

# Economics 5010

**Instructor** Lee Adkins

**Phone** 744-8637

**Office** 303 Business

**Hours** 8:40-11:30 Wednesday

**E-mail** LADKINS@okstate.edu

**WWW** <http://spears.okstate.edu/~ladkins>

**Date** 3 August 2004. This supersedes any previously distributed edition of this syllabus.

## 1 Purpose

The purpose of this course is to prepare entering graduate students in business and economics to use regression analysis given a well-defined problem. Emphasis will be placed on your ability to understand when to adopt a particular model or technique, how to implement it, and how to interpret your results.

## 2 Textbooks

### Required

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, Addison-Wesley, 2004.

### Other Sources

There is an excellent website that is provided by the authors and publisher of your book. It can be found at

[http://wps.aw.com/aw\\_stockwatsn\\_economtrcs\\_1](http://wps.aw.com/aw_stockwatsn_economtrcs_1)

If you purchase a new textbook, the material on this site is free. If you buy a used book, the price is \$14.40. It includes answers to problems in the book, powerpoint slides, sample quizzes, data sets, tutorials, and other stuff. I think it is well worth purchasing if bought your book used.

### 3 Prerequisites

This course requires you to work with basic probability, statistics, algebra, and to use Stata. I will be teaching you a little matrix algebra and I will use a very small amount of calculus. You will not be asked to derive estimators using either of these tools. They are used in order for you to see where the estimators come from (as opposed to believing that they come from the ether). As prerequisites I recommend 2 courses in statistics in addition to a good command of algebra. You should have some notion about what random variables are, what a probability distribution is, what a statistic is, and what a hypothesis test is. These are things that we will cover, but we move through them quickly. If you haven't learned about these before you'll never be able to keep up. It is not necessary that you have any previous experience with linear regression, though this would be *very* helpful.

### 4 Course Outline

The following outline is based on Stock and Watson.

- 1 Economic Questions and Data
  - 1.1 Economic questions: empirical themes of the course
  - 1.2 Causal effects and idealized experiments
  - 1.3 Data: sources and types
- 2 Review of Probability
  - 2.1 Random variables and probability distributions
  - 2.2 Expected values, mean and variance
  - 2.3 Two random variables
  - 2.4 Normal,  $\chi^2$ ,  $F_{m,\infty}$ , and Student-t distributions
  - 2.5 Random sampling and the distribution of the sample average
  - 2.6 Large-sample approximations to sampling distributions

- 3 Review of Statistics
  - 3.1 Estimation of the population mean
  - 3.2 Hypothesis tests concerning population mean
  - 3.3 Confidence intervals for population mean
  - 3.4 Comparing means from 2 populations
  - 3.5 Scatter plots, sample covariance, and sample correlation
  - 3.6 Example: Earnings of Male and Female College Grads in the U.S.
- 4 Linear Regression with One Variable
  - 4.1 Linear regression model
  - 4.2 Estimating the coefficients of the linear regression model
  - 4.3 Least squares assumptions
  - 4.4 Sampling distribution of least squares
  - 4.5 Testing hypotheses about one coefficient
  - 4.6 Confidence intervals for a regression coefficient
  - 4.7 Regression when X is binary
  - 4.8  $R^2$  and the standard error of the regression
  - 4.9 Heteroscedasticity and homoscedasticity
- 5 Linear Regression with Multiple Regressors
  - 5.1 Omitted variable bias
  - 5.2 Multiple regression model
  - 5.3 The OLS estimator
  - 5.4 Least squares assumptions
  - 5.5 Sampling distribution of least squares
  - 5.6 Hypothesis tests and confidence intervals for a single coefficient
  - 5.7 Joint hypothesis tests
  - 5.8 Testing a single restriction of multiple coefficients
  - 5.9 Additional regression statistics
  - 5.10 Omitted variable bias reconsidered
- 6 Nonlinear Regression Functions
  - 6.1 A general strategy for modeling nonlinear regression functions
  - 6.2 Nonlinear functions of a single independent variable
  - 6.3 Interactions between independent variables
  - 6.4 RESET test for nonlinearity

- 6.5 Systematically varying parameter models
- 7 Assessing Studies Based on Multiple Regression
  - 7.1 Internal and external validity
  - 7.2 Threats to internal validity
    - 7.2.1 Omitted variable bias
    - 7.2.2 Misspecification of functional form
    - 7.2.3 Error-in-variables
    - 7.2.4 Sample selection
    - 7.2.5 Simultaneous causality
  - 7.3 Example: Test scores and class size
- 8 Regression with Panel Data
  - 8.1 Panel Data
  - 8.2 Panel data with two time periods: “before and after comparisons”
  - 8.3 Fixed effects regression
  - 8.4 Random effects regression
  - 8.5 Example: Drunk driving laws and traffic deaths
- 9 Regression with Binary Dependent Variables
  - 9.1 Binary dependent variables and the linear probability model
  - 9.2 Probit and logit regression
  - 9.3 Estimation and inference in logit and probit models
  - 9.4 Example: Mortgage denial using Boston’s HDMA data
  - 9.5 Other limited dependent variables models (time permitting)
    - 9.5.1 Censored and truncated regression
    - 9.5.2 Count data
    - 9.5.3 Ordered response
    - 9.5.4 Multiple discrete choice
- 10 Instrumental Variables Regression
  - 10.1 IV estimator with a single regressor and a single instrument (Two stage least squares)
  - 10.2 IV estimator with more than one regressor and multiple instruments
  - 10.3 Checking instrument validity
  - 10.4 Example: demand for cigarettes
- 11 Experiments and Quasi-Experiments

