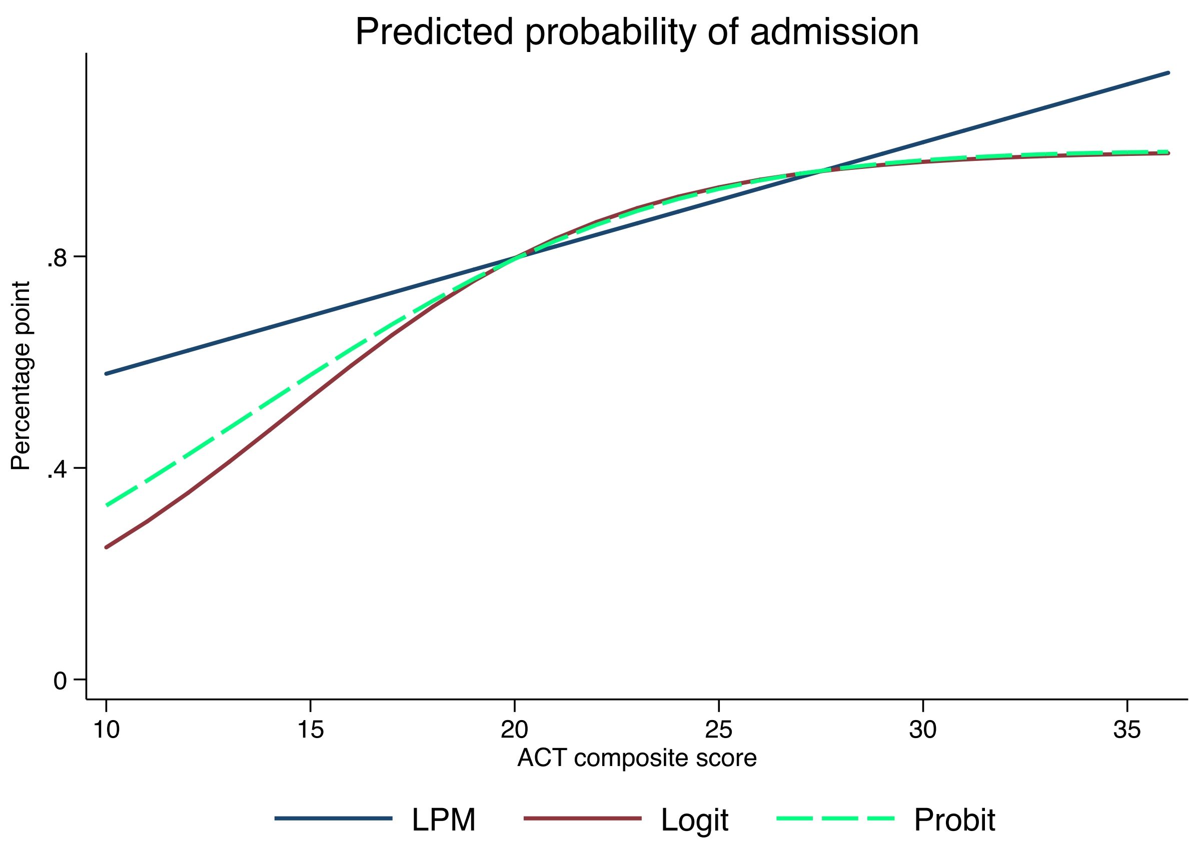
$$Pr(y = 1|X) = \int_{-\infty}^{X\beta} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt \tag{8}$$

3.2 Percentage points vs. percent change

- **Percent (%) is derived from Latin *per centum* (per 100).**
  - If I got a 50% on a quiz, that means that I got 50 out of 100 right
  - Put in the language of probabilities: 50 out of 100 tires were success
- **Percentage point (pp, or %age point) is an absolute difference**
  - %A-%B=pp
  - If Steve's predicted probability of graduation is **90%** and Brian's predicted probability of graduation is **50%**, then Steve is **40 percentage points** more likely to graduate.
- **Percent higher/lower is a relative difference**
  - (%A-%B)/%B
  - Steve's likelihood/probability of graduation is **80%** higher than Brian's.

3.3 Tradeoff between LPM and non-linear models

- LPM
  - It's computationally simpler
  - It's easier to interpret the "marginal effects"
  - It avoids the risk of mis-specification of the "link function"
  - There are complications with Logit or Probit if you have endogenous dummy regressors
  - The estimated marginal effects from the LPM, Logit and Probit models are usually very similar, especially if you have a large sample size.
- Logit or Probit model
  - LPM yields biased and inconsistent estimates in many circumstances; MLE is consistent
  - LPM does poor job in estimating marginal effects in extreme tails
- Check your case by comparing LPM, Logit and Probit
- If we only care about the direction of the estimate, LPM is OK
- Use non-linear models for multinomial/count/truncated outcomes



4 Logit model for a binary world

4.1 Basic estimation

- `Logit` fits a logit model for a binary response by maximum likelihood
- It models the probability of a positive outcome ($Pr(y = 1|X)$) given a set of regressors (X)
- Command: `logit [Y] [X] [if ...]  [weight], [options]`
- Output
  - *Iteration 0: log likelihood*: null log likelihood
  - display results in odds ration by adding `or` in `[options]`, or use `logistic` command
  - `listcoef` postestimation command
- Observations predicted perfectly
  - MLE is not possible when `Y` does not vary within one of the categories of an independent variable
    - Example: All students in the top quartile of ACT composite score are admitted

4.2 Interpretation of log odds and odds

- Relationship between log odds and odds

$$odds = exp(\log odds)) = \frac{p}{1-p} \tag{9}$$

<u>Stata codes</u>

```
// log odds<0, odds<1
// log odds=0, odds=1
// log odds>0, odds>1

preserve
    clear
    set obs 200
    gen odds=0

    forvalues i=1/200 {
        replace odds=`i'*0.1 in `i'
    }

    sum odds
    gen log=ln(odds)

    twoway (connected odds log, xline(0) yline(1))
restore
```

- **Log odds**
  - Each one unit change in `x` is associated with β increase in the log odds (logit) of `Y`, controlling for other variables
  - Log odds is additive
$$\log A + \log B = \log(A * B) \tag{10}$$
  - Log odds is not a measure with intuitive value
- **Odds**
  - Each a unit change in `x` is associated with $\exp(\beta)$ factor change in the odds of the event, *ceteris paribus*
  - Odds is multiplicative
$$\exp(A) * \exp(B) = \exp(A + B) \tag{11}$$
  - Reverse odds $\frac{1}{\exp(\beta)}$ for failure (`listcoef, reverse`)
- **Standardized odds ratio**
  - Change in one SD of `x`
  - $Odds = exp(\beta * StdX)$

4.3 Predicting probability

- Prediction after `logit` estimation

Stata codes

```
// Logit estimation
logit Y X, robust

// Predicting probability
cap drop pr_Y
predict pr_Y, pr

// Linear prediction of log odds
cap drop pr_log
predict pr_log, xb

// Create probability using log odds
cap drop pr
gen pr=exp(pr_log)/(1+exp(pr_log))
```

- Understanding probability and log odds
  - In the null model (without X in the model)
  - The intercept is the log odds of overall probability of Y

5 Additional section

5.1 Additional reading

- Hilger (2016), Parental Job Loss and Children's Long-Term Outcomes: Evidence from 7 Million Fathers' Layoffs. *American Economic Journal: Applied Economics*, *8*(3), 247-283.
  - Causal impacts of parental layoffs on college enrollment, college quality, and early career earnings
  - OLS of binary outcomes, using differences-in-difference
  - Very clear applied paper structure: estimated impacts, robustness checks, heterogeneity and mechanisms

5.2 `Stata` Tutorial

- I feel that I have not provided a complete introduction to `Stata`. For those who might be interested in gaining a better picture of `Stata`, please review the following excellent tutorial materials:
  - http://data.princeton.edu/stata/
  - http://www.ats.ucla.edu/stat/stata/
  - http://blog.stata.com/

5.3 `Stata` tip #3: LOOP with `forvalues` and `foreach`

- A local macro loop command that perform some commands repeatedly and efficiently
- `forvalues`: fastest way to loop over consecutive values (numbers from 1 to k)
- `foreach`: both number list and variable list

Stata codes

```
* (1) forvalues
// forvalues local_name=#/# {
       commands referring to `local_name'
   }

// example
forvalues i=1/3 {
        display "i is now `i'"
   }

* (2) foreach
// foreach local_name in list {
        commands referring to `local_name'
   }
```

```
// example #1
local vlist y x1 x2
foreach v of local vlist {
        display "v is now `v'"
   }

// example #2
local v "3"
display "v is now `v'"

local vlist y x1 x2
foreach v of local vlist {
        display "v is now `v'"
   }

 display "v is now |`v'|"

* (3) More examples from Stata "help foreach"
   // Append a list of files to the current dataset.

   foreach file in this.dta that.dta theother.dta {
         append using "`file'"
   }

   //  Loop over new variables

   foreach var of newlist z1-z20 {
         gen `var' = runiform()
   }
```

EDUC 799 LAB #4
Probability or Likelihood, Which to Maximize?

Xiaoyang Ye
University of Michigan

September 28, 2016

OBJECTIVES OF LAB 4

1. Checklist
2. Assignment 2 feedback
3. Maximum likelihood estimation

1 Checklist

- Questions about lecture and lab?
- Sign-up on Canvas for a meeting with Steve and/or Xiaoyang
- Work on your final project (literature review)
- No assignment this week

2 Assignment 2 feedback

- `Stata` analysis
  - Be care of your sample vs. sub-sample
  - How to describe your variables?
  - Output raw coefficients and odds ration into tables
  - Null likelihood vs. final likelihood
- Write-up
  - Justify your model choice (always in your quantitative papers), and all other choice (data, variable, question)
  - To explain odds ratio or probability (marginal effect)?
  - Word choice (I like to use "I" and "Table * shows that"....)
  - Introduction (why I should read your paper?)/summary (what I can take away from your paper?)

3 Maximum Likelihood Estimation

- **Probability** is used before data are available to describe possible future outcomes given a fixed value for the parameter (or parameter vector)
- **Likelihood** is used after data (X) are available to describe a function of a parameter (or parameter vector, β) for a given outcome (Y)

3.1 MLE with basic calculations

3.1.1 Probability in binomial distribution $X \sim B(n, p)$

- with parameters p (probability of success) and n (number of independent yes/no experiments)
- *Bernoulli experiment:* $n = 1$
- **Probability mass function: k successes in n trails**

$$f(k; n, p) = Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \tag{12}$$

  - In a sample of 10 students who are admitted to CSHPE's Ph.D. program, what's the probability that 5 will eventually enroll, given the historical yield rate is 0.3?
  - $n = 10$, $p = 0.3$, $k = 5$

$$Pr(X = 5) = \binom{10}{5} 0.3^5 (1 - 0.3)^5 = 0.102919 \tag{13}$$

Stata codes

```
// cumulative probability (X<=5)
di binomial(10, 5, 0.3)
bitesti 10 5 0.3

// probability X=5
di binomial(10, 5, 0.3) - binomial(10, 4, 0.3)

// calculation
di ((10*9*8*7*6)/(5*4*3*2*1)) * 0.3^5 * (1-0.3)^5
```

- In a binary outcome model, p is the **parameter** that we are going to estimate

3.1.2 Likelihood: Use data to guess what p is

- Now, assume we have the enrollment data that 5 students have eventually enrolled at CSHPE, we can update the "historical" yield probability.
- We know $n = 10$, $k = 5$, but we do not know $p =$???
- Maximum Likelihood Estimation: Use iteration to guess $\hat{p}$ (model) that best fits the data

– Maximum likelihood that we can observe $k = 3$ in $n = 10$ with our guessed $\hat{p}$

– Start with a random number (guess): $p = 0.5$ (note: $p \in [0, 1]$)

$$Likelihood = L = \binom{10}{5} 0.5^5 (1 - 0.5)^5 = 0.24609375 \tag{14}$$

– Next, make a guess of 0.6, resulting a smaller likelihood

$$Likelihood = L = \binom{10}{5} 0.6^5 (1 - 0.6)^5 = 0..20065812 \tag{15}$$

– So, change to the other direction (around 0.5), a guess of 0.4, a larger likelihood

– Continuing iteration, until the largest likelihood has been found...

– $p_{MLE} = 0.5$ (sample mean)

Stata codes

```
clear
set obs 100

// create guessed phats
gen phat=0
forvalues i=1/100 {
replace phat=`i'/100 in `i'
}

// generate estimated likelihoods
cap drop likelihood
gen likelihood=0
replace likelihood=binomial(10, 5, phat) - binomial(10, 4, phat)

// find the largest likelihood
sum likelihood
list phat if likelihood==r(max)

// graph
twoway (line likelihood phat, xline(0.5, lcolor(gs12)) ///
yline(0.2460938, lpattern(longdash) lcolor(maroon)))
```

3.1.3 From aggregated data to individual data

- What's the likelihood in Table 2?
- What's the log-likelihood? Is it very small?

