Econ 401: Cross Sectional Econometrics

Roger Klein; Fall 2017: MW 2:50-4:10, MU 210

Econometrics is concerned with formulating, estimating, and testing economic relationships. In this course, we will consider estimators and tests for models that differ from the linear model considered in Econ322. As an example of a standard linear model, assume that the log of wages is specified to depend linearly on such explanatory variables as education, experience,... In this case, the coefficient on education is termed the return to education and is constant. In particular, the return to education is assumed to be the same when you graduated high school as when you graduate Rutgers. Yet, it would seem that the return to education may not be the same for all education levels. In non-linear models, the return to education will not be constrained to be constant. Accordingly, such models may provide a better description of the behavior in which we are interested. We will consider such models in this course.

In other models of interest, the sample may be selected in some unusual manner that results in a problem termed sample selection. Testing and correcting for this problem will take us out of the linear model context. In still other examples, the dependent variable of interest may be only partially observed. For example, in a model that explains health status, health may only be observed as a categorical variable in that we observe whether an individual's health is high, average, or low. Again, we will find that a linear model will not be adequate for such data. Departures from the regular linear model will also be required when we seek to jointly explain several dependent variables of interest. For example, in studying murder rates, we might assume that such rates depend in part on some measures of police presence. On the other hand, police are not likely to be randomly assigned to areas and instead our measure of police presence may in turn depend on crime conditions in the area of interest.

OFFICE HOURS:

Office hours will be held in Rm. 311, NJ Hall, CAC and will be set by intersecting the times that you are free with times that I can hold office hours. If you have questions and can not come at the announced times, email me we will arrange another time. The key to doing well in this course is to ask questions both in and outside of class. My email address is

RKlein@economics.rutgers.edu

Since each topic covered depends on previously covered material, I strongly encourage you to **ask questions** as we go along. I would also encourage you to ask any questions that you have about weekly problem sets **before** they are due. To do well in this course, you must be willing to ask questions about anything that you do not understand, and in turn I will be very accessible.

TEXT:

There is no required text for this course. The material described below will be covered in posted handouts Often this class is somewhat heterogeneous, with some students wanting to see more complete proofs than I think are appropriate in class. When the demand exists, I will then post additional and optional handouts.

GOALS AND ASSESSMENT:

Weekly Problem Sets	25%
Interm Exam	15%
Midterm Exam	35%
Course Paper	25%

WEEKLY PROBLEM SETS:

Weekly problem sets will be given that will usually be computer intensive and will involve estimating and/or testing various models of interest. We will be using a software language, Gauss, and doing some programming in

this language. However, no prior experience with this package is assumed. We will go over in class and cover in handouts all material that you will need to know. While a few students typically become interested in writing programs, programming will not be emphasized in this course. You will be given some exposure to programs so that you will have some understanding as to how some of the calculations are actually made. In assigning grades, I will drop the weekly problem set with the lowest score.

You will receive instructions for downloading the software that we will be using in the course: Gauss (likely version 17) and Maxlik 5.0.

You may work together on these weekly problem sets, but each of you should turn in your own write-up. Make sure you understand how to do these problems as approximately 80% of each exam will contain similar problems. You may and are strongly encouraged to ask questions about these weekly problem sets before you turn them in.

EXAMS:

Approximately 80% of the exams will consist of problems similar to those in the weekly problem sets. Accordingly, most of the questions should not be surprising if you have completed and understand the weekly problem sets. Approximately 20% of these problem sets may involve extensions not directly covered in previous problems. Past exams and sample problems will be distributed as a way of studying for the exams. I will offer to have review sessions prior to each exam to go over any questions on such review problems.

PAPER: TO BE DISCUSSED

ATTENDANCE AND MISSED ASSIGNMENTS:

Attendance will not be taken. However, most of the material that we cover will be explained in class and will not be available elsewhere (other than handouts). Consequently, regular attendance will be necessary to do well in this course. If you do miss a class, first get notes from someone else in class and then come to see me to fill in any gaps. As for missed assignments and/or exams, make-ups will usually not be given. Without a "satisfactory" reason for missing an exam or an assignment, these will count as 0 grades .

