ANDREI SIRCHENKO

TEACHING PORTFOLIO

Contents

1. INTRODUCTION

"Education is what remains after one has forgotten everything he learned in school." --- Albert Einstein

"The ultimate aim of education is to enable individuals to become the architects of their own education and through that process to continually reinvent themselves." --- Elliot Eisner

I firmly believe that the only kind of education available is self-education. Actually, all education is self-education. A teacher is only a guide, whose goal is not to fill the students' minds with facts, but rather to teach them to think. I personally prefer an interactive teaching style with reading material ahead and discussing it in the class, a lot of the in-class games, tests and other activities for learning by doing.

I have substantial 7-year teaching experience. I have independently taught Statistics, Econometrics and Game Theory as a lecturer for master and undergraduate students. I have set out the syllabi, problem sets and the tests for all the courses and conducted the examinations. I have also leaded (as a TA) the seminars in the large-enrollment multi-section undergraduate courses in Microeconomics and Statistics for business, economics and finance students. Finally, I have been a TA for a PhD course in Econometrics, taught by Prof. Helmut Lütkepohl.

- **Teaching Assistant**, **European University Institute,** Florence, Italy 2010
  - ✓ *Time Series Econometrics (PhD course by Prof. H. Lütkepohl)*
- **Visiting Instructor, International College of Economics and Finance** at **Higher School of Economics,** Moscow, Russia 2008
  - ✓ *Statistics and Econometrics (undergraduate)*
- **Visiting Lecturer, Economics Education and Research Consortium** 2005 - 08
  - ✓ *Thirty four mini-courses in Econometrics and Game Theory (undergraduate and graduate) at twenty five universities in Ukraine, Belarus, Moldova, Georgia and Armenia*
- **Visiting Lecturer, Kiev School of Economics,** Kiev, Ukraine 2004
  - ✓ *Statistics and Econometrics (graduate)*
  - ✓ *Categorical Dependent Variables (graduate)*
- **Teaching Assistant**, **University of Iowa,** Iowa City, USA 2000 - 03
  - ✓ *Statistics for Business Strategy (undergraduate) - 4 semesters*
  - ✓ *Principles of Microeconomics (undergraduate) - 2 semesters*

Below you can find the teaching materials for the graduate courses on "Discrete and Limited Dependent Variable Models" and "Statistics and Econometrics" I have taught at Kiev School of Economics in Ukraine.

2. CATEGORICAL DEPENDENT VARIABLES

EERC MA Program **January-February 2004**

Syllabus for
Categorical Dependent Variables

Instructor: Andrei Sirchenko

Course description:

We will loosely follow the textbook *Regression Models for Categorical and Limited Dependent Variables* by Scott Long, emphasizing the following main topics:

- ✓ Binary outcomes: the linear probability, probit, and logit regression models
- ✓ Ordinal outcomes: ordered logit and ordered probit analysis
- ✓ Nominal outcomes: multinomial logit, conditional logit and related models
- ✓ Hypothesis testing, goodness of fit, diagnostics and interpretation of regression models for categorical dependent variables
- ✓ Limited outcomes: censored and truncated regression models

Other topics that may be discussed, time permitting, are inference for contingency tables, generalized linear models, regression models for count outcomes, and hazard models.

The statistical package STATA will be used extensively in this course. However, you may work with a software package that you are most comfortable with.

The course will survey theory and methods for the analysis of categorical and limited dependent variables with a focus on applications.

Course organization:

Reading ahead is crucial. Homework problems (mostly applied and computational) and an empirical research project will be an essential part of the course. You are encouraged to work together to help each other in understanding the course material and completing the homework problems. However, each student has to write up his/her own solutions. Projects must be done individually. Quizzes will be given on a regular basis, and all of them will be announced beforehand. There will be one final examination.

Course objective:

Those of you who successfully complete this course will have a good understanding of the underlying theoretical principles of categorical and limited dependent variable analysis. You will be familiar with a variety of methods for analyzing discrete or limited data and understand how they can be used for empirical research. You will obtain good working knowledge of the STATA software.