Stata codes

```
clear
set obs 10

// generate ID
gen ID=0
forvalues i=1/10 {
replace ID=`i' in `i'
}

// generate y
gen y=0
replace y=1 if ID<6

// generate p
gen p=0.5

// generate p^y
gen py=p^y

// generate (1-p)^(1-y)
gen py1=(1-p)^(1-y)

// generate p^y * (1-p)^(1-y)
gen total=py*py1

// generate (log) likelihood
egen ll=total(ln(total))
/* (exp(logx1+logx2)=x1*x2) */
gen L = exp(ll)
```

Table 2: Calculation of likelihood using individual data

ID	y_i	p_i	$p_i^{y_i}$	$(1-p_i)^{(1-y_i)}$	$p_i^{y_i} * (1-p_i)^{(1-y_i)}$
1	1	0.5	0.5	1	0.5
2	1	0.5	0.5	1	0.5
3	1	0.5	0.5	1	0.5
4	1	0.5	0.5	1	0.5
5	1	0.5	0.5	1	0.5
6	0	0.5	1	0.5	0.5
7	0	0.5	1	0.5	0.5
8	0	0.5	1	0.5	0.5
9	0	0.5	1	0.5	0.5
10	0	0.5	1	0.5	0.5

- Example: Logit model

```
clear
set obs 4

// generate ID
gen ID=0
forvalues i=1/4 {
replace ID=`i' in `i'
}

// generate y
gen y=0
replace y=1 if ID<3

// generate x
gen x=2
replace x=5 in 2
replace x=3 in 3

// generate b0
gen b0=0.5

// generate b1
gen b1=0.5

// generate b0+bx (linear XB)
gen xb=b0+b1*x

// generate phat [LOGIT MODEL]
gen phat=1/(1+exp(-xb))

// generate y^phat
gen yp=phat^y

// generate (1-phat)^(1-y)
gen yp1=(1-phat)^(1-y)

// generate log likelihood
egen ll = total(ln(yp*yp1))
egen tmp = total(y*ln(phat)+(1-y)*ln(1-phat))

// ITERATION: change b0 and b1 to find the maximum log likelihood, give x and y
```

3.2 How MLE works?

1. The link function (cumulative distribution function, CDF) of logit model

$$Pr(y=1|X) = \mathbf{F}(X\beta) = \frac{e^{X\beta}}{1+e^{X\beta}} = \frac{1}{1+e^{(-X\beta)}} \tag{16}$$

2. Likelihood function: in the sample there are N observations

$$L = \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{1-y_i} = \prod_{i=1}^{N} \mathbf{F}(X\beta)^{y_i}(1-\mathbf{F}(X\beta))^{1-y_i} \tag{17}$$

3. Then, the log-likelihood function is

$$lnL = l = \sum_{i=1}^{N}[y_i ln\mathbf{F}(X\beta) + (1-y_i)ln(1-\mathbf{F}(X\beta))] \tag{18}$$

   - Logit model: The first derivative (probability distribution function, PDF) of $\mathbf{F}(X\beta) = \frac{e^{X\beta}}{1+e^{X\beta}}$ is $\mathbf{F}(X\beta)(1-\mathbf{F}(X\beta))$

4. The first order conditions are nonlinear and non-analytic, that we obtain the MLE using numerical optimization methods (the Newton-Raphson method)

$$\widetilde{\beta_{n+1}} = \widetilde{\beta_n} - \left[\frac{\partial^2 l}{\partial\beta\partial\beta'}\right]^{-1}_{\beta=\widehat{\beta_n}} \left[\frac{\partial l}{\partial\beta}\right]_{\beta=\widehat{\beta_n}} \tag{19}$$

   - Hessian matrix & score matrix
   - Until $|\widetilde{\beta_{n+1}} - \widetilde{\beta_n}| < \varepsilon$

5. MLE theorem (asymptotic normal)

$$\sqrt{N(\widetilde{\beta_{ML}} - \beta)} \xrightarrow{asy} \left(0, N\left(\mathbf{E}\left[\left[\frac{\partial^2 l}{\partial\beta\partial\beta'}\right]^{-1}_{\beta=\widehat{\beta_n}}\right]\right)\right) \tag{20}$$

4 Additional section

4.1 Additional reading

- Long (2004), How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, *121*(1), 271-296.
- Posselt et al. (2012), Access without equity longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004. *American Educational Research Journal*, *49*(6), 1074-1111.
  - Cross-cohort comparison using NCES datasets
  - How to treat compositional change vs. mechanism change?

4.2 Non-`Stata` tip #4: What are we clustering for?

- Non-independence of data
  - Standard error with independently distributed observations (each n contributes the same amount of information)
$$se = \frac{sd}{\sqrt{N}} \tag{21}$$
  - To adjust the cases when an extra observation does not contribute the same amount of information because of cluster
    - E.g., students in the same classroom can be very similar
  - Effective sample size may be smaller, even though the sample is very large
    - ICC and sample size (average # per cluster) matters

Stata codes

```
*===================
* Numeric example *
*===================

// input data
input cluster x
a 1
a 2
b 2
c 3
c 4
d 4
d 5
e 5
e 6
f 6
g 7
g 8
h 8
h 9
i 9
i 10
j 10
j 11
k 11
k 12
end

// Intraclass correlation (ICC)
loneway x cluster
     // ICC is 0.49
```

```
// VIF (variance inflation factor) = 1 +(mean # per cluster - 1)ICC
di "VIF = " 1 +(4-1)*0.49  = 2.47

// Naive s.e.
sum x
di "se = sd/(sqrt(N)) = " r(sd)/sqrt(8)
reg x

// From ICC that N=8 is wrong (2.47 times high)
sum x
di r(sd)/sqrt(8/3.15)
reg x, cluster(cluster)
/* regression with cluster also adjusts for robust s.e. */

*===========================================
* Empirical example: Mincerian equation  *
* % increase in wage per 1 schooling year*
*===========================================

// sys. data set
webuse regsmpl, clear

// OLS without s.e. correction
 regress ln_wage grade tenure tenure2 i.year


 // OLS with robust s.e.
 regress ln_wage grade tenure tenure2 i.year, robust

 // OLS with cluster s.e.
 regress ln_wage grade tenure tenure2 i.year, vce(cluster id)

 // Panel data
 xtset id year

 xtreg ln_wage grade tenure tenure2 i.year
   xtreg ln_wage grade tenure tenure2 i.year, robust
     xtreg ln_wage grade tenure tenure2 i.year, cl(id)
```

EDUC 799 LAB #5
A Tale of Two Brothers: Logit vs. Probit

Xiaoyang Ye
University of Michigan

October 4, 2016

OBJECTIVES OF LAB 5

1. Checklist
2. Maximum likelihood estimation continued: Probit
3. Marginal effects

1 Checklist

- Questions about lecture and lab?
- Use office hours as research meetings
- Work on your final project (question, literature, data)
- Assignment 3 is due next Tuesday (Oct 11)

2 Maximum Likelihood Estimation Cont'd

2.1 Logit and probit model

1. The link function (cumulative distribution function, CDF) of logit/probit model

$$\begin{aligned} Pr(y=1|X) &= Pr(y*>0) \\ &= Pr(X\beta+\varepsilon>0) \\ &= Pr(\varepsilon>-X\beta) \\ &= Pr(\varepsilon<X\beta)[symmetry of normal distribution] \\ &= \mathbf{F}(X\beta) \end{aligned} \tag{22}$$

- Logit

$$\mathbf{F}(X\beta)=\frac{e^{X\beta}}{1+e^{X\beta}}=\frac{1}{1+e^{(-X\beta)}} \tag{23}$$

- Probit

$$\mathbf{F}(X\beta)=\Phi(X\beta) \tag{24}$$

2. Likelihood function: in the sample there are N observations

$$L = \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{1-y_i} = \prod_{i=1}^{N} \mathbf{F}(X\beta)^{y_i}(1-\mathbf{F}(X\beta))^{1-y_i} \tag{25}$$

3. Then, the log-likelihood function is

$$lnL = l = \sum_{i=1}^{N}[y_i ln\mathbf{F}(X\beta) + (1-y_i)ln(1-\mathbf{F}(X\beta))] \tag{26}$$

2.2 MLE with `Stata`

- Null model and log likelihood in the logit model
  - Intercept: overall probability (log odds) of the positive outcome
  - Recall that
$$L = \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{1-y_i} = \prod_{i=1}^{N} \mathbf{F}(X\beta)^{y_i}(1-\mathbf{F}(X\beta))^{1-y_i} \tag{27}$$
  - Without information in X, P is given by the sample average
- Logit model with many categories
  - Include one X variable: female
  - Intercept: overall probability for male
- Maximum likelihood estimation: `ml`
  - In `Stata`, type "`help ml`"
- OLS is a particular case of MLE under the normality assumption for the error terms
  - OLS: Choose a set of b to minimize sum of squared residuals (SSR):

$$SSR = \sum_{i=1}^{n}(y_i = x_i'b)^2 = (y-Xb)'(y-Xb) \tag{28}$$

  - Estimated coefficients:

$$\hat{\beta} = argminSSR(\beta) = (\frac{1}{n}\sum_{i=1}^{n} x_i x_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} x_i y_i' \tag{29}$$

  - Normality assumption: $Y_i = X_i\beta + \varepsilon_i$, $\varepsilon_i \sim N(0,\delta^2)$

$$L(Y_i;\beta,\delta^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\delta^n}exp(-\frac{1}{2\hat{\delta}^2})(\sum_{i=1}^{n}(Y_i - X_i\beta)^2) \tag{30}$$

3 Probability and marginal effects

- Predicted probabilities change given a discrete change (e.g., 0 to 1 for a dummy X) or the instantaneous rate of change for continuous variables
- We can compute the marginal effects by hand or using `Stata` commands
  - `margins`
  - `mchange`, `mgen`, `mtable`
- Three types of marginal effects
  - Average marginal effects (AME) are computed by averaging over the sample.
  - Marginal effects at the mean (MEM)
  - Marginal effects at representative value (MER)

4 Additional section

4.1 Additional reading

- Ebenstein, Lavy, and Roth (2016), "The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution." *American Economic Journal: Applied Economics*, 8(4): 36-65.
  - Transitory PM2.5 exposure is associated with a significant decline in student performance
  - Then is negatively correlated with postsecondary educational attainment and earnings

4.2 `Stata` tip #5: A way to leaner, faster graphs

- By Patrick Royston: http://www.stata-journal.com/sjpdf.html?articlenum=gr0013

```
If you have many variables, consider doing a preserve of the data and dropping
several of them before drawing a graph. This greatly speeds up production.

Take plotting fitted values from a model as an example. If there are many tied
observations at each value of the predictor and therefore many replicates of
the fitted values, the size of the graph file can be large, also making the
plotting time large.

A construction like this can save resources:
. preserve
. bysort x: drop if _n > 1
. line f1 f2 f3 x, sort clp(l - _) saving(graph, replace)
. restore
```

```
Here is another real example: with 15,156 variables and 50 observations, I
wanted a dotplot of variable v15155 by v15156. The time taken with all data
present was 10.66 seconds, but with preserve and all irrelevant variables
dropped, it was 0.69 seconds.
```

EDUC 799 LAB #6
If You Ever Feel Useless, Just Remember That Someone Is A Lifeguard at the Olympics Swimming Event.

Xiaoyang Ye
University of Michigan

October 12, 2016

OBJECTIVES OF LAB 6

1. Checklist
2. Last lesson (of life) on binary response models
3. Discussion: How your model fits your data?

1 Checklist

- Questions about lecture, lab and office hours?
- Work on your final project (question, literature, data)
- Assignment 4 and Literature review are due next **Wednesday (Oct 19)**

General comments on Assignment #2

- Choose a title to present your key information (and introduction/conclusion)
- You are estimating "the enrollment model"
- *Ceteris paribus*
- When you have a new finding, you may want to explain it

2 Binary Response models: A summary

2.1 Set-up

In real life, we often make yes/no choices that usually take two values for the outcome (1 and 0). Such variables are called *dummy* or *dichotomous* or *binary response* variables.

- To be, or not to be?
- Love me, or her?

- To enroll at U Michigan, or not?

If the number of your options exceeds two, we will be using other multi-responses models, which are based on the binary response models.

2.2 Probability

The expected value of a dummy variable $y_i \in \{1, 0\}$ is the probability of positive outcome (taking the value 1):

$$E(y_i) = 1 * P(y_i = 1) + 0 * P(y_i = 0) = P(y_i = 1) \tag{31}$$

2.3 Linear Probability Model

$$y = x\beta + \varepsilon, E(\varepsilon|x) = 0; P(y_i = 1) = x\beta \tag{32}$$

Problems with LPM: Estimates and statistical inferences

- The error term ε is heteroscedastic

$$Var(\varepsilon|x) = P(y = 1)(1 - P(y = 1)) = x\beta(1 - x\beta) \tag{33}$$

  - Can be addressed using WLS
- The error term is not normally distributed (t-test is not guaranteed)
- Predicted $\hat{y}$ can lie outside $[0, 1]$ that does not represent a probability
- Incorrect linearity assumption
  - Estimates will be highly sensitive to the range of data observed in the sample
  - May understate/overstate the magnitude of the true effects
  - As the underlying estimation function is similar $P(y = 1) = \mathbf{F}(x\beta)$, LPM gives the correct sign of the effect

Mostly harmless econometrics: Probit better than LPM?

- Practical advantages of LPM
  - Easier to calculate
  - The parameters are directly interpretable
  - Fixed effects and instrumental variables estimators can easily be implemented
    - Note: Adding fixed effects as dummy variables in the probit or logit model will yield biased estimates (when within group N is small)
    - The statistical problem is that, as the number of groups tends to infinity, the number of estimated parameters increases at the same rate. The estimates are not consistent.
    - See more in Greene (2004), The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, *7*(1), 98-119.

- Comments on model choice from Jorn-Steffen Pischke:

```
We care about the marginal effects. The LPM will do a pretty good job estimating
those. If the CEF (conditional expectation function) is linear, as it is for a
saturated model, regression gives the CEF - even for LPM. If the CEF is
non-linear, regression approximates the CEF. Usually it does it pretty well.
Obviously, the LPM won't give the true marginal effects from the right nonlinear
model. But then, the same is true for the "wrong" nonlinear model!

The fact that we have a probit, a logit, and the LPM is just a statement to the
fact that we don't know what the "right" model is. Hence, there is a lot to be
said for sticking to a linear regression function as compared to a fairly
arbitrary choice of a non-linear one! Nonlinearity per se is a red herring.
```

2.4 Logit and Probit

- Binary response models directly describe the probability $P(y_i = 1)$ of the dependent binary variable y_i, using an index function (latent variable):

$$P(y = 1|x) = P(y* > 0|x) = P(x\beta + \varepsilon > 0) = \mathbf{F}(x\beta) \tag{34}$$

