AFRE 835: Introductory Econometrics

Ch. 1: Introduction

Spring 2017

The Basics

- Instructor: Prof. Joseph Herriges
  - Office: 220C Marshall-Adams Hall
  - Office Hours: MW 9-10 am or by appointment
  - email address: jah@msu.edu
- Class Time: MW 10:20-11:40 am
- Class Location: 48 Ag Hall
- Textbook: Wooldridge, Jeffrey M. 2011. *Introductory Econometrics: A Modern Approach, Fifth Edition*. South-Western College publishing Company.
- Software package: Stata
- Class information is available through Desire2Learn

About Me

- B.S. from Marquette University in Economics and Math/Stat
- M.S. and Ph.D. for the University of Wisconsin, Madison in Economics (Major field: Econometrics)
- 10 years work for Laurits R. Christensen, Inc. - Economic consulting in electric power industry
- 26 years at Iowa State University, specializing in
  - Applied Econometrics
  - Environmental and Natural Resource Economics
- This is my third year at MSU

Objectives

- This is a masters-level econometrics course.
- The purpose of the course is to provide students with a core set of empirical skills to test economic theories and to address real-world economic problems.
- Much of the underlying statistical theory and estimation techniques will be covered.
- However, the emphasis of the course will be placed on the empirical applications with a focus on model specification, estimation, testing, and interpretation of results.
- Applications are drawn from
  - Environmental Economics
  - Agricultural Economics
  - Labor
  - Transportation
  - Health
  - Education
  - Energy Economics

Tentative Lecture Schedule - Part I

Ch.	Topic	Lectures	Date
1	Introduction	1	1/9
1A.	Review of Probability, Statistics, and Matrix Alg.	1-3	1/9-1/18
2.	Simple Linear Regression	4	1/23
2A.	Stata Review Session	5	1/25
3.	Multiple Regression Analysis: Estimation	6	1/30
4.	Multiple Regression Analysis: Inference	7	2/1
5.	Multiple Regression Analysis: OLS Asymptotics	8	2/6
6.	Multiple Regression Analysis: Further Issues	8	2/6
7.	Multiple Regression Analysis: Binary Variables	9	2/8
8.	Heteroskedasticity	10-11	2/13-2/15
9.	Miscellaneous Issues	12	2/20
10.	Regression Analysis with Time Series Data	13	2/22
	Review for Midterm	14	2/27
	Midterm Exam	15	3/1

Tentative Lecture Schedule - Part II

Ch.	Topic	Lectures	Date
11	Further Issues with Time Series Data	16	3/13
12	Serial Correlation and Heteroskedasticity	17	3/15
13	Simple Panel Data Methods	18	3/20
14	Advanced Panel Data Methods	19	3/22
15	Instrumental Variables and 2SLS	20-21	3/27-3/29
16	Simultaneous Equation Models	22-23	4/3-4/5
17	Limited Dependent Variable Models	24-25	4/10-4/12
18	Advanced Time Series Topics	26-27	4/17-4/19
19	Program Evaluation (DinD and RD)	28	4/24
	Review for Final	29	4/26
	Final Exam	10am-noon	5/3

Grade Composition

- Midterm exam: 30%
- Final exam: 45%
- Homework: 25% (6 total, top 5 count)
  - Full credit if turned in *during class* on due date
  - Half credit if turned in late, but prior to next class

Other Matters

- Academic Integrity
- Accommodations
- Attendence
- Electronic Devices
- E-mail
- Emergencies (www.altert.msu.edu)
- Missed exams