Textbooks:

The **required** textbook is *Regression Models for Categorical and Limited Dependent Variables* by J. Scott Long (SAGE Publications, Inc., 1997).

The **recommended** supplementary textbook is *Regression Models for Categorical Dependent Variables using STATA* by J. Scott Long (STATA Corporation, 2003, revised edition).

Other books on categorical data analysis that are more theoretical and are well worth reading, even though they are not required on this course:

Limited-Dependent and Qualitative Variables in Econometrics by G. S. Maddala (Cambridge University Press, 1983);

Categorical Data Analysis by Alan Agresti (John Wiley & Sons, Inc., second edition, 2002)

Evaluation:

Grades will be calculating as follows:

- Homework 15 % (the lowest homework score will be dropped)
- Quizzes 15 % (the lowest quiz score will be dropped)
- Project 35 %
- Final Exam 35 %

You will not be "graded on a curve", and hence you are not competing with fellow students. Therefore, you are not penalized for working together to better learn and understand the material.

Other information:

The EERC and the University policies on academic integrity covering cheating, plagiarizing, and other acts of academic dishonesty will be adhered to very closely.

More information on the course guidelines, policies, and grading procedures will be given during the first meeting of the class.

Special arrangements can be made for students with disabilities.

Please feel free to contact me, should you have any questions or comments.

Quizes

Available upon request.

Homework 1

1. Identify each variable as qualitative (nominal or ordinal) or quantitative (discrete or continuous):

a. Political party affiliation (Democrat, Republican, Green, other)
b. Anxiety rating (none, mild, moderate, severe, very severe)
c. Number of children in a family
d. Income
e. Clinic location (Kyiv, Ternopil, L'viv, Yalta, Poltava)
f. Response of tumor to chemotherapy (complete elimination, partial reduction, stable, growth progression)
g. Favorite beverage (beer, wine, vodka, gorilka, kefir)
h. Number of sexual intercourses during a given day (none, one, two, three, four or more, don't remember)

2 (optional). Russian Roulette (which originally came to Russia from France, ironically) consists of putting a bullet in one of the six chambers of a pistol, spinning the chambers to select one at a random, and then firing the pistol once at a one's head.

a. One played this game six times and survived. Find the probability of this outcome.
b. Suppose that he kept playing this game until the bullet fired. Let Y denote the number of the game on which it fires. Show the probability mass function for Y.

If you are not good enough with STATA, go through Chapter 2 of "Regression Models for Categorical Dependent Variables using STATA" by Long and Freese (on reserve in the library). Go to http://www.stata-press.com/data/regmodcdvs-rev.html and download data and do-files used in the book.

3 (not to be submitted). Use file binlfp2 with data on labor force participation. Replicate the logit and probit estimations made in the section 4.2. Familiarize yourself with STATA commands, options and outputs for binary outcomes. Work together if possible!

All other data sets for the homeworks can be downloaded from the folder ASIRTCHENKO on the EERC file server Kitty.

4. Use data (see the file "Data Hw1.xls") from a study of nesting horseshoe crabs. Each female horseshoe crab had a male crab resident in her nest. The study investigated factors affecting whether the female crab had any other males, called satellites, residing nearby. Explanatory variables are:

- C - the female crab's color (1, light medium; 2, medium; 3, dark medium; 4, dark)
- S - spine condition (1, both good; 2, one worn or broken; 3, both worn or broken)
- Wt - weight (g)
- W - carapace width (cm)

The response outcome for each female crab is her number of satellites (Sa). Let Y=1 if a crab has at least one satellite, and Y=0 otherwise.

I. Regress Y on X=[weight]:
   a. Use the linear probability model. Interpret the parameter estimates. Find the estimated probability at the highest observed weight (5.20 kg). Comment.
   b. Fit the logistic regression model. Find the fitted probability at 5.20 kg.
   c. Fit the probit model. Find the fitted probability at 5.20 kg.