- The index function maps the single index into dichotomous choice $[0, 1]$

$$\mathbf{F}(-\infty) = 0, \mathbf{F}(\infty) = 1 \tag{35}$$

  - $\mathbf{F}(x\beta)$ is NOT a linear of β

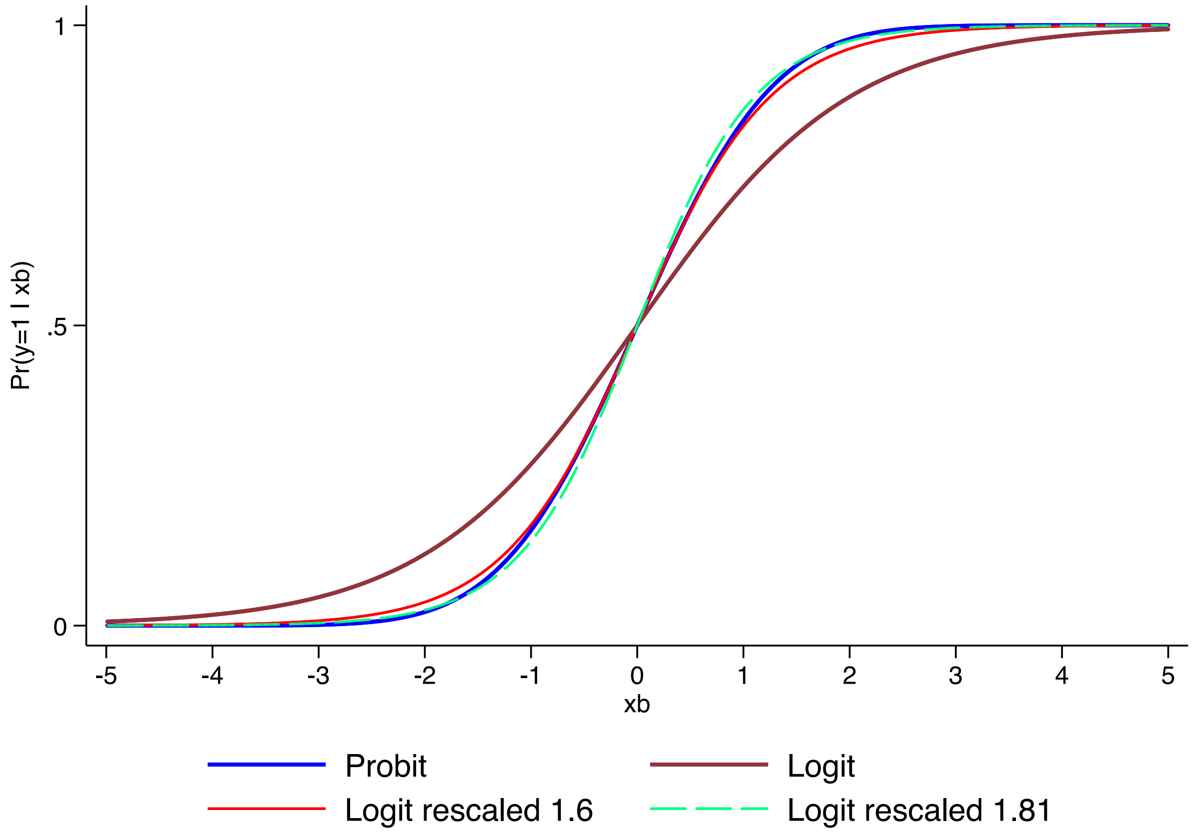
- Logit and Probit models are almost identical and the model choice is usually arbitrary
  - Logit: $P(y = 1|x) = \mathbf{F}(x\beta) = \frac{1}{1+\exp^{-x\beta}}$
  - Probit: $P(y = 1|x) = \mathbf{F}(x\beta) = \Phi(x\beta) = \int_{-\infty}^{x\beta} \frac{1}{\sqrt{2\pi}} \exp^{-\frac{1}{2}t^2} dt$

Stata codes

```
* Compare Logit and Probit models
// global date "10_12_2016"

// set observations
clear
set obs 1000

// create a sequence of numbers (-5, 5)
gen xb = _n*0.01 - 5
sum xb
```




```
// create logit and probit
gen logit = 1/(1+exp(-xb))
sum logit

gen probit = normal(xb)
sum probit

// create scaled logit
gen logit_scale1 = 1/(1+exp(-xb*1.6))
gen logit_scale2 = 1/(1+exp(-xb*1.81))

// graph
twoway  (line probit xb, lwid(medthick)) lcolor(blue)   ///
(line logit xb, lwid(medthick))   ///
(line logit_scale1 xb, lcolor(red))    ///
(line logit_scale2 xb, lcolor(mint)   ///
lpattern(longdash)  ///
xtitle("xb", size(small))  ///
ytitle("Pr(y=1 | xb)", size(small))  ///
ylabel(0(.5)1, nogrid labs(small) angle(h))    ///
xlabel(-5(1)5, angle(0) labs(small))  ///
```

```
legend(order(1 2 3 4 ) label(1 "Probit") label(2 "Logit")    ///
label(3 "Logit rescaled 1.6") label(4 "Logit rescaled 1.81")    ///
cols(2) region(color(none)) margin(zero))    ///
graphregion(color(white)) bgcolor(none))

graph save "logit_probit_${date}.gph", replace
graph export "logit_probit_${date}.pdf", replace
erase "logit_probit_${date}.gph"
```

2.5 Estimation using Maximum Likelihood

- The likelihood function is:

$$\mathcal{L} = \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{1-y_i} = \prod_{i=1}^{N} \mathbf{F}(x\beta)^{y_i}(1-\mathbf{F}(x\beta))^{1-y_i} \tag{36}$$

- The corresponding log likelihood function is:

$$log\mathcal{L} = \sum_{i=1}^{N}[(1-y_i)log(1-\mathbf{F}(x\beta)) + y_i log\mathbf{F}(x\beta)] \tag{37}$$

- No analytical solution to the FOCs and numerical optimization routines are used (iterations)

2.6 Marginal effects

- LPM
  - Raw coefficients (β) are average marginal effects
  - Important to report robust s.e. due to heteroscedasticity
- Logit and Probit
  - Raw coefficients are NOT marginal effects
    - Logit: log odds (log of odds ratio)
    - Probit: z-score
  - Marginal effect of $x_{ik} = \frac{\partial P(y_i=1|x_i)}{\partial x_{ik}}$, changes in predicted probabilities given a marginal change in x
    - Logit: $\frac{exp(x\beta)}{(1+exp(x\beta))^2}\beta_k$
    - Probit: $\phi(x_i\beta)\beta_k$
  - Marginal effects depend on all $x_i k$ that any individual has a different marginal effect
  - Average marginal effect; marginal effect at mean/median; marginal effect at any values

2.7 More topics

- `cloglog`, assuming the error term is log-log distributed (rather than logit/probit)
- `scobit`, relaxing the assumption that the marginal effect is greatest when $P(y = 1) = 0.5$
- `hetprobit`, heteroscedasticity in probit model
- `ivprobit`, endogenous regressors
- `biprobit`, seemingly unrelated regression
- `xtlogit` or `xtprobit`, panel data

3 Hypothesis testing and model fit

3.1 Inference based on the log likelihood function

- By definition, $\hat{\beta}$s maximize the likelihood of observing the sample. All other parameter values will lower the log likelihood value.
- How much log likelihood will change if we move $\hat{\beta}$s away from the ML estimators
- **LR test**:
  - Under null hypothesis (validity of restriction), LR is asymptotically a chi-squared distribution with q degrees of freedom, where q is the number of restrictions

$$LR = 2(ln\mathcal{L}_{unrestricted} - ln\mathcal{L}_{restricted}) \tag{38}$$

    * The LR test is often used to test whether a sub-set of Xs can be omitted from the (unrestricted/full) model
    * `Stata` computes critical values for chi2 of degree 1 at 0.05 level: `invchi2tail(`q`, 0.05)`
    * If the log likelihood ratio falls very slightly (the function is flat), we may not be able to reject the null

3.2 T-test: Standard errors for parameters and marginal effects*

- The conventional estimator for the covariance matrix for parameters is based on the inverse of the negative Hessian
- The robust S.E. is computed by adding a **score** vector in a "sandwich" formula
- S.E. for marginal effects are computed using the **delta method** (Taylor series approximation)

3.3 Goodness of fit #1: Percent correctly predicted

- Using ML estimators of β, we can predict $P(y = 1)$ for every observation in the sample (out of sample observations as well)
- We then set the predicted outcome of y_i as 1 if $\hat{p} > 0.5$
- The percentages of times the predicted $\hat{y}_i$ matches the actual y_i is the percent correctly specified
- It is a good idea to report for both outcomes (0, 1)

3.4 Goodness of fit #2: Pseudo R-squared

- One of the most used pseudo R-squared measures

$$\tilde{R}^2 = 1 - \frac{ln\mathcal{L}_{unrestricted}}{ln\mathcal{L}_{restricted}} \tag{39}$$

  - Notice: $\tilde{R}^2$ uses the same information of LR test
  - If the full model performs poorly as the null model, $\tilde{R}^2 = 0$
  - If the full model performs perfectly, $ln\mathcal{L} = 0$, then $\tilde{R}^2 = 1$

3.5 Goodness of fit #3: Akaike's information criterion (AIC) & Bayesian information criterion (BIC)

- The **best** model is the one with smallest AIC/BIC
- The better fit, created by making a model more complex by requiring more parameters, must be considered in light of the **penalty** imposed by adding more parameters
- Compare two models, delta AIC/BIC (<2, support the candidate model; >10, against the candidate model)
  - AIC

$$AIC = -2ln\mathcal{L} + 2k \tag{40}$$

  - BIC

$$BIC = -2ln\mathcal{L} + kln(N) \tag{41}$$

4 Additional section

4.1 Additional reading

- Scott-Clayton and Schudde (2016), Performance standards in need-based student aid. *NBER Working Paper No.22713.*

- Satisfactory Academic Progress (SAP) requirements failure has heterogeneous effects in the short term, with negative impacts on persistence but positive effects on grades for students who remain enrolled.
- After three years, the negative effects appear to dominate

- Bulman et al. (2016), Parental resources and college attendance: Evidence from lottery wins. *NBER Working Paper No.22679*
  - Lottery data!
  - Modest, increasing, and only weakly concave effects of resources: wins less than $100,000 have little influence on college-going (i.e., effects greater than 0.3 percentage point can be ruled out) while very large wins that exceed the cost of college imply a high upper bound (e.g., wins over $1,000,000 increase attendance by 10 percentage points).
  - The effects are smaller among low-SES households.

4.2 `Stata` tip #6: What do you mean by saying "holding all else equal?

```
(1) D. C. Hoaglin
"Regressions are commonly misinterpreted"
http://www.stata-journal.com/article.html?article=st0419

(2) J. W. Hardin
"Regressions are commonly misinterpreted: Comments on the article"
http://www.stata-journal.com/article.html?article=st0420

(3) J. S. Long and D. M. Drukker
"Regressions are commonly misinterpreted: Comments on the article"
http://www.stata-journal.com/article.html?article=st0421

(4) D. C. Hoaglin
"Regressions are commonly misinterpreted: A rejoinder"
http://www.stata-journal.com/article.html?article=st0422
```

EDUC 799 LAB #7
Understanding and Estimating Causal Effects: An Application of Non-Linear Models

Xiaoyang Ye
University of Michigan

October 19, 2016

OBJECTIVES OF LAB 7

1. Checklist
2. Potential outcomes and treatment effects
3. `teffects`

1 Checklist

- Questions about lecture, lab and office hours?
- Work on your final project (question, literature, data)
- Quiz

General comments on Assignment #3

- Choose a title to present your key information (and introduction/conclusion)
- Significance can be different between coefficients and AMEs
- Be careful

2 Potential Outcome and Treatment Effects

2.1 Neyman-Holland-Rubin Causal Model

- **Causality** is ambiguous: "what is the effect of college education on earnings?" or "what is the effect of race (or gender) on earnings?"
- Definition in the *Merriam-Webster Dictionary:*

> CAUSE/TREATMENT - Something or someone that produces an effect, result, or condition
>
> EFFECT - An event, condition, or state of affairs that is produced by a cause

- Holland (1986):

> ...the potential (regardless of whether it can be achieved in practice or not) for exposing or not exposing each unit to the action of a cause.

  - If we agree with Holland (1986), then it is clear that many questions are not causal.

Potential outcomes

- Causality is defined by potential outcomes, not by realized (observed) outcomes
- *Counterfactual*

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0) \tag{42}$$

  - T_i: a dummy variable indicating whether individual i receives treatment ($T_i = 1$) or not ($T_i = 0$);
  - $Y_i(1)$: the outcome of individual i if she receives treatment;
  - $Y_i(0)$: the outcome of individual i if she does not receive treatment.
- A valid causality question must involve well-defined causes (treatments, manipulations), and the counterfactuals should be unambiguously defined.
- For any vaguely-defined problem, It is always good to consider whether there is a treatment/manipulation/thought experiment that could produce meaningful counterfactuals.

Fundamental problem of causal inference

- *Individual treatment effect*

$$\tau_i = Y_i(1) - Y_i(0) \tag{43}$$

- We can only observe one of the two potential outcomes
  - *Missing data problem*: Any statistical method dealing with treatment effects necessarily imputes the counterfactual part of the data.
  - Extensions: **Fisher's sharp null hypothesis** and **Permutation Test**

Average treatment effects

- The **average treatment effect (ATE)** is defined as

$$\tau_{\texttt{ATE}} = \mathbb{E}[Y_i(1) - Y_i(0)]. \tag{44}$$

  - Extensions: Quantile treatment effects

- The **average treatment effect on the treated (ATT)** is defined as

$$\tau_{\texttt{ATT}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 1]. \tag{45}$$

- The **average treatment effect on the untreated (ATUT)** is defined as

$$\tau_{\texttt{ATU}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid T_i = 0]. \tag{46}$$

2.2 Treatment Assignment Mechanism

- Estimation of causal effects starts with the assignment mechanism.
- In general, the **assignment mechanism** could be denoted by the following mapping:

$$(i, \mathbf{X}_i, \mathbf{Y}_i, u_i) \ \rightarrow T_i, \tag{47}$$

  where $\mathbf{X}_i$ includes individual characteristics, $\mathbf{Y}_i = (Y_i(0), Y_i(1))$ are potential outcomes and u_i is come orthogonal error term which allows for randomization.

