

ECONOMETRICS II

SPRING 2005

SUSTAINABLE DEVELOPMENT U9611
Thursday 4:10-6:00
404 International Affairs Building

PROFESSOR SHARYN O'HALLORAN
Email: so33@columbia.edu
OFFICE: 727 IAB
Phone: 854-3242

TEXTS: *The Statistical Sleuth*, Fred Ramsey and Daniel Schafer (R)
 Statistics with Stata, Version 8, Lawrence C. Hamilton (O)
 A Guide to Econometrics, Peter Kennedy (O)

WEB RESOURCES: *Student Solutions Manual*:
 <http://www.proaxis.com/%7Epanorama/home.htm>
 Stata Online Tutorial:
 <http://www.ats.ucla.edu/stat/stata/modules/default.htm>

Note: Required = (R), Optional = (O)

This course is the follow-up to Quantitative Analysis I for students in the Sustainable Development doctoral program. In many respects, it is a typical second-semester econometrics course; we will cover the linear regression model, introduced at the end of the previous course, in more detail, and extend regression analysis to some important application areas: time series analysis, qualitative dependent variables (variants of logit and probit models), and time-to-event studies, or survival models.

But in other respects, this course differs significantly from the way such a course is usually taught. In the past, statistics was presented, and often practiced, as a series of formulas, and the student's job was to know which formulas applied to which situations. But with advances in computer speeds, graphics, and software, statistics has become much more interactive. Researchers now spend much more time examining their data before running any regressions, and then they look carefully at the results to make sure that what they did makes sense.

The present course will reflect this new, integrated approach to data analysis, with emphasis placed on such topics as model diagnostics, data visualization, simulations, checking distributions, and model building. The point is to see statistics as a process of asking questions of the data, getting answers, and then using this information to ask further questions. It is not only more interesting that way; it is also closer to what practicing statisticians do when confronted with a data set.

Unfortunately, most textbooks are still written following the old paradigm, without much (or any) attention paid to the subjects listed above. One of the exceptions is our text, *The Statistical Sleuth*, which by and large does an excellent job of integrating graphical techniques, statistical formulas, and process-oriented explanations in building students' statistical repertoire.

As for course mechanics, the lecture schedule is below. There will be homework problem sets due most weeks, a midterm, and a final paper. The final paper should build off the data collection exercise completed in the previous semester, and use the analytical techniques developed in this course to explore a policy question. You should aim to make them about 15-20 pages long, plus any figures, tables and appendices. The grades will be based on your work as follows:

Homework:	35%
Midterm:	25%
Final Paper and Presentation:	35%
Participation:	5%

COURSE OUTLINE

January 20: Introduction: Data and Models

- Topics:
 - Statistical inference: How to lie with numbers?
 - Asking questions of data
 - Data management and statistical programs
- READINGS: Handouts

January 27: Assumptions of Linear Regression Model

- Topics:
 - Graphical tools of assessment
 - Transforming variables
 - Assessment of fit
- READINGS: Chapter 8 (skip 8.5) and Chapter 3 pp. 68-73

February 3: Multiple Regression

- *Assignment 1 due*
- Topics:
 - Interpreting output
 - Graphing results
 - Data exploration and presentation
- READINGS: Chapter 9

February 10: Inferential Tools for Multiple Regression

- *Assignment 2 Due*
- Topics:
 - Inference about coefficient estimates
 - Comparison of models
 - Goodness of fit
- READINGS: Chapter 10

February 17: Model Checking and Refinement

- *Assignment 3 Due*
- Topics:
 - Residual plots
 - Robust regression
 - Model refinement
- READINGS: Chapter 11, Chapter 3 pp. 65-68

February 24: Strategies for Variable Selection

- *Assignment 4 due*
- Topics:
 - The problem with too many variables
 - Variable selection techniques
 - Model selection
- READINGS: Chapter 12

March 3: Review: Multiple Regression in Practice

- *Assignment 5 Due*
- Topics:
 - Applications
 - Practical Solutions
 - Assessment and review
- READINGS: Handouts

March 10: Midterm

- *Assignment 6 Due*

March 17: Spring Break

March 24: Serial Correlation

- Topics:
 - Standard errors in correlated series
 - Auto regressive models AR(1)
 - Tests and diagnosis
- READINGS: Chapter 15

March 31: Logistical Models with Binary Data

- *Assignment 7 Due*
- Topics:
 - Generalized linear models
 - Estimation of logistical coefficients
 - Probit models
- READINGS: Chapter 20

April 7: Logistical Model with Multilevel Data

- *Assignment 8 Due*
- Topics:
 - Logistical regression for binomial responses
 - Model assessment
 - Ordered Logit and Probit models
 - Multinomial logit
 - Nested logit
- READINGS: Chapters 21.1-.4 and Handouts

April 14: Hazard Models

- *Assignment 9 Due*
- Topics:
 - Survival analysis
 - Estimating hazard rates
 - Cox Proportional Hazard Model
- READINGS: Handouts

April 21: Research Design and Model Testing: Student Presentations I

- *Assignment 10 Due*
- READINGS: Chapter 23

April 28: Research Design and Model Testing: Student Presentations II

Paper due May 5th 5:00PM