

James Raymond Vreeland  
Assistant Professor  
E-Mail: james.vreeland@yale.edu  
Tel: 203-432-5252  
Office hours: Wed. 10am to 12pm

Dept. of Political Science  
Yale University  
Room 300  
124 Prospect Avenue  
New Haven, CT 06520

## **Statistical Methods for research in International Relations and Comparative Politics**

Generic problems in comparative research concern the origins or the consequences of some systemic features of a country. Such problems often entail discrete or limited dependent variables and endogenous sampling.

The course will cover statistical models of discrete and limited dependent variables, leading to the problem of non-random selection and the appropriate ways of handling it in statistical research as well as in case studies. The focus will be on selection models. We will gradually learn all the building blocks – probit, logit, and tobit – and then put things together. Applications will include the origins of democracy, bilateral agreements, and the impact of political regimes and institutions (such as the IMF) on economic performance.

The emphasis will be practical. Students will be using a cross-national data file (ACLP) and a software package (LIMDEP). The course will be useful for students who are doing or think of doing analyses involving limited or qualitative dependent variables and for students working on policy evaluation.

Prerequisite: OLS, basic calculus.

### **Textbooks:**

Kennedy, Peter. 1998. *A Guide to Econometrics*. 4th Edition. Cambridge, MA: MIT Press.

King, Gary. 1989. *Unifying Political Methodology*. Cambridge: Cambridge University Press.

Greene, William H. 1997. *Econometric Analysis*. 3rd Edition. New York: Macmillan.

Amemiya, Takeshi. 1994. *Introduction to Statistics and Econometrics*. Cambridge, MA: Harvard University Press.

Amemyia, Takashi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

You will also use a software manual:

Greene, William H. 1995. *Limdep Version 7.0. User's Manual*. Bellport, NY: Econometric Software Inc.

Note: To distinguish Greene's textbook from the LIMDEP manual, below I refer to the first as "Greene" and to the second as "LIMDEP."

## 1. Introduction

Discrete and limited dependent variables. "Qualitative" variables in comparative research. Endogenous sampling and truncation. Selection as a problem in comparative research. Preview of the course.

### LIMDEP and its quirks

Ways of operating LIMDEP: editor and batch modes. Basic syntax. Data input and output. Controlling the sample. Transforming data.

*Do:* Study the ACLP Data Codebook. Browse through LIMDEP, Chapters 3-5 and 7.

*Do:* Calculate average rates of economic growth (G) for democracies and dictatorships, and then separately for presidential, mixed, and parliamentary democracies. Do the same for education (EDT). Run an OLS regression with G as the dependent variable and any list of independent variables but including REG (regime dummy).

## 2. Maximum Likelihood

Distribution of discrete and continuous variables. Density functions. Maximum likelihood. First order conditions. Second order conditions. Properties of MLE. Examples. Methods of maximization. Convergence. Problems.

King, Gary. 1989. *Unifying Political Methodology*. Cambridge: Cambridge University Press. Chapters 2 and 4.

Amemiya, Takeshi. 1994. *Introduction to Statistics and Econometrics*. Cambridge, MA: Harvard University Press. Sections 7.3 and 7.4.

Greene, Sections 4.5 and 5.5. (If you do not understand it, study Chapter 3).

or

Cramer, J.S. 1986. *Econometric applications of Maximum Likelihood methods*. Cambridge: Cambridge University Press. Chapters 2 and 5.

*Do:* Write the likelihood function for linear regression. Use LIMDEP's MAXIMIZE do what you did with REGRESS. Interpret the results. (You will need to read LIMDEP, Chapter 33).

*Do:* Redo the maximization using different optimization methods and convergence criteria (You will need to read LIMDEP, Chapter 8).

### **3. Binomial Probit and Logit**

Qualitative variables: binomial, multinomial, ordered, univariate, multivariate. Endogenous and exogenous justifications. Binomial Probit and Logit. MLE estimation. Marginal effects. Ordered probit. Bivariate probit. Batch file: PROBIT.LIM, ORDERED.LIM, BIVARIATE.LIM.

Kennedy, Chapter 15.

Greene, Sections 19.1 through 19.4.

LIMDEP, Sections 21.1 through 21.5, Chapters 22 and 23.

Do: Use PROBIT and LOGIT to predict political regimes. Try to get the best specification by the criterion of percentage of correct predictions. Interpret the results. Write the likelihood function for probit and use MAXIMIZE to estimate it. (Keep the model small; it takes a long time.)

### **4. Dynamic Probit**

Markov chains. Transition probabilities. Estimation. Batch file: DPROBIT.LIM.

Amemyia, Takashi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press. Sections 11.1.1 through 11.1.3.

### **5. Dynamic Bivariate Probit**

Batch file DBIVA.SEL

Przeworski, Adam and James R. Vreeland. 1998. "A Statistical Model of Bilateral Cooperation." Manuscript.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *The American Political Science Review* 93: 279.

### **6. Multinomial Logit**

Multinomial variables. Individual and group data. MLE estimation. Marginal effects. Independence of irrelevant alternatives. Different Logit models and the terminology. Batch files: LOGIT.LIM, LOGITIND.LIM, LOGITGRP.LIM.

Greene, Sections 19.6 and 19.7.