- Since T_i is a binary random variable (familiar?), the assignment mechanism could be represented by the **propensity score function**:

$$p(i, \mathbf{X}_i, \mathbf{Y}_i) = \mathbf{P}\left[T_i = 1 \mid i, \mathbf{X}_i, \mathbf{Y}_i\right] \tag{48}$$

- Here are come important constraints about the assignment mechanism:
  - **Individualistic Assignment**:

$$p(i, \mathbf{X}_i, \mathbf{Y}_i) = p(\mathbf{X}_i, \mathbf{Y}_i) = \mathbf{P}\left[T_i = 1 \mid \mathbf{X}_i, \mathbf{Y}_i\right] \tag{49}$$

  - **Probabilistic Assignment** (overlapping):

$$0 < p(i, \mathbf{X}_i, \mathbf{Y}_i) < 1 \tag{50}$$

  - **Unconfoundedness**:

$$p(i, \mathbf{X}_i, \mathbf{Y}_i) = p(\mathbf{X}_i) = \mathbf{P}\left[T_i = 1 \mid \mathbf{X}_i\right]. \tag{51}$$

    * Conditional on covariates (both observed and unobserved), the assignment mechanism is similar to a randomized experiment

* **Omitted variable bias**: In many cases the covariates $\mathbf{X}_i$ are not rich enough, meaning that the assignment mechanism could still be related to the potential outcomes, so that the estimated treatment effects could be biased

2.3 Selection Bias

- **Prima facie causal effect** by Holland (1986):

$$
\begin{aligned}
&\mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0] \\
&= \underbrace{\mathbf{E}[Y_i(1)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 1]}_{\tau_{\text{ATT}}} + \underbrace{\mathbf{E}[Y_i(0)|T_i = 1] - \mathbf{E}[Y_i(0)|T_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

  - Without further assumptions, however, the selection bias could be arbitrary large/small which makes the above quantity uninformative about the average treatment effect

- Rubin causal model is a special case of **switching regression / general Roy model**:

$$
\begin{aligned}
\text{Potential Outcomes:} \quad & Y_i(0) = \mathbf{X}_i\beta(0) + u_i(0) \\
& Y_i(1) = \mathbf{X}_i\beta(1) + u_i(1) \\
\text{Selection/Assignment Mechanism:} \quad & \mathbf{1}_{\{T_i=1\}} = F(\mathbf{X}_i\gamma) + \epsilon_i
\end{aligned} \tag{52}
$$

- The identification is:

$$\mathbf{X}_i \perp (u_i(0), u_i(1), \epsilon_i).$$

- The unconfoundedness assumption is simply:

$$\text{Unconfounded Assignment:} \quad (u_i(0), u_i(1)) \perp \epsilon_i, \tag{53}$$

  which has the intuitive meaning that after controlling the covariates, the errors in determining the potential outcomes are orthogonal to the errors in the selection equation/assignment mechanism

- If the above does not hold, then the potential outcomes will be correlated with the selection equation/assignment mechanism even after partialling out the covariates, hence selection bias will affect estimating the average treatment effect

2.4 Causal Research Designs

- This section summarizes the excellent chapter in DiNardo and Lee (2011). You may want to read more in:
  - Heckman and Vytlacil (2007a), *Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation*
  - Heckman and Vytlacil (2007b), *Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments*

- Abbring and Heckman (2007), *Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation*
- Angrist and Pischke (2008), *Most Harmless Econometrics*
- Murnane and Willett (2010), *Methods Matter*

1. By knowledge of **Assignment Mechanism**
   - Random assignment (RCT)
     - Gold standard
   - Regression discontinuity (RD)
     - A close "cousin" of RCT
2. By **Self-Selection**
   - Difference-in-differences (DID)
     - Influence of "other factors" fixed
   - Selection on unobservables and instrumental variables (IV)
     - Conditional on covariates, instrument "as good as randomly assigned"
     - Instrument is uncorrelated with outcomes
     - Another structural approach: Heckman selection model
   - Selection on observables and matching (Matching)
     - Conditional on covariates, treatment "as good as randomly assigned"
     - Including Blinder/Oaxaca decomposition, matching, propensity score matching, re-weighting

- If you are interested in knowing how these research designs have been used to study the core question in the economics of education: **the causal effects of schooling on wages**, please read Chapter 6 in Angrist and Pischke (2014)

3 `teffects`: "Treatment effects" using propensity score

- Read more in http://www.stata.com/manuals14/te.pdf
- "Treatment effects" using matching (impute missing potential outcomes):
  1. Blinder/Oaxaca Decomposition
     - Take each individual in the treated sample, impute the missing outcomes $Y|T = 0, X = x$ given the individual's characteristics X
     - Regression Adjustment estimators run separate regressions for each group, then calculate means of predicted outcomes using all the data and the estimated coefficients to impute potential outcome means
     - Concern: the regression used to predict $\hat{Y}$ may be a bad approximation of the true conditional expectation

2. Matching
   - To simply use the sample mean of Y_i for all individuals in the non-treated sample that have exactly the same value for X as the individual i
   - Exact matching on discrete covariates is the same as regression adjustment
   - X can be multi-dimensional
   - nearest matching, non-parametric matching (kernel, local polynomial)
3. Propensity score matching
   - To match on the propensity score, rather than on the observed X
4. Re-weighting
   - To "re-weight" the control sample so that the re-weighted distribution of X matches that in the treated population
   - One can re-weight the data and examine other aspects of the distribution, such as quantiles, variances, etc. by computing the desired statistic with the appropriate weight

- `teffects` estimators:

```
teffects ra        Regression adjustment
teffects ipw       Inverse-probability weighting
teffects aipw      Augmented inverse-probability weighting
teffects ipwra     Inverse-probability-weighted regression adjustment
teffects nnmatch   Nearest-neighbor matching
teffects psmatch   Propensity-score matching
```

4 Additional section

4.1 Understanding OVB in regressions

4.1.1 What is omitted variable bias (OVB)?

- **One of the most common and vexing problems in regression analysis**
  - It occurs when we omit an important variable in the model
  - The variable is correlated with both the dependent *and* one or more independent variables
- **Example: Returns to schooling**
  - The basic OLS model (excluding ability measures):

$$Wage_i = \beta_0 + \beta_1 SchoolingYears_i + \varepsilon_i \tag{54}$$

- Estimated $\hat{\beta_1} = 0.15$ using a random sample of the population

- Omitted variable: ability
- It is an confounding factor as ability is (positively) correlated with one's years of schooling, as those with greater ability seek more education
- It is also related with wages, as those with greater ability earn more

- The correct model (including ability measures):

$$Wage_i = \alpha_0 + \alpha_1 SchoolingYears_i + \alpha_2 Ability_i + \mu_i \tag{55}$$

- Estimated $\hat{\alpha_1} = 0.10$ using the same random sample of the population

4.1.2 Upward or downward OVB?

- **Why $\hat{\beta_1} > \hat{\alpha_1}$, which suggests upward bias?**

$$\mathbf{E}(\hat{\beta_1}) = \alpha_1 + \alpha_2 [\frac{\sum SchoolingYears_i * Ability_i}{\sum SchoolYears_i{}^2}] \tag{56}$$

- **Determine the direction of bias**
  (1) Direction of a: correlation between $Ability$ and $Wage$
  (2) Direction of $sumSchoolingYears_i * Ability_i$: correlation between $Ability$ and $SchoolingYears$
- **Predict potential OVB?**
  - Many times you cannot predetermine the direction of OVB
    * Downward bias in returns to education: Overeducated
  - Use theory/intuition/context... to guide your hypothesis

4.1.3 Possible solutions to OVB?

- **Observed (to researchers) omitted variables**
  - Include it in the regression
- **Unobserved omitted variables**
  - **Selection bias** in the causal inference literature
  - Experimental and quasi-experimental methods

4.2 `Stata` tip #7: Linear and non-linear model choice matters!

Here is a great `Stata` post that uses simulation to show that LPM could produce very different marginal effects than logit/probit models: regress, probit, or logit? by Enrique Pinzon.

- Simulation design

```
clear all
cls
local N = 10000    // Sample size
local R = 5000     // Simulation repetitions
local L = 20000000 // Sample size for approximate true value calculations
local S = 10       // Repetitions for approximate true value calculations
set seed 111

program define mkdata
    syntax, [n(integer 1000)]
    clear
    quietly set obs `n'

    // 1. Generating data from probit, logit, and misspecified
    generate x1    = rchi2(2)-2
    generate x2    = rbeta(4,2)>.2
    generate u     = runiform()
    generate e     = ln(u) -ln(1-u)
    generate ep    = rnormal()
    generate xb    = .5*(1 - x1 + x2)
    generate y     =  xb + e > 0
    generate yp    = xb + ep > 0

    // 2. Computing probit & logit marginal and treatment effects
    generate m1   = exp(xb)*(-.5)/(1+exp(xb))^2
    generate m2   = exp(1 -.5*x1)/(1+ exp(1 -.5*x1 )) - ///
    exp(.5 -.5*x1)/(1+ exp(.5 -.5*x1 ))
    generate m1p  = normalden(xb)*(-.5)
    generate m2p  = normal(1 -.5*x1 ) - normal(.5 -.5*x1)

    // 3. Computing marginal and treatment effects at means
    quietly mean x1 x2
    matrix A        = r(table)
    scalar a        = .5 -.5*A[1,1] + .5*A[1,2]
    scalar b1       =  1 -.5*A[1,1]
    scalar b0       = .5 -.5*A[1,1]
    generate mean1  = exp(a)*(-.5)/(1+exp(a))^2
    generate mean2  = exp(b1)/(1+ exp(b1)) - exp(b0)/(1+ exp(b0))
    generate mean1p = normalden(a)*(-.5)
    generate mean2p = normal(b1) - normal(b0)
end
```

- Simulation codes (Logit, Average Marginal Effects)

```
mkdata, n(`L')
local values "m1 m2 mean1 mean2 m1p m2p mean1p mean2p"
local means  "mx1 mx2 meanx1 meanx2 mx1p mx2p meanx1p meanx2p"
local n : word count `values'
forvalues i= 1/`n' {
    local a: word `i' of `values'
    local b: word `i' of `means'
    sum `a', meanonly
    local `b' = r(mean)
}

postfile lp y1l y1l_r y1lp y1lp_r y2l y2l_r y2lp y2lp_r ///
using simslp, replace

forvalues i=1/`R' {
    quietly {
        mkdata, n(`N')

        logit  y x1 i.x2, vce(robust)
        margins, dydx(*) post  vce(unconditional)
        local y1l = _b[x1]
        test _b[x1] = `mx1'
        local y1l_r   = (r(p)<.05)
        local y2l = _b[1.x2]
        test _b[1.x2] = `mx2'
        local y2l_r   = (r(p)<.05)

        regress  y x1 i.x2, vce(robust)
        margins, dydx(*) post  vce(unconditional)
        local y1lp = _b[x1]
        test _b[x1] = `mx1'
        local y1lp_r   = (r(p)<.05)
        local y2lp = _b[1.x2]
        test _b[1.x2] = `mx2'
        local y2lp_r   = (r(p)<.05)

        post lp (`y1l') (`y1l_r') (`y1lp') (`y1lp_r') ///
        (`y2l') (`y2l_r') (`y2lp') (`y2lp_r')
    }
    if (`i'/50) == int(`i'/50) {
        di ".                 `i'"
    }
    else {
        di _c "."
    }
}
postclose lp
use simslp, clear
sum
```

4.3 `Stata` tip #8: Heteroskedasticity and robust standard errors in MLE

This tip comes from Two faces of misspecification in maximum likelihood: Heteroskedasticity and robust standard errors by Enrique Pinzon.

```
For a nonlinear model with heteroskedasticity, a maximum likelihood estimator
gives misleading inference and inconsistent marginal effect estimates unless
you model the variance.

Using a robust estimate of the variance-covariance matrix will not help you
obtain correct inference.

-- use "hetprobit"
```

4.4 `Stata` tip #9: Generate random numbers in Stata

```
clear
set obs 10000

/* set sed to make numbers reproducible */
set seed 0926

/* random number betwee 0 and 1 */
gen u1 = runiform()
// mean of a (0,1)-uniform is 0.5, sd is (1/12)^(0.5)=0.289
sum u1

/* random number between 1 and 2 */
gen u2 = runiform(1, 2)
// mean is 1.5, sd is 0.289
sum u2

/* random integers  between 0 and 100 */
gen u3 = runiformint(0, 100)
// mean is 50, sd is ((101^2-1)/12)^(0.5)=29.155

/* normal distribution (0,1) */
gen x = rnormal()
// other distribution: rbeta, rweibull
```

EDUC 799 LAB #8
Modeling multi-valued choices: Ordered

Xiaoyang Ye
University of Michigan

October 26, 2016

OBJECTIVES OF LAB 8

1. Checklist
2. Introduction of NCES data sets
3. Ordinal regression

1 Checklist

- Questions about lecture, lab and office hours?
- **Close-to-end course evaluation** and paper review sign-up
- Work on your final project (from theories to empirics)
- Assignment 5 is due next Tuesday; Revised literature review and methodology write-up is due in two weeks

General comments on Assignment #4 and final projects

- Very interesting (causal) questions!
- When do we need to test model fit?
  - Description
  - Causal inference (*ex post*)
  - Prediction (*ex ante*); much more complicated when including general equilibrium effects (structural econometric models)
- Show your audience that you are not only **methodological-trained** (Stata runner?) but also **empirical-minded**
- Literature review: to find open questions with limited evidence
  - Why/What should readers care about your paper?
  - Most (if not all) empirical questions are policy-relevant

2 A brief introduction to NCES data sets

- National Center for Education Statistics: http://nces.ed.gov/
  - The primary federal entity for collecting and analyzing data related to education in the U.S. and other nations
- Assessments, surveys, and administrative data
- Reports, e.g.:
  - The Condition of Education 2016
  - Revenues and Expenditures for Public Elementary and Secondary Education: School Year 2013-14 (Fiscal Year 2014) (NCES 2016-301)

2.1 Assessment data

- **NAEP** (National Assessment of Educational Progress): http://nces.ed.gov/nationsreportcard/
  - The largest nationally representative and continuing assessment of what America's students know and can do in various subject areas
  - Results on subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., all fourth-graders) and groups within those populations (e.g., female students, Hispanic students)
    * Include many subjects, including mathematics, reading, science, writing, the arts, civics, economics, geography, and U.S. history
    * Each subject is assessed at grades 4, 8, and 12
    * Four of these subjects (mathematics, reading, science, and writing) are reported at the state level
    * At the state level, assessment is in public schools only
    * State assessments began in 1990
- International Assessment: http://nces.ed.gov/surveys/international/
  - **TIMESS** (Trends in International Mathematics and Science Study)
  - **PISA** (Program for International Student Assessment)