II. Fit the logistic regression of Y on X=[width]. Determine the odds increase from x=26.3 to 27.3.

5. Use data (see the file "Data Hw1.xls") from an epidemiological survey of 2484 subjects to investigate snoring as a risk factor for heart disease. Those surveyed were classified according to their spouses' report of how much they snored. Check out if the probability of heart disease is related to the level of snoring. Use linear, logit and probit models.

I. Use (0, 2, 4, 5) score for the level of snoring, treating the last two levels as closer than the other adjacent pairs.

II. Now refit the models using the scores:
   a. (0, 2, 4, 6)
   b. (0, 1, 2, 3)
   c. (1, 2, 3, 4)

Compare regression parameters **b** for the four choices. Compare fitted values. Summarize the effect of linear transformations of scores in part II, which preserve relative sizes of spacing between scores. Summarize the effect of nonlinear transformation of scores in case I and case IIa.

Enjoy!

Homework 2

1. (NOT to be submitted) Go through chapter 3 (Estimation, Testing, Fit, and Interpretation) of "Regression Models for Categorical Dependent Variables using STATA" by Long and Freese (on reserve in the library). Familiarize yourselves with STATA commands for estimation, testing, fit and interpretation and replicate them.
2. (Optional) Consider the choice between two options, such as two product brands. Let U_o denote the utility of outcome y = 0 and U_1 denote the utility of y = 1. For y = 0 and 1, suppose that $U_y = \alpha_y + \beta_y x + \varepsilon_y$, using a scale such that ε_y has some standardized distribution. A subject selects y = 1 if $U_1 > U_o$ for that subject. If ε_0 and ε_1 are independent $N(0,1)$ random variables, show that $Pr(Y=1)$ satisfies the probit model.
3. (NOT to be submitted) Read chapter 4 (Models for Binary Outcomes) of Long & Freese. Reproduce all STATA results shown in the chapter.

Remark. *Whenever you are asked to perform a test, you have to provide null hypothesis, test statistics and your conclusion.*

Remark. *Use data from the file* "Data HW2.xls" in the folder ASIRTCHENKO on Kitty

4. The data at page "Intercourse" in the file appeared in a national study of 15- and 16-year-old adolescent. The event of interest is ever having sexual intercourse. Analyze the data using logit regression:
    a. Test significance of individual slope coefficients using z- (or t-) test.
    b. Test significance of individual slope coefficients using Wald (chi-squared) test.

c. How are z- and chi2-tests related?

d. How many times are males more likely to have sexual intercourse than females? What is the observed ratio of frequencies of having intercourse for males and females? Comment the difference between predicted and observed ratios.

e. How many times are blacks more likely to have sexual intercourse than whites? What is the observed ratio of frequencies of having intercourse for blacks and whites? Comment the difference between predicted and observed ratios.

f. What are count and adjusted count R^2?

g. Compare predicted probabilities and observed frequencies for each combination of explanatory variables (there are four of them).

h. Report Efron's, McFadden's, adjusted McFadden's, McKelvey and Zavoina's, Maximum likelihood, and Cragg and Uhler's R^2.

i. Test hypothesis that all slopes are 0 with both Likelihood Ratio and Wald tests.

j. Report Pearson goodness-of-fit test.

k. Report classification table.

Remark. *The following three tasks will be explained in the class on Monday, Feb. 2.*

l. Report receiver operating characteristic (ROC) curve.

m. Report concordance index.

n. Report sensitivity and specificity plots

5. (Optional) For binary observations, consider the model $\Pr(y=1|x) = \frac{1}{2} + (1/\pi)\tan^{-1}(\alpha + \beta x)$. Which distribution has cdf of this form? Explain when this model might be more appropriate than logistic or probit regressions.