- 11.1 Idealized experiments and causal effects
- 11.2 Potential problems with experiments in practice
- 11.3 Regression estimators of causal effects using experimental data
- 11.4 Quasi-experiments
- 11.5 Problems with quasi-experiments
- 11.6 Experimental and quasi-experimental estimates in heterogenous populations

## 5 Software

The principle software used in the course is *Stata* (?). *Stata* is currently available in the CBA labs and on the CBA Trading Floor.

For those interested in what *Stata* can do, here is a link to a *Stata* brochure:

<http://www.stata.com/products/Stata8.pdf>

and to a brief list of *Stata*'s statistical capabilities

<http://www.stata.com/capabilities/statisticalcap.html>

For a comparison to SAS and SPSS visit:

[http://www.ats.ucla.edu/stat/stata/faq/compare\\_packages.htm](http://www.ats.ucla.edu/stat/stata/faq/compare_packages.htm)

## 6 Data sets

Here is a list and brief description of the data sets used in the course. This gives one an idea about the kinds of applications discussed. They are available for download from Stock and Watson's website as *Stata* data files, Microsoft Excel spreadsheets, and (at least for some) EViews data format.

**CPS data** This consists of 11130 observations on 3 variables from the current population survey.

**California Test Score** The California Standardized Testing and Reporting (STAR) data set contains data on test performance, school characteristics and student demographic backgrounds. The data used here are from all 420 K-6 and K-8 districts in California with data available for 1998 and 1999.

**Economics Journal Subscriptions** This data set contains data on 180 economics journals for the year 2000.

**Massachusetts Test Scores** The Massachusetts data are district-wide averages for public elementary school districts in 1998. The test score is taken from the Massachusetts Comprehensive Assessment System (MCAS) test, administered to all fourth graders in Massachusetts public schools in the spring of 1998.

**Auto Fatalities** The data are for the "lower 48" U.S. states (excluding Alaska and Hawaii), annually for 1982 through 1988.

**Boston Mortgage Applications** The Boston HMDA data set was collected by researchers at the Federal Reserve Bank of Boston. The data set combines information from mortgage applications and a follow-up survey of the banks and other lending institutions that received these mortgage applications. The data pertain to mortgage applications made in 1990 in the greater Boston metropolitan area. The full data set has 2925 observations, consisting of all mortgage applications by blacks and Hispanics plus a random sample of mortgage applications by whites.

**Cigarette** The data set consists of annual data for the 48 continental U.S. states from 1985 - 1995. The series include annual per capita cigarette sales in packs per fiscal year, the average retail cigarette price per pack, per capita income, the average cents per pack tax (which includes the sales tax), the tax applied to cigarettes only, and the cpi.

**Tennessee STAR experiments** The Project STAR public access data set contains data on test scores, treatment groups, and student and teacher characteristics for the four years of the experiment, from academic year 1985-86 to academic year 1988-89.

## 7 Exams

There will be 3 exams in the course. The final (exam 3) may include a take-home portion that will be due at the beginning of the period at the time of our regularly scheduled final exam.

All exams must be taken at the designated time. No make up exams will be given. If you miss an exam you will receive a grade of zero.

## 8 Grades

Your grade in this class will be based on your performance on 3 exams, and homework assignments.

Grades will carry the following weights and be measured according to the accompanying scale.

### Grade Weights

Exam 1	25%
Exam 2	25%
Exam 3	30%
Homework	20%

### Grades

92%–100%	A
78%–92%	B
65%–78%	C
50%–65%	D
< 50%	F

Note: Exam 3 may contain a take-home portion that will require you to do some work using Stata. This is one reason why it carries a bit more weight than the other exams.

## 9 Homework

There will be a significant amount of homework in the course. The best way to learn econometrics is to do econometrics. A large portion of your homework will require you to use a computer. The computer software we are using is Stata. Stata is a Windows program that operates under the Windows 2000 or Windows XP operating system on the microcomputers in the CBA lab.

I will not accept late homework under any circumstance. I expect homework to be legible and well organized. I encourage you to work with others in the class while doing homework, and may turn in assignments in groups of 2. The homework receives style points, so identical answers may receive different grades. I am predisposed to look favorably upon work that is well organized and legible.

Unless you are specifically told otherwise by me, all homework must be turned

in at the beginning of the class period on the date that it is due. Homework will not be accepted if late.

## 10 Attendance

Regular attendance is expected. You are responsible for any material you miss because of absence. In general, I do not permit students to copy my notes. If you miss class and need a copy of the notes, please obtain them from one of your classmates.

## 11 Cheating Policy

Cheating will not be tolerated. Any violation of the University's academic dishonesty policy will be prosecuted according to University regulations. You will receive a grade of 0 on any test or assignment you are caught cheating on. In addition, you are responsible for the security of your work (in other words, if someone copies your work, you will also receive a zero on the test or assignment).

**Econometrics is Fun!**