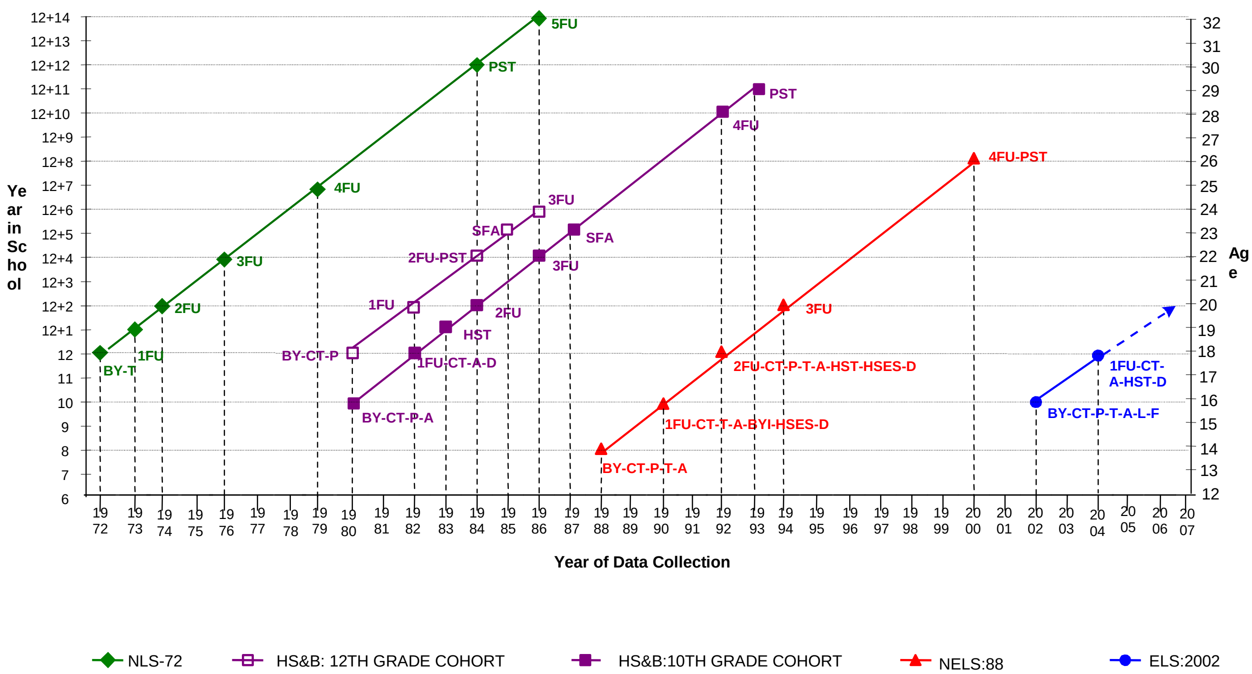
2.2 Survey data

- Early Childhood
  - ECLS (Early Childhood Longitudinal Study)
    * **ECLS-B** (2001-02; follow-up in 2003-04 2-yr; 2005-06 4-yr)
    * **ECLS-K** (1998-99; follow-up in 1999-2000 1st grade, 2002 3rd grade, 2004 5th grade, 2007 8th grade)
    * **ECLS-K: 2011** (2010-2011; follow-up in 2011-12 1st grade, 2012-13 2nd grade, 2014 3rd grade, 2015 4th grade, 2016 5th grade)

– **NHES** (National Household Education Survey)

  ∗ Surveys: 1991, 1993, 1995, 1996, 1999, 2001, 2003, 2005, 2007

- K-12 (main longitudinal surveys)
  – **NLS-72** (National Longitudinal Study of the High School Class of 1972, senior)
  – **HS&B** (High School and Beyond, 1980 senior and sophomore cohorts)
  – **NELS:88** (National Education Longitudinal Study of 1988, 8th grade)
  – **ELS:2002** (Education Longitudinal Study of 2002, sophomore)
  – **HSLS:09** (High School Longitudinal Study of 2009, 9th grade)
  – **MGLS:2017** (Middle Grades Longitudinal Study 2017-18)



Notes: "Figure 1.2. - Longitudinal design for the NCES high school cohorts" in NELS 88 Base-Year to Fourth Follow-up Data File User's Manual.

- Postsecondary
  – **NPSAS** (National Postsecondary Student Aid Study, 87, 90, 93, 96, 00, 04, 08, 12)
  – **BPS** (Beginning Postsecondary Student, 90/94, 96/01, 04/09, 12/17)
  – **B&B** (Baccalaureate and Beyond Longitudinal Study, 93/03, 00/01, 08/12)
- Other survey data
  – **SER** (State Education Reforms)
  – **PSS** (Private School Universe Survey, similar to CCD, bi-annually since 1989-90)

- **SASS** (School and Staffing Survey, 7 times between 1987-2001, renamed to **NTPS**, National Teacher and Principal Survey since 2016)
- **SABS** (School Attendance Boundary Survey)
- **HST** (High School Transcript Studies, 82 (HS&B), 87, 90, 92 (NELS), 94, 98, 00, 04 (ELS), 09)
- **CSS** (Crime & Safety Surveys, 89, 95, 99, 01, 03, 05, 07, 09, 11, 13)

2.3 Administrative data

- **CCD** (Common Core of Data, data on U.S. public schools and districts)
- **IPEDS** (Integrated Postsecondary Education Data System, data on U.S. colleges)

2.4 An example: NELS 1988

- Sample
  - Nationally representative sample of eighth-graders were first surveyed in the spring of 1988
  - A sample of these respondents were then resurveyed through four follow-ups in 1990, 1992, 1994, and 2000

Wave	Year	Cohort grade	Variable name
Base	1988	8th grade	by___(e.g., byses)
First follow-up	1990	10th grade	f1___(e.g., f1ses)
Second follow-up	1992	12th grade	f2___(e.g. f2ses)

- Questionnaires: http://nces.ed.gov/surveys/nels88/questionnaires.asp
- User Manual http://nces.ed.gov/pubs2002/2002323.pdf
- Variables
  - "How important is or was each of the following in choosing a school you would like to attend'?'
  - Question 59 in Second Follow-up Student Survey
  - College expenses: `f2s59a`

3 Ordinal regression

3.1 The model

- Extensions of binary response model: multi-valued outcomes
  - Ordered, with even gaps
  - Ordered, with uneven (and unknown) gaps - Ordinal regression (ordered logit/probit)
  - Unordered, with case-specific covariates - Multinomial logit/probit
  - Unordered, with alternative-specific covariates
- Ordered vs. Multinomial
  - Ordered: A binary model with a series of thresholds/cut-points (latent intervals)
    - Parallel assumption
  - Multinomial: A series of binary models

3.2 Estimation

3.2.1 Binomial distribution

- Logit model ($Y = 0, 1$)

$$\mathbf{P}(Y_i = 1) = \mathbf{F}(X\beta) = \frac{e^{a+X\beta}}{1+e^{a+X\beta}} = \frac{1}{1+e^{(-a-X\beta)}} \tag{57}$$

- Ordered logit model ($Y = 1, 2, 3, ..., M$)

$$\mathbf{P}(Y_i > j) = \mathbf{F}(X\beta) = \frac{e^{\alpha_j+X\beta}}{1+e^{\alpha_j+X\beta}}, j = 1, 2, ..., M-1 \tag{58}$$

  - Latent variable

$$Y_i = 1 \quad \iff \tau_i \leq Y{*}_i \leq \tau_{i+1} \tag{59}$$

  - Specific estimation models

$$\mathbf{P}(Y_i = 1) = \frac{1}{1+e^{-\tau_1+X\beta}} \tag{60}$$

$$\mathbf{P}(Y_i = j) = \frac{1}{1+e^{-\tau_j+X\beta}} - \frac{1}{1+e^{\tau_{j-1}+X\beta}}, j = 2, ..., M-1 \tag{61}$$

$$\mathbf{P}(Y_i = M) = 1 - \frac{1}{1+e^{-\tau_{M-1}+X\beta}} \tag{62}$$

- Generalized ordered logit model ($Y = 1, 2, 3, ..., M$)

$$\mathbf{P}(Y_i > j) = \mathbf{F}(X\beta) = \frac{e^{\alpha_j+X\beta_j}}{1+e^{\alpha_j+X\beta_j}}, j = 1, 2, ..., M-1 \tag{63}$$

  - Logit and ordered logit models are special cases of the generalized model

- The generalized model allows **some** of the β coefficients to be the same, while others can differ (partial proportional/parallel odds)
- Note: Multinomial models assume the probability distribution is given by **multinomial** distribution

3.2.2 Parallel assumption: Constrained β

- When the "Parallel Assumption" is violated, try all the below and decide what to do based on the results

  (1) Do nothing if the potential bias is minimal (a fairly common practice)

  (2) Use multinomial models with efficiency loss
    - Because you don't use the order information and you may estimate more parameters than is necessary
    - Increased risks of getting insignificant results
    - Estimates are unbiased (you may get biased estimates using ordered model to estimate unordered outcomes)

  (3) Dichotomize the outcome and use binary models

  (4) Use an ordered model without Parallel Assumption: `gologit2`, `oglm`

  (5) Use Stereotype logistic model (`slogit`)
    - If you are unsure of the relevance of the ordering of the outcome, or the order is indistinguishable
    - Estimate fewer parameters than multinomial models

3.3 Prediction and marginal effects: Using the *Challenger* example

- The example is from http://www3.nd.edu/ rwilliam/statafiles/shuttle2.dta, clear

- The space shuttle ***Challenger*** disaster of January 28, 1986, might have been averted had NASA officials heeded the warning signs.

- Data
  - 25 flights including Challenger; 23 are used for estimation, Flight 4 (STS-4 in 1982) - "O-ring condition unknown; rocket casing lost at sea."

- Variables
  - **Distress**: Outcome - number of "thermal distress incidents" in which hot gas damaged the joint seals of a flight's booster rockets -> led to the Challenger disaster
  - **Temp**: The calculated joint temperature at launch time. Temperature depends largely on weather. Colder temperatures cause the rubber o-rings sealing the booster rocket joints to become less flexible and hence more likely to have problems
  - **Date**: Date, measured in days elapsed since January 1, 1960 (an arbitrary starting point). The rationale for this variable is that undesirable changes in the shuttle program and aging hardware may have caused launches to become more risky across time

```
use http://www3.nd.edu/~rwilliam/statafiles/shuttle2.dta, clear
(First 25 space shuttle flights)

###outcome
tab distress

Thermal      |
distress     |
incidents    |      Freq.     Percent        Cum.
-------------+-----------------------------------
None         |          9       39.13       39.13
1 or 2       |          6       26.09       65.22
3 plus       |          8       34.78      100.00
-------------+-----------------------------------
Total        |         23      100.00

###create date
cap drop date1
gen date1 = mdy(month, day, year)

###estimate ologit model
ologit distress date temp

Iteration 0:   log likelihood = -24.955257
Iteration 1:   log likelihood = -18.871284
Iteration 2:   log likelihood =  -18.79755
Iteration 3:   log likelihood =  -18.79706
Iteration 4:   log likelihood =  -18.79706

Ordered logistic regression                     Number of obs     =         23
                                                LR chi2(2)        =      12.32
                                                Prob > chi2       =     0.0021
Log likelihood    =  -18.79706                  Pseudo R2         =     0.2468

------------------------------------------------------------------------------
distress |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
date     |    .003286   .0012662     2.60   0.009     .0008043    .0057677
temp     |  -.1733752   .0834475    -2.08   0.038    -.3369293   -.0098212
---------+--------------------------------------------------------------------
/cut1    |   16.42813   9.554822                     -2.298978    35.15524
/cut2    |   18.12227   9.722302                      -.933092    37.17763
------------------------------------------------------------------------------

### Pseudo R2
di e(chi2)/(-2*e(ll_0))
.2467695

### Predicted probability of flight 13
list date temp if flight==13

     +---------------+
     | date    temp |
     |---------------|
13.  | 9044      78 |
     +---------------+

cap drop xb
gen xb = _b[date] * 9044 + _b[temp] * 78
sum xb  /* 16.19526 */

di 1/(1+exp(r(mean)-_b[/cut1]))
```

```
.55795573

di 1/(1+exp(r(mean)-_b[/cut2])) - 1/(1+exp(r(mean)-_b[/cut1]))
.31496255

di 1/(1+exp(r(mean)-_b[/cut2])) - .55795573
.31496255

di 1- 1/(1+exp(r(mean)-_b[/cut2]))
.12708172

### Predicted probability of Challenger (out of sample)
list date temp if flight==25

     +---------------+
     | date    temp |
     |---------------|
25.  | 9524      31 |
     +---------------+

cap drop xb
gen xb = _b[date] * 9524 + _b[temp] * 31
sum xb  /* 25.92117 */

di 1/(1+exp(r(mean)-_b[/cut1]))
.00007537

di 1/(1+exp(r(mean)-_b[/cut2])) - 1/(1+exp(r(mean)-_b[/cut1]))
.00033465

di 1- 1/(1+exp(r(mean)-_b[/cut2]))
.99958998

### Prediction using "predict" command
predict xb, xb         /* xb */
predict y1 y2 y3, pr   /* probability */


###Sample selection
 sum temp, d

          Joint temperature, degrees F
-------------------------------------------------------------
 Percentiles      Smallest
 1%           31             31
 5%           53             53
10%           57             57       Obs                  25
25%           67             58       Sum of Wgt.          25

50%           70                      Mean              68.44
                    Largest           Std. Dev.      10.52806
75%           75             78
90%           79             79       Variance         110.84
95%           80             80       Skewness      -1.888471
99%           81             81       Kurtosis       7.511047

* incorrect prediction
twoway  (lfit distress temp if distress>1 & flight<24) ///
(scatter distress temp if distress>1  & flight<24)
```

- Pseudo R2 (McFadden R2): $ModelLRChi2/(-2 * L_0) = 12.32/(-2 * -24.955257) = 49.91$

- **Predicted probability**: Flight #13 - temp=78, date=9044
  - $X\hat{\beta} = 16.19526$

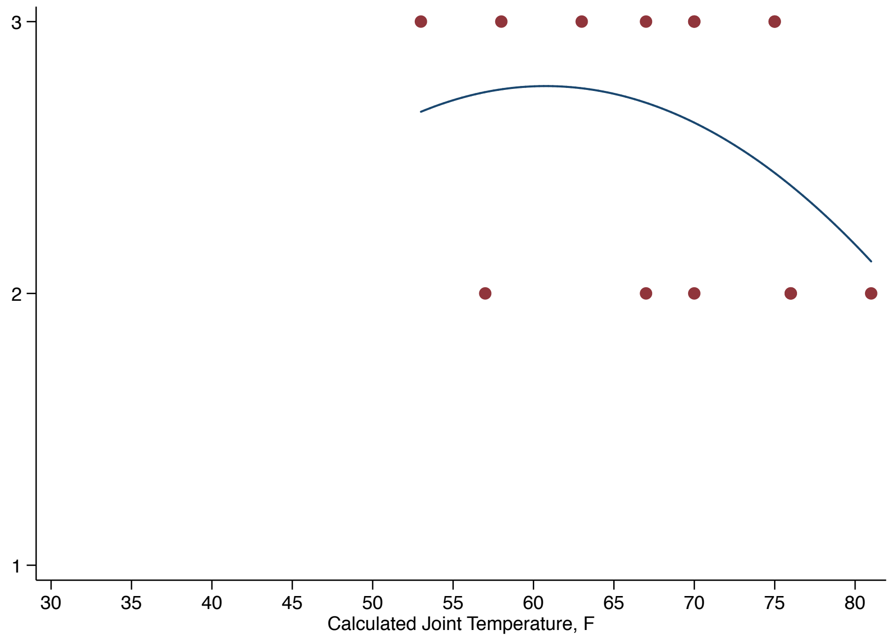
$$\mathbf{P}(Y_i = 1) = \frac{1}{1 + e^{-16.42813+16.19526}} = .55795573 \quad (64)$$
$$\mathbf{P}(Y_i = 2) = \frac{1}{1 + e^{-18.12227+16.19526}} - \frac{1}{1 + e^{-16.42813+16.19526}} = .31496255 \quad (65)$$
$$\mathbf{P}(Y_i = 3) = 1 - \frac{1}{1 + e^{-18.12227+16.19526}} = .12708172 \quad (66)$$