COURSE OUTLINE

The following outline gives the topics that we will cover and the approximate dates for each topic. More or less time may be required than as is indicated for several of the listed topics. Accordingly, you should view the dates listed below as tentative.

- **Course Introduction; Review of Expectations (Population Means) and Sample Means:** Sept. 6, 11, 13

In econometrics, we are typically interested in conditional expectations. Viewing an expectation as an average in the relevant population, we may be concerned with the average value for wages in the population of individuals with 12 years of education and 6 years of experience. Such a population average is termed a conditional expectation in that it is conditioned on those individuals with 12 years of education and 6 years of experience. A sample mean, provides one estimator for such an expectation in that it is "likely" to be close to the corresponding population mean (expectation).

- **Partially Linear Models:** Sept. 18, 20, 25

In a linear model, the impact of each explanatory variable on the dependent variable is constant. However, there are instances where we might believe that the impact is not constant. As a direct generalization of a linear model, a partially linear model addresses this concern.

- **Binary Response Models:** Sept. 27, Oct. 2, 4

In many applications (e.g. labor force participation or the decision of whether or not to purchase health insurance), the dependent variable is limited to a few values. In the binary case, the dependent variable takes on one of two variables (e.g. one if the individual joins the labor force and zero otherwise). Here, we will describe and implement an alternative estimation method termed maximum likelihood.

Review Session: TBA
Interm Exam: Oct. 9

- **Categorical Models:** Oct. 11, 16, 18, 23, 25

In categorical models, the dependent variable takes on a small number of distinct values. For example, health may be reported as being poor, average, or good. As a very different example, bonds are assigned such categorical ratings as AAA, AA, etc. These categorical models are a generalization of the binary case above.

- **Endogenous Variable Bias: The Linear Case** Oct. 25, 30 Nov. 1, 6, 8

In estimating a model in which several dependent variables are jointly determined . For example, the crime rate in an area my depend on some measure of police presence. However, police are not likely to be randomly assigned to areas but rather to areas with high crime rates. In such cases, the OLS estimator is not appropriate, and we will employ an alternative estimators: instrumental variables, two-stage least squares, or control methods.

- **Endogenous Treatment Models:** Nov. 8, 13, 15

To illustrate this type of nonlinear model, assume that individuals must decide whether or not to participate in a job training program (the "treatment"– discrete choice component). Further suppose that the log-wage (another equation) depends on whether or not the individual participates in a job training program. This two equation model combines aspects of simultaneous equations and discrete choice equations. We will discuss how to estimate an endogenous treatment model when the treatment effect is constant and when it depends on other variables. An often employed method, two-stage-least-squares, will not be appropriate in this nonlinear context.

Review Session: TBA
Midterm Exam: Nov. 20

Thanksgiving Break

- **Censored Regression (Tobit)** Nov. 27, 29

In some models, the dependent variable may take on the same value for many individuals and be continuous otherwise. . For example, with survey data, many individuals may report a demand of zero for the product, while others report a continuum of positive demands.

- **Sample Selection**: Dec. 4, 6

Frequently we are interested in estimating an equation of interest with a relatively simple structure when we have a random sample from the entire population. For example, the equation may be the same type of linear model previously studied. However, rather than having a random sample from the entire population, the sample may be selected so that we only have data on "unusual individuals". Employing the selected sample, in this section we discuss the nature of the sample selection problem and methods for dealing with it.

- **Panel Data and Individual Specific Effects,** Dec. 11, 13

Some data may consist of repeated observations on individuals over time. In this case it is possible to allow for "individual specific" effects that are not observed and that may or may not be related to explanatory variables. We will discuss these effects and a test for whether or not they are related to the explanatory variables.