6. Use the data at page "Crab" (the same data as in the HW#1). Analyze the data using probit regression:

Part I.

a. Treat color as qualitative nominal variable. Regress y on color and width.

b. On the same graph plot predicted probabilities of y=1 versus width for all four colors. Interpret effects of width and color.

c. Test whether color contributes significantly to your model, i. e. test H_o: slopes of all color indicators are 0. Use Wald test.
d. Does the model permitting (color x width) interaction provide an improved fit? Use LR test.
e. Now use AIC.
f. Finally use BIC.

Part II. Now treat color as quantitative. Use scores as they are given in the table (1, 2, 3, 4).

a. Regress y on color and width
b. Using LR test, compare this fit to the more complex model in part I-a having a separate parameter for each color.
c. Now use AIC and BIC.
d. Change the scores for color. Let 0 be a score for color you dropped from the model in part I. For other three colors let scores be equal the coefficients of corresponding indicator variables from probit model in part I.
e. Compare the fit of the models with new and old scores using Hosmer-Lemeshow goodness-of-fit test.
f. Finally use the entire battery of tests provided by **fitstat** command with **dif** option (see section 4.5 of Long & Freese). Comment.
g. Choose the best model, add spine condition as a new regressor and check significance of its effect.

Part III. Now fit a model using weight and width as predictors. Conduct

a. Likelihood-ratio test of H_o: $\beta_1 = \beta_2 = 0$.
b. Separate tests for the partial effect of each regressor.
c. Why does neither test in part (b) show evidence of an effect when the test in part (a) shows strong evidence?

Part IV. Now consider two models:

- model A with width and color as explanatory variables;
- model B with weight and color as explanatory variables.

Use logit regression for both models:

a. Plot standardized Pearson residuals against width (weight) and identify outliers.
b. Plot Cook's distances against width (weight) and identify influential observations. Are they the same for both models?
c. Plot Delta chi-squared influence statistic against width (weight) and identify influential observations. Are they the same for both models?
d. Compare Hosmer-Lemeshow goodness-of-fit test for both models.
e. Finally use the entire battery of tests provided by **fitstat** command with **dif** option to choose your best model for this data. You are not constrained to the above models. If you can suggest the model superior to both A and B, go ahead!

7. (Optional) Fit your model from previous problem using EViews:
   a. Report and explain the differences in estimation results between STATA and EViews if any.
   b. In EViews try different optimization algorithms (Quadratic Hill Climbing, Newton-Raphson, BHHH) and comment differences in estimation results if any.

Homework 3

1. (NOT to be submitted) Read chapters 5 (Models for Ordinal Outcomes) and 6 (Models for Nominal Outcomes) of Long & Freese. Reproduce all STATA results shown in the text.

Remark. *Use data from the file "Data HW3.xls" in my folder on Kitty*.

2. The data at page "Party" refer to the effect on political party identification of gender and race. Fit a baseline-category logit model.
   a. What is the estimated odds ratio of preferring Democrat instead of Republican for whites and blacks holding gender constant?
   b. What is the estimated odds ratio of preferring Democrat instead of Republican for males and females holding race constant?

c. Find the prediction equation for log of odds of Independent vs. Republican identification.
d. What are estimated probabilities of all three party identifications for white males?
e. What are estimated probabilities of all three party identifications for black males?
f. Provide a table of predicted probabilities of all outcomes for all combinations of a set of race and gender.
g. Test the significance of interaction effect using LR and Wald tests.
h. Test the independence of irrelevant alternatives (IIA) assumption using the Hausman test.

Course Project

In this empirical project you are expected to apply your knowledge on ordinal probit and ordinal logit analysis. Your goal is to perform a categorical data analysis of the manner in which the US Federal Open Market Committee (FOMC) sets the level of the federal funds rate target – one of the most publicized and anticipated indicators in the financial world.

One half of the class, namely students whose last names start at any letter between (and including) A and M, has to use ordered probit model. The other half of the class, whose family names start at any letter between (and including) N and Z, is supposed to do ordered logit regression.