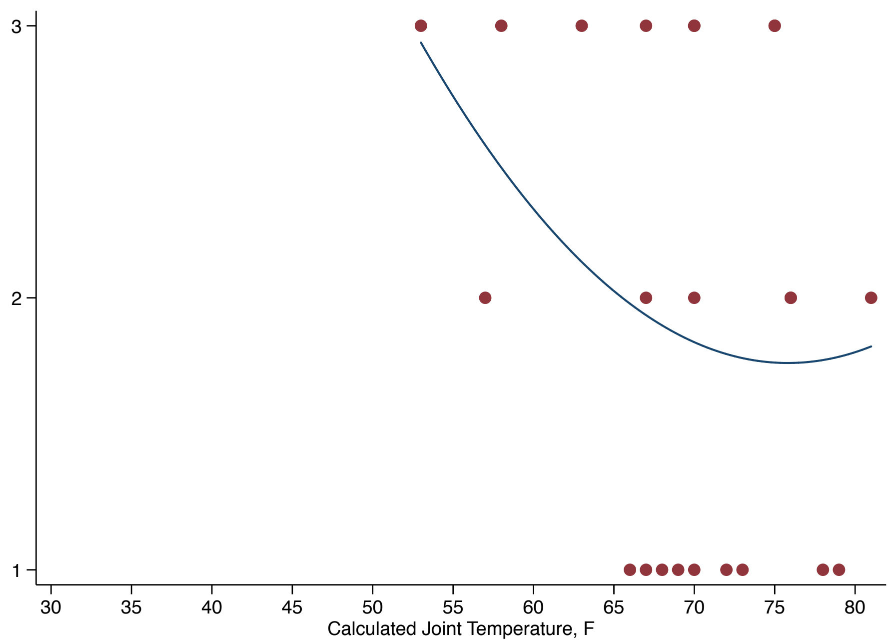
  - For Flight 13, which occurred more than a year earlier than Challenger under much warmer conditions, the most likely outcome was that there would be no damage to the booster joints
    - In fact, Flight 13 did not have any problems
  - For Flight 13, $X\beta$ (16.20) is our best guess for the value of Y^* using the model and data. This value places Flight 13 in the $Y = 1$ threshold ($Pr(Y = 1)$ has the largest predicted probability)
  - But, because of the random disturbance term (other unmeasured factors), there is at least some chance that Y^* is larger than 16.20, e.g.,
    - the Y^* value for Flight 13 could actually be 17, in which case $Y = 2$
    - or it might be 18.5, then $Y = 3$
- **Out-of-sample prediction, Challenger**
  - If NASA had used this empirical study to guide their work
  - Date was 9524, temperature was 31 (so low!)

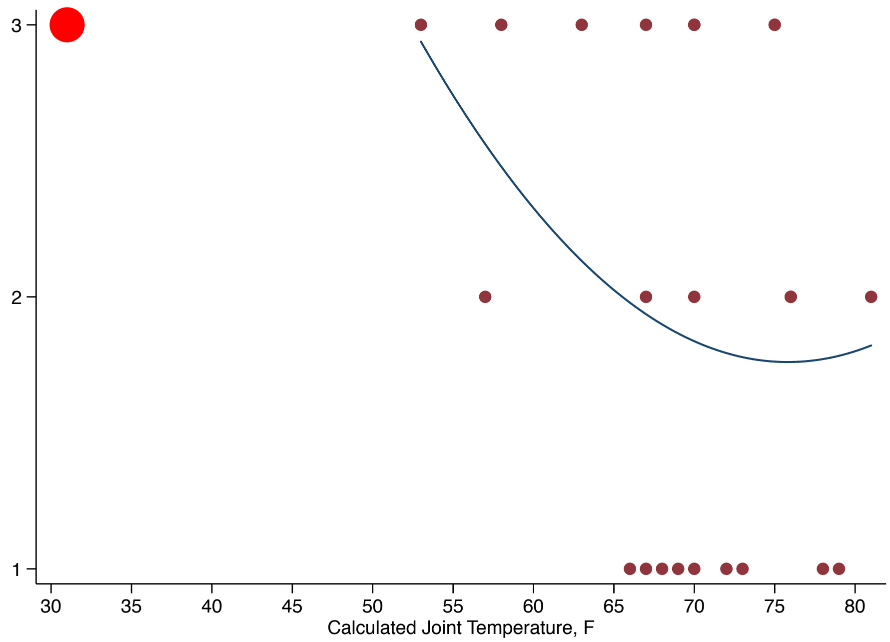
$$\mathbf{P}(Y_i = 1) = .00007537 \quad (67)$$
$$\mathbf{P}(Y_i = 2) = .00033465 \quad (68)$$
$$\mathbf{P}(Y_i = 3) = .99958998 \quad (69)$$

  - Based on the experience from the previous 23 flights, there was virtually no chance that Challenger would experience no damage to its joint seals
  - Indeed, it was a virtual certainty that Challenger would experience 3 or more damage incidents
- An additional note on sample selection: "Nobody knew what would happen to the O-rings on a day where the temperature was 15 degrees colder than that of any previous launch."
  - This note is from Johnson and Gluck (2016), *Everydata: The Misinformation Hidden in the Little Data You Consume Every Day*. Routledge. (Read sample chapters on Canvas "Lab 8" folder.)
  - "The team recognized that they didn't have data below 53 degrees, and decided to look at all cases where there had been signs of O-ring distress, regardless of temperature."

– "But the problem, as you'll see, was that they did not study the *right* data for the question they needed to answer. In this case, they should have looked at *all* of the data on O-ring performance - not just cases where there were signs of distress."







- **Another additional note: Marginal effects**
  - Estimate and interpret in the same way as for logit/probit models

4 Additional section

4.1 Additional reading

- Liu and Xie (2016), "Why do Asian Americans academically outperform Whites? - The cultural explanation revisited." *Social Science Research*, 58, 210-226.
  - An interactive approach to examining the role of culture and SES in explaining Asian Americans' achievement
  - The cultural orientation of Asian American families is different from that of white American families in ways that mediate the effects of family SES on children's academic achievement
    (1) SES's positive effects on achievement are stronger among white students than among Asian-Americans
    (2) The association between a family's SES and behaviors and attitudes is weaker among Asian-Americans than among Whites
    (3) A fraction of the Asian-White achievement gap can be accounted for by ethnic differences in behaviors and attitudes, particularly ethnic differences in family SES's effects on behaviors and attitudes
  - ELS 2002 data
  - A related study: Hsin and Xie (2014)

4.2 `Stata` tip #10: With or without reference (in dummy indicators)

```
This is Stata tip 106 by Maarten L. Buis
Please see http://maartenbuis.nl/publications/ref_cat.pdf
```

EDUC 799 LAB #9
Modeling multi-valued choices: Unordered and case-specific

Xiaoyang Ye
University of Michigan

November 2, 2016

OBJECTIVES OF LAB 9

1. Checklist
2. Multinomial logit model
3. Many `Stata` tips

1 Checklist

- Questions about lecture, lab and office hours?
- Close-to-end course evaluation
- Work on your final project (**empirical analysis plan done?**)
- Revised literature review and methodology write-up is due this Friday 6pm
- Assignment 6 will be distributed this Friday evening
- Steve's Indiana paper .do file is now available on Canvas

Final projects

- Start from the simple models (OLS/binary logit or probit)
- Focus on the central question
  - What's your most important outcome?
  - What's the key explanatory variable?
- If you are going to answer a causal question, try to causally address it
  - Prior literature?
  - Theoretical/conceptual framework?
  - Empirical analysis and limitation?
- Present your results in tables and figures

NO LAB ON NOVEMBER 9

- Association for the Study of Higher Education 2016 Conference (Nov 9-12)
- Have fun if you are going to Columbus, OH
- Represent UMich, CSHPE, and Educ 799!
- Association for Education Finance and Policy 2017 Conference CFP is due on November 11

2 Multinomial logit/probit models

2.1 The model

- Extensions of binary response model: multi-valued outcomes
  - Ordered, with even gaps
  - Ordered, with uneven (and unknown) gaps - Ordinal regression (ordered logit/probit)
  - Unordered, with case-specific covariates - Multinomial logit/probit
  - Unordered, with alternative-specific covariates
- Ordered vs. Multinomial
  - Ordered: A binary model with a series of thresholds/cut-points (latent intervals)
    * Parallel assumption
  - Multinomial: A series of binary models
    * The estimates from the binary model are close to those from the multinomial models, but not exactly the same
    * This is because a series of binary logits fit with `logit` does not impose the constraints among coefficients that are implicit in the definition of the multinomial logit model

2.2 Estimation

2.2.1 Binomial distribution

- Logit model ($Y = 0, 1$)

$$\mathbf{P}(Y_i = 1) = \mathbf{F}(X\beta) = \frac{e^{a+X\beta}}{1+e^{a+X\beta}} = \frac{1}{1+e^{(-a-X\beta)}} \tag{70}$$

- Ordered logit model ($Y = 1, 2, 3, ..., M$)

$$\mathbf{P}(Y_i > j) = \mathbf{F}(X\beta) = \frac{e^{\alpha_j+X\beta}}{1+e^{\alpha_j+X\beta}}, j = 1, 2, ..., M-1 \tag{71}$$

- Generalized ordered logit model ($Y = 1, 2, 3, ..., M$)

$$\mathbf{P}(Y_i > j) = \mathbf{F}(X\beta) = \frac{e^{\alpha_j + X\beta_j}}{1 + e^{\alpha_j + X\beta_j}}, j = 1, 2, ..., M - 1 \tag{72}$$

2.2.2 Multinomial distribution

- The probability distribution

$$\mathbf{P}(Y_{i1} = y_{i1}, ..., Y_{iJ} = y_{iJ}) = \binom{n_i}{y_{i1}, ..., y_{iJ}} P_{i1}{}^{y_{i1}} ... P_{iJ}{}^{y_{iJ}} \tag{73}$$

  - If $J = 2$, it is binomial distribution

- Log odds

$$ln \frac{\mathbf{P}_i}{\mathbf{P}_j} = \mathbf{X_i}(\beta_i - \beta_j) \tag{74}$$

  - For reference group, β_j are 0

- Probability of each outcome I

$$\mathbf{P}(Y = i) = \frac{exp(X\beta_i)}{\sum_{k=1}^{J} exp(X\beta_k)} \tag{75}$$

  - How can we derive Equation 75 from Equation 74

- Similarly, we use maximum likelihood estimation to compute the parameters
- Similarly, we can test the model fit using `fitstat`
- Similarly, we test the coefficient hypotheses using Wald or likelihood ratio tests
  - But we can also test for combining alternatives
- Similarly, we produce marginal effects for all alternative outcomes
  - `margins`, `mchange`, `mgen`, `mtable`
- The assumption: **independence of irrelevant alternatives**
  - Adding or dropping alternatives does not affect the odds among the remaining alternatives
  - Recall the idea of OVB
  - `mlogtest`: Hausman-McFadden test (1984); Small-Hsiao test (1985)

3 Additional section

3.1 Additional reading

- Heckman, Humphries, and Veramendi (2016), "Dynamic treatment effects." *Journal of econometrics, 191*(2), 276-292.
  - This paper develops robust models for estimating and interpreting treatment effects arising from both ordered and unordered multi-stage decision problems.
  - Treatment effects are decomposed into direct effects and continuation values associated with moving to the next stage of a decision problem.
  - This paper is really **hard**, but will be popular in the coming 10-15 years.

3.2 `Stata` tip #11: Output marginal effects for multi-valued outcomes

- `outreg2` will use landscape tables for marginal effects for multi-valued outcomes (`ologit` and `mlogit`)

```
* OUTPUT MARGINAL EFFECTS FOR OLOGIT

* data
use "http://www3.nd.edu/~rwilliam/statafiles/shuttle2.dta", clear

* ologit model
ologit distress date temp
est store ologit

* marginal effect
margins, dydx(*)

* output (run margins separately for each outcome category)

// outcome=1
est restore ologit
margins, predict(outcome(1)) dydx(*) post

outreg2 using "reg1.xls", replace dec(3) ctitle("Y=1")

// outcome=2 & 3
forvalues i=2/3 {
     est restore ologit
     margins, predict(outcome(`i')) dydx(*) post

     outreg2 using "reg1.xls", append dec(3) ctitle("Y=`i'")
}
```

3.3 `Stata` tip #12: Plot bar graphs for multiple variables in %

```
* input data
clear
input var1 var2 var3 var4 var5 var6 var7 var8
1 30 70 . . . . .
2 . . 20 30 50 . .
3 . . . . . 50 50
end
// you can generate these statistics from real data

* label X variable
cap label drop label
label define label 1 "Country" 2 "Age group" 3 "Gender"
label values var1 label

* Bar graph
graph bar (asis) var2 var3 var4 var5 var6 var7 var8 , stack over(var1) ///
legend(order(1 2 3 4 5 6 7) label(1 "U.S.") label(2 "China")     ///
label(3 "Below 30") label(4 "30-50") label(5 "Above 50")     ///
label(6 "Female") label(7 "Male"))

// modify the graph based on your preference
// e.g. size, color, titles
// horizontal bar: "graph hbar ..."
```

3.4 `Stata` tip #13: Label variable using variable names

```
* add variable name B1 to the beginning of its label
foreach var of varlist B1 {

// store label
local label: variable label `var'

// add variable name to the label
label variable `var'   `"`=substr("`var'", 1, 20)': `label'"'
}
```

EDUC 799 LAB #10
Modeling multi-valued choices: Unordered and alternative-specific

Xiaoyang Ye
University of Michigan

November 16, 2016

OBJECTIVES OF LAB 10

1. Checklist
2. Conditional logit model

1 Checklist

- Questions about lecture, lab and office hours?
- Work on your final project (**hands dirty?**)
- Share your experience at ASHE?

1.1 Close-to-end class survey: Thank you!

- Cons
  1. Too much lecture in the lab
  2. Not enough time to do the exercises and ask questions
  3. Disconnection between "high-level" discussions in lecture to the "very detailed" lab discussions
  4. More hands on instead of having all commands prepared (more self-directed)
- Suggestions
  1. Enough time for both the final project and assignments
     - So you know what's going on with Assignment 7....
  2. More practices to try things out
     - Try things on you own and come to me with questions
     - More practices in the last month
- Pros
  - ......