LIMDEP, Sections 24.1 and 24.2.

Do: Use LOGIT to predict democratic institutions (Presidential, Mixed, and Parliamentary: the variable is INST), first by using Dictatorships as the first choice and then without it. Interpret the results.

## **7. Tobit**

Truncated samples. Moments of truncated distributions. Tobit. MLE estimation.

Kennedy, references to TOBIT in Chapter 16.

Greene, Sections 20.1. through 20.3.

or

Amemyia, Takashi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press. Chapter 10.

LIMDEP, Sections 27.1. - 27.4.

Do: Use TOBIT to estimate economic growth under dictatorship, setting the values under democracy to zero. Interpret the results.

## **8. Selection: Identification**

The generic problem of selection. Endogenous sampling. Endogenous and exogenous selection. Terminology of models.

Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA.: Harvard University Press. Chapters 1, 2, and 6.

Pudney, Stephen. *Modelling Individual Choice*. Cambridge: Blackwell. Sections 2.1.1, 2.1.2, 2.4. and 2.5.1.

## **9. Selection based on Probit: Estimation**

The algebra of selection. Selection equations. Performance (outcome) equations. Varieties of estimation methods. Heckman (1976) OLS-IV estimator. MLE estimator. Panel methods. Batch files PROBIT.SEL, TESTSEL.SEL, PANEL.SEL.

Przeworski, Adam, and Fernando Limongi. 1992. "Selection, Counterfactuals, and Comparisons." Manuscript. University of Chicago.

OR

Przeworski, Adam and Fernando Limongi. 1997. "Modernization: Theories and Facts." *World Politics* 49. Pages 155-183.

Greene. Section 21.4.

LIMDEP, Sections 28.1, 28.2, 28.6.

Do: Estimate a PROBIT based SELECTION model of the growth of per capita income (G). Interpret the results.

### **10. Selection based on Logit: Applications**

Batch file LOGIT4.SEL.

LIMDEP, Section 28.3.3.

Alvarez, Mike, and Adam Przeworski. 1995. "Presidentialism and Parliamentarism: Which Works? Which Lasts?" Working Paper #7, Chicago Center on Democracy. University of Chicago.

Do: Estimate a LOGIT based SELECTION model of economic growth among different types of democracies. Interpret the results.

OR

### **10. Selection based on Dynamic Bivariate Probit: Applications**

Batch file DBIVA.SEL

Przeworski, Adam and James R. Vreeland. 1998. "The Effect of IMF Programs on Economic Growth." Forthcoming in *The Journal of Development Economics*.

### **11. Implications for Case Studies**

Do: Take any case study of two countries. Identify the most important causes of the difference that is being explained. Assume that these causes have been interchanged. Would you expect the consequences to follow?

### **12. Survival**

Survival model. Censoring. Non-parametric. Proportional hazards. Parametric models: exponential, Weibull, logistic. Estimation.

Kiefer, Nicholas M. 1988. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature* 56: 646-679.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press. Chapter 11.

LIMDEP, Chapters 48, 49, and 50.

Do: Run Life Tables, Cox's Proportional Hazards, and Exponential Parametric Model for the survival of democracies.

## Bibliography

- Achen, Christopher H. 1986. *The Statistical Analysis of Quasi-Experiments*. Berkeley: University of California Press.
- Amemiya, Takeshi. 1984. "Tobit models: a survey." *Journal of Econometrics* 24: 3-61.
- Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Cramer, J.S. 1986. *Econometric applications of Maximum Likelihood methods*. Cambridge: Cambridge University Press.
- Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43: 169-195.
- Goodman, Nelson. 1979. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press. Fourth Edition.
- Heckman, James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 46: 931-959
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-161.
- Heckman, James J. 1988. "The Microeconomic Evaluation of Social Programs and Economic Institutions," in *Chung-Hua Series of Lectures by Invited Eminent Economists, No. 14*. Taipei: The Institute of Economics, Academia Sinica.
- Heckman, James J. 1990. "Selection bias and self-selection." In *The New Palgrave. Econometrics*. Edited by John Eatwell, Murray Milgate and Peter Newman. New York: W.W. Norton.
- Johnson and Kotz. 1970. *Distributions in Statistics: Continuous Univariate Distributions, Vols. 1 and 2*. Boston: Houghton Mifflin.
- Judge, George S., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*. New York: John Wiley and Sons. Second Edition.
- Lee, L.F. 1978. "Unionism and wage rates: a simultaneous equations model with qualitative and limited dependent variables." *International Economic Review* 19: 415-433.
- Lewis, David. 1973. "Counterfactuals and Comparative Possibility." *Journal of Philosophical Logic* 2: 418-446.

Lijphart, Arend. 1975. "The Comparative-Cases Strategy in Comparative Research." *Comparative Political Studies* 8: 158-177.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: John Wiley and Sons.

Pudney, Stephen. 1989. *Modelling Individual Choice. The Econometrics of Corners, Kinks and Holes*. Cambridge: Blackwell.

Roy, A.D. 1951. "Some thoughts on the distribution of earnings." *Oxford Economic Papers* 3: 135-146.

Skocpol, Theda and Margaret Somers. 1980. "The Uses of Comparative History in Macrosocial Inquiry." *Comparative Studies in Society and History* 22: 174-197.