Before you start your research it is worthwhile to refresh your knowledge of tools and conduct of monetary policy (you may want to review chapter 17 and 18 of Mishkin's *The Economics of Money, Banking, and Financial Markets* to get an overall idea).

Another important source of information is Federal Reserve Banks' web-pages. Go to http://www.federalreserve.gov, where you will see a lot of useful links (check out http://www.federalreserve.gov/policy.htm, http://www.federalreserve.gov/rnd.htm, Personal Financial Education, http://www.newyorkfed.org/education/).

The project should be done individually and submitted both as hard copy and via e-mail. Deadline is TBD.

All of you should regress the same dependent variable – the federal funds target rate (target, for short). The data for project is available in my folder on Kitty – open file "Data Project.xls". There are 178 observations in your sample which reflect FOMC meetings during 1989-2004 period. At that period FOMC changes target in discrete amounts by multiples of 25 basis points (with few exceptions). The level of target in our sample has a range from 1% to 9.75% with more than 30 categories. Our sample size is not enough to perform categorical data analysis of target level per se. However, first differences, or target changes from meeting to meeting, fall into 5 main categories: -0.5%, -0.25%, 0%, 0.25%, 0.5% (one 0.75% change and few changes smaller than 0.25% can be merged with adjacent category).

Your task is to find relevant explanatory variables and combine them into so-called *federal funds reaction function* – your regression equation. Thus, you have to build an econometric model of the target. Hence, you are expected to deal with all aspects of econometric modeling: data search, model evaluation, diagnostics and selection, reporting of results. Later on this week you will be given some guides and minimum requirements for the project. However, you are encouraged to apply your way of thinking, choice of methodology and creativity.

There are some free sources of US economic data you need for analysis:

- ✓ The Federal Reserve Board at http://www.federalreserve.gov/rnd.htm;
- ✓ The Federal Reserve Bank of St. Louis' FRED II® (Federal Reserve Economic Data) – probably, the best one for our purposes;
- ✓ Bureau of Labor Statistics (BLS) at http://www.bls.gov/;
- ✓ Bureau of Economic Analysis (BEA);
- ✓ Census Bureau at http://www.census.gov/

FINAL EXAMINATION

Available upon request.

3. STATISTICS AND ECONOMETRICS

EERC MA Program **January-February 2004**

Syllabus for
Statistics and Econometrics III

Instructor: Andrei Sirchenko

Course description:

We will cover the following topics:

- ✓ Specification errors
- ✓ Model evaluation and diagnostics
- ✓ Maximum likelihood, generalized least squares, and instrumental variable estimators
- ✓ Univariate time series modeling
- ✓ Autoregressive distributed lag relationships

Other topics may be discussed as far as time permits.

Textbooks:

We will use the following textbooks:

- Johnston and DiNardo (1997), *'Econometric Methods'*, 4th edition, McGraw-Hill
- Gujarati, D. (1995), *'Basic Econometrics'*, 3rd edition, McGraw-Hill

The lectures will be based on chapters 4, 5, 7, 8 from Johnston and DiNardo and chapters 13, 14 from Gujarati.

Additional source of information is '*Econometric Analysis'* by W. Green.

Evaluation:

- Homework 15 % (the lowest homework score will be dropped)
- Quizzes 15 % (the lowest quiz score will be dropped)
- Midterm Exam 30 %
- Final Exam 40 %

Other information:

The EERC and the University policies on academic integrity covering cheating, plagiarizing, and other acts of academic dishonesty will be adhered to very closely.

More information on the course guidelines, policies, and grading procedures will be given during the first meeting of the class.

Special arrangements can be made for students with disabilities.

Please feel free to contact me, should you have any questions or comments.

Quizes

Available upon request.