1.2 My teaching plan in the lab

- Fish, fishing, or fish?
- Not a Stata (or R) runner, but a researcher with quantitative skills
- Sharpening the axe will not interfere with the cutting of firewood

1.3 Assignments and final projects

- Page limits
  - Many journal/conference/proposal submissions require page limits
    - **Journal of Higher Education**: Manuscripts submitted to The Journal of Higher Education (JHE) must meet the 10,000-word limit, inclusive of all text, reference list, tables, figures, and notes (excluding the abstract).
    - **ASHE proposal**: Proposals must not exceed 2,500 words (approximately four single-spaced pages). References, tables, and figures are not included in the word limit. Proposals exceeding 2,500 words will be rejected without review.
    - **Rackham Research Grant proposal**: The research proposal is limited to 1,500 words.
  - Writing: as short and clear as possible, conditional on having every argument evidence-based
  - Tables and figures: keep only the relevant information
    - Most (if not all) results presented in the text should be in tables/figures [in the text direct your readers to your tables/figures]
    - Tables/figures should have enough information for readers
    - Prepare your own formats, don't copy and paste the `Stata` results
    - Online appendix
- Missing data
  - "for simplicity, missing cases are dropped" - the bottom line for simplicity is that missing data do not largely impact your estimates
  - When you drop cases, show where, why and how man you have dropped
- Be careful

Reminder: Workflow of data analysis using `Stata`

- Steps in workflow (Long et al., 2009)
  1. Have a good idea for a project
  2. Prepare and clean the data for analysis
     - General goals and analysis plan
     - An organized director structure

- Uniform formats for .do files
- Look for the variables of interest
- Variables must be carefully named, labeled, and cleaned
  * Categorical vs. continuous (e.g., code education)
  * Missing data
  * Truncated data
  * Drop cases
  * Random number generator with seed

3. Conduct analysis
   - Estimate models
   - Postestimation
   - Create graphs
4. Present results
   - Formats of tables and figures
5. Protecting files
   - Replication is impossible without your data and .do files
   - Never rewrite original data files
   - Version control of .do files (+date and documentation)

2 Conditional logit/probit models

2.1 The model

- Extensions of binary response model: multi-valued outcomes
  - Ordered, with even gaps
  - Ordered, with uneven (and unknown) gaps - Ordinal regression (ordered logit/probit)
  - Unordered, with case-specific covariates - Multinomial logit/probit
  - Unordered, with alternative-specific covariates - Conditional logit/probit (Mcfadden, 1974)

2.2 Estimation

2.2.1 Multinomial distribution

- Probability of each outcome I in **multinomial logit/probit**:

$$\mathbf{P}(Y = i) = \frac{exp(X\beta_i)}{\sum_{k=1}^{J} exp(X\beta_k)} \tag{76}$$

  - I alternative choices

- Case-specific covariates: X (same covariates for all alternatives)
- Alternative-specific coefficients: β_i

- Probability of each outcome I in **conditional logit/probit**:

$$\mathbf{P}(Y = i) = \frac{exp(Z_i\beta)}{\sum_{k=1}^{I} exp(Z_k\beta)} \tag{77}$$

  - I alternative choices
  - Alternative covariates: Z_i (different alternatives have different covariates)
  - Same coefficients across alternatives: β

- **Mixed model**:

$$\mathbf{P}(Y = i) = \frac{exp(X\gamma_i + Z_i\beta)}{\sum_{k=1}^{I} exp(X\gamma_k + Z_k\beta)} \tag{78}$$

2.2.2 (Post)Estimation

- Similarly, MLE/model fit/coefficients/marginal effects/

3 Interpreting results of conditional logit/probit models

3.1 Marginal effects

- Old command: `clogit`
- Newer command: `asclogit`
  - Allows both alternative/case-specific variables
  - Not available: `margins` (and `mtable`)
  - *The predicted probability of an individual choosing one option depends on characteristics of all of the options, thus the predictions sum to 1 within each individual*
  - `estat mfx`
  - *holds variables to their alternative-specific means*

3.2 Hoxby and Avery (2013)

- The hidden supply of high-achieving, low-income students
  - The vast majority of low-income high achievers do not apply to any selective college
- College application behavior
  - Students can apply to all college in the United States but decide to apply only to some (random utility model)
    * Conditional logit model: No preference ordering within the colleges applied

        * Rank-ordered logit model
    - Results from the conditional logit model
        * The coefficients are expressed as odds ratios so that a coefficient greater than 1 means that an increase in the covariate is associated with an increase in the probability that the student applies to the school, all other covariates held constant.

Table 3. Conditional Logit Regressions Explaining High-Achieving Students' College Applications[a]

Factor	*Low-income students*	*High-income students*
College is a peer school[b]	1.015	76.214***
College is a safety school[c]	3.009***	14.895***
College is nonselective	0.748***	1.6e-9***
Tuition before discount (thousands of dollars)	0.865***	1.176***
Average tuition discount (percent)	1.091**	0.925**
Could live at family home (college is <10 miles away)	4.942***	0.810***
Could go home often (college is <120 miles away)	1.556***	1.185***
Distance in miles to college	0.996	0.998
Square of (distance in miles/1,000)	1.056**	1.283***
College is in-state	2.595***	1.206***
College is private	0.838***	1.002
College is for-profit	0.834***	0.012***
Highest degree offered is 2-year	0.925**	0.009***
College is a university	0.997	0.567***
College is a liberal arts college	0.717***	0.973*

3.3 Jacob, McCall, and Stange (2016)

- To estimate student willingness to pay for various attributes of college
- Conditional logit model with preference heterogeneity
    - Potential OVBs and solutions (much of the existing college-choice literature does not address them)
        * Fixed unobserved differences between schools: Institutional fixed effects
        * Individual-specific price discounting: Net price (predicted) that varies across students, schools, and time
        * Selective admissions (choice sets): Weighted conditional logit model with weights equal to the likelihood that a given alternative is contained in an individual's choice set (similar to Inverse Probability Weighting)
- Interpretation using model estimates

- Willingness-To-Pay (Table 4, Figure 1)
  * Negative ratio of the coefficient on the attribute to the coefficient on log(net tuition)
  * Postestimation `test` to compute s.e.
  * Predicted WTP distribution graph: WTP for spending and distance can be interpreted as the percent increase in cost students are willing to pay to attend a college with a 1% increase in spending or 1% further away.
- Prediction graph (Figure 2)
  * The distribution of the percent change in total enrollment at each individual college if this college were to change a single characteristic

Table 4: Estimates of the Predictors of College Choice, No Preference Heterogeneity (Odds Ratios Reported)

	Dept Variable: College Chosen by High School Graduates in 1992 and 2004			
	(1)	(2)	(3)	(4)
Log (Tuition, Fees, Room & Board)	0.137 ***	0.195 ***	0.046 ***	0.389 ***
	(0.0039)	(0.0054)	(0.0018)	(0.0145)
Log (Distance)	0.327 ***	0.324 ***	0.315 ***	0.484 ***
	(0.0020)	(0.0020)	(0.0019)	(0.0042)
Log (Spending on Consumption Amenities/FTE)	2.032 ***	1.592 ***	1.402 ***	1.137 ***
	(0.0536)	(0.0417)	(0.0374)	(0.0294)
Log (Spending on Academics/FTE)	1.158 ***	1.484 ***	0.873 ***	0.880 ***
	(0.0375)	(0.0510)	(0.0306)	(0.0292)
School Mean SAT (percentile)	1.013 ***	1.018 ***	1.011 ***	1.006 ***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Institution state unemployment rate				0.948 ***
				(0.0138)
Log(high school grads in institution state)				0.997
				(0.0209)
College located in the student's home state				8.269 ***
				(0.3642)
College located in the student's census region				2.061 ***
				(0.0910)
Log (Lagged first time freshman enrollment)	Yes	Yes	No	No
Accounting for Probability of Admissions	No	Yes	Yes	Yes
Log (Predicted net price) used as cost measure	No	Yes	Yes	Yes
College Fixed Effects	No	No	Yes	Yes
Willingness-to-Pay (s.e.)				
Log (Distance)	-0.563	-0.689	-0.376	-0.768
	(0.0097)	(0.0140)	(0.0052)	(0.0318)
Log (Spending on Consumption Amenities/FTE)	0.357	0.284	0.110	0.136
	(0.0137)	(0.0168)	(0.0086)	(0.0275)
Log (Spending on Academics/FTE)	0.074	0.241	-0.044	-0.136
	(0.0161)	(0.0200)	(0.0110)	(0.0365)
School Mean SAT (percentile)	0.007	0.011	0.004	0.007
	(0.0004)	(0.0005)	(0.0003)	(0.0009)

3.4 Odds ratio in `outreg2`

- When `outreg2` outputs odds ratios, it does not output standard errors
- "seeform in parentheses"
- http://fmwww.bc.edu/repec/bocode/o/outreg2.ado

Table 3: Rural-urban differences in college choice (Source: Loyalka et al., 2016)

	Admitted by College *J*	
	(1)	(2)
Log(tuition)	0.868***	0.924
	(0.034)	(0.048)
*Urban student	1.124**	1.070
	(0.065)	(0.061)
Other controls	No	Yes

4 Additional section

4.1 Additional reading

- Muñoz et al. (2016), "Democratization and diversion: The effect of Missouri's A+ Schools Program on postsecondary enrollment." *Journal of Higher Education* 87 (6):801–830.
  - Democratizing effects of "**free community college**" on college enrollment rate by increasing two-year enrollment, but decreasing four-year enrollment
  - Difference-in-differences policy evaluation

4.2 `Stata` tip #14: Reshape data

```
UCLA IDRE Stata Learning Module
Reshape data wide to long, or long to wide

*http://www.ats.ucla.edu/stat/stata/modules/reshapel.htm
*http://www.ats.ucla.edu/stat/stata/modules/reshapew.htm
```

EDUC 799 LAB #11
Thanksgiving

Xiaoyang Ye
University of Michigan

November 23, 2016

OBJECTIVES OF LAB 11

1. Checklist
2. Conditional logit model: Practice in `Stata`
3. Research clinic

1 Checklist

- Questions about lecture, lab and office hours?
- Work on your final project
  - Last due: Nov 29, Tuesday, lit review, revised methodology, draft findings

Preliminary plan between Thanksgiving and Christmas, and beyond

- Lecture
  - Nov 29, Interaction effects
  - Dec 6, Introduction to causal inference
- Lab
  - Nov 30, Interaction effects, missing data in practice
  - Dec 7, Review, practice presentations, work on your final paper
- Dec 13/14: 15-min presentations
  - Do you prefer extending the lecture time on Dec 13 or using a two-day agenda?
  - Final paper due: Dec 14 Noon (hard deadline)
- After fall 2016
  - Present your papers in seminars and conferences
  - Submit them to journals

2 Conditional logit/probit models using `Stata`

- Old command: `clogit`
- Newer command: `asclogit`
  - Allows both alternative/case-specific variables
  - Not available: `margins` (and `mtable`)
  - *The predicted probability of an individual choosing one option depends on characteristics of all of the options, thus the predictions sum to 1 within each individual*
  - `estat mfx`
  - *holds variables to their alternative-specific means*

3 Research clinic

- Share your work in progress in pairs
- I will talk to the class if there are common questions

4 Additional section

4.1 Additional reading

- Hillman (2016), "Geography of college opportunity The case of education deserts." *American Educational Research Journal*, forthcoming.
  - Much of the college choice literature does not engage with the importance of geography in shaping educational destinations
    - Students often stay in close proximity to home and work when making college choice
  - The number of local colleges varies along lines of race and class
    - Communities with large Hispanic populations and low educational attainment have the fewest alternatives nearby, while White and Asian communities tend to have more.
  - These can result in education deserts, or places where opportunities richly available for some communities are rare (or even nonexistent) in others.

4.2 `Stata` tip #15: Report the baseline odds (exponentiated constant)

- Odds ratios have a bad reputation for being hard to interpret
- Using baseline odds ratios, we can assess the "effect size" o the coefficients

```
* Stata Journal (2012)
* Stata tip 107: The baseline is now reported
* By

* http://www.stata-journal.com/sjpdf.html?articlenum=st0251
```

EDUC 799 LAB #12
Interaction Effects in Non-linear Models

Xiaoyang Ye
University of Michigan

November 30, 2016

OBJECTIVES OF LAB 12

1. Interaction effects
2. Missing data in practice
3. How to give an applied micro talk

1 Checklist

- Questions about lecture, lab and office hours?
- Work on your final project
  - Final presentation and paper: Dec 13 Tuesday

2 Interaction effects

- Interaction effects are extensively used in applied research
- The best reference is Norton et al. (2004), "Computing interaction effects and standard errors in logit and probit models". *Stata Journal*, *4*, 154-167.

2.1 Linear model

- A linear regression model example: Impacts of family income on math score (β)?

$$Math_i = Income_i * \beta + \varepsilon_i \tag{79}$$

```
\\ NELS 1988
. reg math income

      Source |       SS           df       MS      Number of obs   =    11,294
-------------+----------------------------------   F(1, 11292)     =     52.71
       Model |  9460.06038         1  9460.06038   Prob > F        =    0.0000
    Residual |  2026768.55    11,292  179.487119   R-squared       =    0.0046
-------------+----------------------------------   Adj R-squared   =    0.0046
       Total |  2036228.61    11,293  180.308918   Root MSE        =    13.397


------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0362836   .0049978     7.26   0.000      .026487    .0460802
       _cons |   52.67856   .1536833   342.77   0.000     52.37732    52.97981
------------------------------------------------------------------------------

\\ Each a percentile increase in famliy income ranking is statistically
\\ significantly correlated with 0.04 SD increase in math score
```

- Interaction effects in linear models
  - Whether the impacts are different between male and female students?

$$Math_i = Income_i * \beta_1 + Female_i * \beta_2 + Income_i * Female_i * \beta_3 + \varepsilon_i \tag{80}$$

  - If β_3 (interaction effect) is **statistically significant**, there is a differential impact ("heterogeneous effects" in the treatment effect literature)

```
. reg math c.income##i.female

      Source |       SS           df       MS      Number of obs   =    11,294
-------------+----------------------------------   F(3, 11290)     =     27.19
       Model |  14604.2951         3  4868.09837   Prob > F        =    0.0000
    Residual |  2021624.31    11,290   179.06327   R-squared       =    0.0072
-------------+----------------------------------   Adj R-squared   =    0.0069
       Total |  2036228.61    11,293  180.308918   Root MSE        =    13.381


------------------------------------------------------------------------------
        math |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      income |   .0152906   .0073894     2.07   0.039      .000806    .0297752
    1.female |  -1.617467   .3076807    -5.26   0.000    -2.220575    -1.01436
             |
     female#|
    c.income |
          1  |    .038749   .0100221     3.87   0.000     .0191039    .0583941
```

```
               |
         _cons |   53.54196   .2245558   238.43   0.000     53.10179    53.98213
------------------------------------------------------------------------------

\\ Each a percentile increase in famliy income ranking is statistically
\\ significantly correlated with 0.02 SD increase in math score for male,
\\ and 0.05 SD for female students
```

2.2 Nonlinear model

- *The intuition from linear regression models does not extend to nonlinear models!*
  - The marginal effect of a change in both interacted variables is **not equal** to the marginal effect of changing just the interaction term
  - The sign may be different for different observations
  - Most applied researchers misinterpret the coefficient of the interaction term in nonlinear models (Ai and Norton, 2003)

2.2.1 Understanding the mechanisms

- A probit model with interaction terms

$$\mathbf{E}[y|x_1, x_2] = \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2) = \Phi(\mu) \tag{81}$$

  where Φ is the standard normal cumulative distribution.