Homework 1

From Johnston & DiNardo textbook:

Problems: 4.1, 4.2, 4.3, 4.8, 4.9

Computational problem:

Use file auto1.wf1 located in the folder Chap1 from J&D's data disk. Take sample from 1959:1 through 1973:3. Your response variable is y - ln of per capita real expenditure on gasoline and oil. Your explanatory variables are x2 (ln of the real price of gasoline and oil, where 10 to the

power of 4 – is an arbitrary scaling factor) and x3 (ln of per capita real disposable income). Use EViews to perform the following (provide all estimation outputs):

Part I.

1. Regress y on x2 and x3. Do you see any signs of parameter inconstancy from regression statistics?
2. a) Perform the Chow forecast test. Use 1959:1-1971:3 period for estimation and 1971:4-1973:3 period for testing. Comment.

   b) Compute and report the vector of prediction errors d. Compute and report matrix X2*inv(X1'*X1)*X2', where X1 and X2 are matrices of the explanatory variables for estimation and testing period correspondingly. Now using eq. (4.9) calculate F statistics of the Chow forecast test and compare it with one from part a). You can use either Matlab or Excel for matrix manipulations.
3. a) Perform the Chow breakpoint test using 1971:4 observation as a breakpoint. Comment.

   b) Using the next formula after eq. (4.36) compute and report the restricted and the unrestricted RSSs, and F statistics of the Chow breakpoint test. Compare it with one from part a).
4. Perform and comment the following tests based on recursive estimation:
   a) recursive residuals (one-step ahead prediction errors);
   b) recursive coefficients;
   c) CUSUM test;
   d) CUSUMSQ test.
5. Perform and comment Ramsey RESET test. What happens if you include more than one fitted term and why?

Part II.

1. Regress y on x2, x3 and interaction term x2*x3. Comment.

2-5. Do the same stability tests as in Part I (except 2b and 3b).

Enjoy!

Homework 2

From Gujarati textbook:

Problems: 13.10, 13.14, 13.15, 13.16, 13.17, 13.18 (optional), 14.7, 14.8. Enjoy!

Homework 3

From Johnston & DiNardo textbook:

Problems: 5.1, 5.3, 5.4 (also obtain the asymptotic distribution of the ML estimator of θ).

Computational problem:

Use data from "Data hw3.xls". Regress FEDFUNDS on intercept and CPIAUCSL, PCEIfePI, UNRATE, TCU, M1SL, INDPRO and PPIACO. This is your unrestricted model.

1. Using formulae 5.12, 5.15 and 5.18 perform LR, Wald and LM tests of H_0: coefficient of CPIAUCSL is zero. Report all three statistics and verify that W >= LR >= LM.
2. Now perform LR test of the same restriction using built-in software test (you can do this with EViews or Stata). Report the results.
3. Calculate LM statistic using formula on page 150: LM = nR^2. Compare with LM obtained in part 1). Comment.
4. Using formulae 5.12, 5.15 and 5.18 perform LR, Wald and LM tests of H_0: coefficient of TCU plus coefficient of PCEIfePI is zero. Report all three statistics and verify that W >= LR >= LM.
5. Now perform LR test of the same restriction using built-in software test (you can do this with Stata). Report the results.
6. Calculate LM statistic using formula on page 150: LM = nR^2. Compare with LM obtained in part 4). Comment.

Enjoy!

Homework 4

From Johnston & DiNardo textbook: problems: 7.1, 7.2, and 7.3.

Computational problem:

Use the data from "HW4.xls".

Obtain graphs and sample correlograms up to 36 lags for all 7 time series. Intuitively, which one(s) of these time series seem to be stationary?

Obtain graphs and sample correlograms up to 36 lags for first differences of all 7 time series. Intuitively, which first difference(s) seem to be stationary?

Now perform formal tests for stationarity using augmented Dickey-Fuller (ADF) test for all series. In order to choose the lag order of lagged first differences of the dependent variables check out the absence of serial correlation among residuals using Q-statistics. Report the order of integration for each series.

Enjoy!

Midterm exam

Available upon request.

Final Exam

Available upon request.