  - Suppose that x_1 and x_2 are continuous, the marginal effect of just the interaction terms is

$$\frac{\partial \Phi(\mu)}{\partial (x_1 x_2)} = \beta_3 \Phi'(\mu) \tag{82}$$

    Most applied researchers interpret this as the interaction effect
  - However, the full interaction effects is the cross-partial derivative

$$\frac{\partial^2 \Phi(\mu)}{\partial (x_1) \partial (x_2)} = \beta_3 \Phi'(\mu) + (\beta_1 + \beta_3 x_2)(\beta_2 + \beta_3 x_1)\Phi''(\mu) \tag{83}$$

  - Four implications
    1. The interaction effect could be nonzero, even is $\beta_3 = 0$ ($\beta_1 \beta_2 \Phi''(\mu)$)
    2. The statistical significance of the interaction term cannot be tested with a simple t test on the coefficient of the interaction term β_3
    3. Unlike in linear models, the interaction effect is conditional on the independent variables
    4. The interaction effect may have different signs for different values of covariates

2.2.2 Understanding the odds ratio

- Read Ai and Norton (2003) for detailed formulas
- Odds ratio in logit models

$$P(Y=1|X) = \frac{1}{1+exp(-X\beta)} \tag{84}$$

  Odds ratio:

$$odds = \frac{p}{1-p} = exp(X\beta) \tag{85}$$

- Odds ratio when there is an interaction between two dummy variables, x_1 and x_2
  - The common interpretation is that the odds ratio for the interaction term equals $exp(\beta_3)$.
  - THIS IS NOT TRUE.
  - Odds ratio for $x_1|x_2 = 1$:

$$\frac{Pr(y=1|x_1=1,x_2=1)/(1-Pr(y=1|x_1=1,x_2=1))}{Pr(y=1|x_1=0,x_2=1)/(1-Pr(y=1|x_1=0,x_2=1))} = \frac{exp(\beta_1+\beta_2+\beta_3)}{exp(\beta_2)} \tag{86}$$

  - Odds ratio for $x_1|x_2 = 0$:

$$\frac{Pr(y=1|x_1=1,x_2=0)/(1-Pr(y=1|x_1=1,x_2=0))}{Pr(y=1|x_0=0,x_2=0)/(1-Pr(y=1|x_0=0,x_2=0))} = \frac{exp(\beta_1)}{exp(0)} \tag{87}$$

  - Ratio of odds ratio for x_1 and x_2

$$\frac{exp(\beta_1+\beta_2+\beta_3)}{exp(\beta_2)} / \frac{exp(\beta_1)}{exp(0)} = exp(\beta_3) \tag{88}$$

  $exp(\beta_3)$ is not a risk ratio, or even an odds ratio. It's the **ratio of odds ratios**

  - Recall that, odds ratio is the ratio of odds for two different observations that differ only in the value of one explanatory variable

2.2.3 Estimating interaction effect using `Stata`

- This is a post-estimation procedure to compute the interaction effect (and its standard error) after nonlinear regressions
- The command `predictnl` can be used to compute nonlinear predictions
- Norton et al. (2004) develops the `inteff` command
  - `inteff3`: triple dummy interactions by Cornelißen, Sonderhof et al. (2009)
- Excellent posts
  - http://www.ats.ucla.edu/stat/stata/seminars/interaction_sem/interaction_sem.htm
  - Interaction effects in logistic and probit regression by CRMportals Inc.

- Interaction effects can be very much complicated
  - Bai (2009), "Panel data models with interactive fixed effects". *Econometrica, 77*(4), 1229-1279.
  - Two endogenous interacted variables (dynamic treatment effects of high school math by Steve, Brian, & Xiaoyang)

3 Missing data in practice

- Missing data problem is annoying
  - Missing X variables
  - Missing Y variables
    - A special case: **The fundamental problem of causal inference**
- There is not a best way to address the problem
  - Robustness checks using different approaches

3.1 Example

- Dobbie and Fryer (2015), "The medium-term impacts of high-achieving charter schools". *Journal of Political Economy, 123*(5), 985-1037.
  - Estimating the effects of high-achieving charter schools on human capital (high school graduation, college enrollment), risky behaviors, and health outcomes
  - **Empirical strategy:** Charter school lottery as the instrumental variable
- **Problems with missing data** in Dobbie and Fryer (2015)
  - Sample attrition (missing Y variable)
    - Why
      - Lottery winners were 11.8 percentage points more likely to respond to our survey.
      - If lottery losers who did not respond to the survey differ in some important way, this could invalidate our empirical design by creating unobserved differences between the treatment and control groups
    - Solutions ("In 36 out of 46 cases, our qualitative results are unchanged")
      1. Compare treatment effects on the subsample of students who completed our survey and the larger set of students for whom we have administrative data (see Table A3, regressions on each outcome using maximum observations)
      2. Lee bounds (Lee, 2009)
      3. Impute outcomes for youths who did not respond to the survey (using minimum values) and estimate median regressions

- Non-response in surveys (missing Y variable)
  - Show evidence on no differential survey response between lottery winners and losers on observable baseline characteristics (see footnote 11 and Table 2)
- Missing X variables
  - Average test score of three subjects. If youths are missing one or two of these exams, calculate the average using just the nonmissing scores
    - "Results are nearly identical dropping these observations or imputing missing values"
  - "When appropriate, we also include indicators for missing variables to prevent attrition."
  - *Note for two additional approaches: You may want to do multiple imputations to test the robustness of the results; or by regressing without the covariates with missing values*

4 Presentations

- "How to give an applied micro talk" by Jesse M. Shapiro
  - Some tips: Use your talk to tell your story
    - Your audience does not care about your topic, you have 1-2 slides to change their minds
    - State what your paper does and why?
    - State clearly your data/sample/variable/estimation model
    - Discuss the most important vulnerabilities of your modeling approach
    - Every word is precious, use figures/tables wherever possible
  - By the end of your talk, **make sure that your audience**
    - Cares about your research question
    - Understands how you answer it
    - Knows why they should believe you
    - Walks out of the room knowing what you learned

5 Additional section

5.1 Additional reading

- Fryer (2016), "Information, non-financial incentives, and student achievement: Evidence from a text messaging experiment". *Journal of Public Economics, 144*, 109-121.

– A field experiment in Oklahoma City Public Schools in which students were provided with free cellular phones and daily information about the link between human capital and future outcomes via text message in one treatment and minutes to talk and text as an incentive in a second treatment.
– Students' reported beliefs about the relationship between education and outcomes were influenced by the information treatment.
– There were no measurable changes in student effort, attendance, suspensions, or state test scores, though there is evidence that scores on college entrance exams four years later increased.

5.2 `Stata` tip #16:

- Christmas Tree in `Stata`

```
****************************************************************************
**  PROGRAM FOR CHRISTMAS TREE                                            **
****************************************************************************
* Xiaoyang Ye
* Jan 1, 2014
* Revised from Internet resource
* http://www.stata.com/statalist/archive/2005-12/msg00710.html
* Updated: Nov 30, 2016

* Standard setup
version 14
capture clear
capture log close
set more off, perm

* Drop existng program
capture program drop MerryChristmas

* Define program
program define MerryChristmas
 di "{break}"
 forval i=0(4)16 {
  forval j=1(2)9 {
   local n=`i'+`j'
   local tree=" "
   forval k=1/`n' {
    if `n'==1 {
     local tree="`tree'"+"{res}*"
    }
    else if uniform()>0.8 {
     local tree="`tree'"+"{error}*"
    }
    else {
```

```
     local tree="`tree'"+"{text}*"
    }
   }
   di "{center:`tree'}"
  }
 }
 di "{center:{input}2016}"
 di "{center:{input}~~ Merry Christmas ~~}"
 di "{center:{input}## Happy New Year ##}"
 di "{center:{input}2017}"
 di "{hline}"
 di "{center:{input}*Xiaoyang Ye*}"
end
/* You may want to change contents in "" */

* Run program
MerryChristmas

* END AND HAPPY NEW YEAR 2017 IN ADVANCE
```

Questions? Suggestions?

References

Abbring, Jaap H and James J Heckman. 2007. "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation." *Handbook of econometrics* 6:5145–5303.

Ai, Chunrong and Edward C Norton. 2003. "Interaction terms in logit and probit models." *Economics letters* 80 (1):123–129.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

———. 2014. *Mastering 'metrics: The path from cause to effect*. Princeton University Press.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica* 77 (4):1229–1279.

Bettinger, Eric P, Brent J Evans, and Devin G Pope. 2013. "Improving college performance and retention the easy way: Unpacking the ACT exam." *American Economic Journal: Economic Policy* 5 (2):26–52.

Bulman, George, Robert Fairlie, Sarena Goodman, and Adam Isen. 2016. "Parental resources and college attendance: Evidence from lottery wins." *NBER Working Paper No.22679* .

Cameron, Adrian Colin and Pravin K Trivedi. 2009. *Microeconometrics using Stata*, vol. 5. Stata press College Station, TX.

Cornelißen, Thomas, Katja Sonderhof et al. 2009. "Partial effects in probit and logit models with a triple dummy-variable interaction term." *Stata Journal* 9 (4):571.

DesJardins, Stephen L, Dennis A Ahlburg, and Brian P McCall. 2006. "An integrated model of application, admission, enrollment, and financial aid." *Journal of Higher Education* 77 (3):381–429.

DiNardo, John and David S Lee. 2011. "Program evaluation and research designs." *Handbook of labor economics* 4:463–536.

Dobbie, Will and Roland G Fryer. 2015. "The medium-term impacts of high-achieving charter schools." *Journal of Political Economy* 123 (5):985–1037.

Ebenstein, Avraham, Victor Lavy, and Sefi Roth. 2016. "The long run economic consequences of high-stakes examinations: evidence from transitory variation in pollution." *American Economic Journal: Applied Economics* 8 (4):36–65.

Fryer, Roland G. 2016. "Information, non-financial incentives, and student achievement: Evidence from a text messaging experiment." *Journal of Public Economics* 144:109–121.

Greenberg, Edward. 2012. *Introduction to Bayesian Econometrics*. Cambridge University Press.

Greene, William. 2004. "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects." *The Econometrics Journal* 7 (1):98–119.

Hayashi, Fumio. 2000. *Econometrics*. Princeton University Press.

Heckman, James J, John Eric Humphries, and Gregory Veramendi. 2016. "Dynamic treatment effects." *Journal of econometrics* 191 (2):276–292.

Heckman, James J and Edward J Vytlacil. 2007a. "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation." *Handbook of econometrics* 6:4779–4874.

———. 2007b. "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments." *Handbook of econometrics* 6:4875–5143.

Hilger, Nathaniel G. 2016. "Parental job loss and children's long-term outcomes: Evidence from 7 million fathers' layoffs." *American Economic Journal: Applied Economics* 8 (3):247–283.

Hillman, Nicholas W. 2016. "Geography of College Opportunity The Case of Education Deserts." *American Educational Research Journal* :0002831216653204.

Holland, Paul W. 1986. "Statistics and causal inference." *Journal of the American statistical Association* 81 (396):945–960.

Hoxby, Caroline and Christopher Avery. 2013. "The missing" one-offs": The hidden supply of high-achieving, low-income students." *Brookings Papers on Economic Activity* 2013 (1):1–65.

Hsin, Amy and Yu Xie. 2014. "Explaining Asian Americans' academic advantage over whites." *Proceedings of the National Academy of Sciences* 111 (23):8416–8421.

Imbens, Guido W and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Jacob, Brian, Brian McCall, and Kevin M Stange. 2016. "College as country club: Do colleges cater to students' preferences for consumption?" *Journal of Labor Economics* (forthcoming).

Johnson, John H and Mike Gluck. 2016. *Everydata: The misinformation hidden in the little data you consume every day*. Routledge.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *The Review of Economic Studies* 76 (3):1071–1102.

Liu, Airan and Yu Xie. 2016. "Why do Asian Americans academically outperform Whites? - The cultural explanation revisited." *Social Science Research* 58:210–226.

Long, Bridget Terry. 2004. "How have college decisions changed over time? An application of the conditional logistic choice model." *Journal of Econometrics* 121 (1):271–296.

Long, J Scott et al. 2009. *The workflow of data analysis using Stata*. Stata Press College Station.

Manski, Charles F. 2009. *Identification for prediction and decision*. Harvard University Press.

Mcfadden, D. 1974. "Conditional logit analysis of qualitative choice behavior." *Frontiers in Econometrics* :105–142.

Muñoz, José, James Harrington, Bradley R Curs, and Mark Ehlert. 2016. "Democratization and diversion: The effect of Missouri's A+ Schools Program on postsecondary enrollment." *Journal of Higher Education* 87 (6):801–830.

Murnane, Richard J and John B Willett. 2010. *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Norton, Edward C, Hua Wang, Chunrong Ai et al. 2004. "Computing interaction effects and standard errors in logit and probit models." *Stata Journal* 4:154–167.

Poirier, Dale J. 1995. *Intermediate statistics and econometrics: A Comparative approach*. Mit Press.

Posselt, Julie Renee, Ozan Jaquette, Rob Bielby, and Michael N Bastedo. 2012. "Access without equity longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004." *American Educational Research Journal* 49 (6):1074–1111.

Scott-Clayton, Judith and Lauren Schudde. 2016. "Performance standards in need-based student aid." *NBER Working Paper No.22713* .

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.