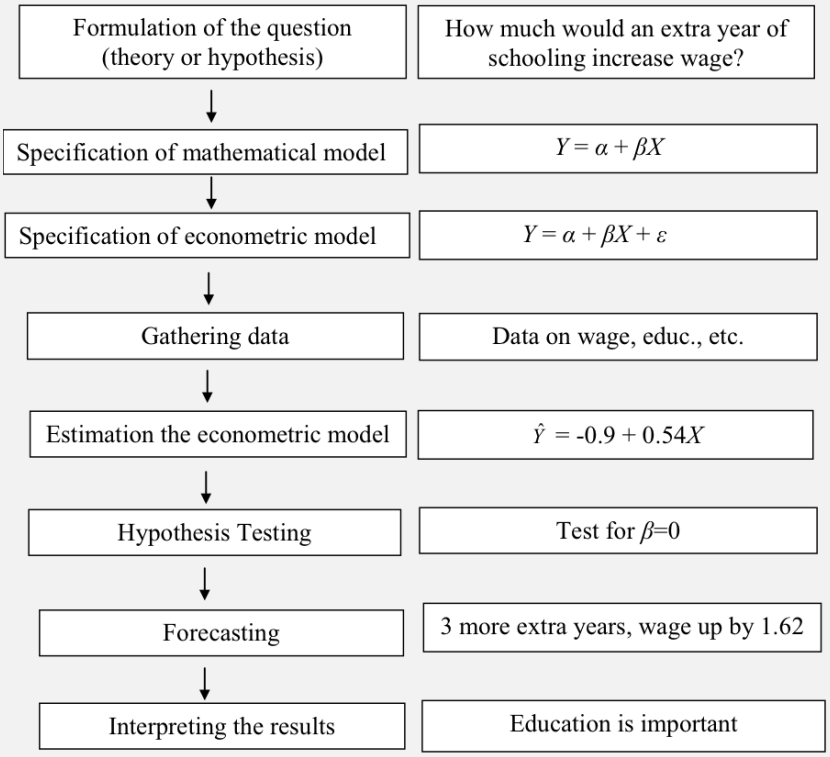
What is Econometrics?

- "Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy" (Woodridge 2009).
- "Econometric may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference" (Samuelson 1954).
- "Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to analysis of economic phenomena" (Theil, 1971).
- "At a broad level, econometrics is the science and art of using economic theory and statistical techniques to analyze economic data" (Stock and Watson 2007)

What Distinguishes Econometrics as a Discipline

- *Economic theory* provides qualitative relationships among economic variables
  - . . . though it is rarely informative about functional forms or definitive about the sign or even magnitude of effect.
- *Economic statistics* focuses on collecting, processing and presenting data in the form of charts and tables, descriptive statistics rather than inferential statistics.
- *Mathematical statistics* foundation for theoretical econometrics and the analysis experimental data
  - Economists rarely have available experimental data (though there are exceptions);
  - Experiments are often prohibitively expensive or morally repugnant;
  - Instead, we rely on *nonexperimental* or *observational* data;
  - Sometimes *natural experiments* emerge, approximating experimental conditions.

Steps in Economic Analysis



Data Types

There are four broad types of data we will typically encounter in empirical research:

1. Cross-Sectional
2. Time series
3. Pooled Cross-Sections
4. Panel (or longitudinal)

Data Type 1:Cross-Sectional

- Data available for different economic agents (e.g. rural households, farmers, firms, etc.) at one particular time period
- Typically, it is assumed that the sample is drawn randomly from the underlying population. . . though this need not be the case.

Cross-Sectional Example

A cross-section rural household survey (demographic portion)

Household no.	Head's age	Head's educ.	Head's main occupation	Household size
1	24	6	1	4
2	53	1	1	7
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
1000	36	4	2	4

Data Type 2:Time Series

- Data available for many different time periods for the same economic agent (e.g. national GDP or unemployment rate, or inflation rate for U.S. from 1975-2005).
- Typically more challenging to model than cross section data as observations are rarely independent across time.
- More complex techniques are required to address
  - serial correlation
  - seasonality
  - trends
- The nature of these problems can change with the frequency of the data.

Time Series Example

GDP, unemployment, and inflation rate for U.S. from 1972-2007

Observation	Total GDP	Unemp. Rate	Infl. Rate
1975			
1976			
.			
.			
.			
2005			

Data Type 3: Pooled Cross-Sections

- Data having both time series and cross-sectional features;
- Combines cross-sections gathered in distinct time periods, but not for the same agents (except by chance);
- The advantage over a single cross-section is that one can examine changes over time.
- This can be particularly useful in studying the impact of policy changes.

Pooled Cross-Section Example: American Time Use Survey

		Time per day Spent		
Individual	Year	Working	Sleeping	Recreating
1	1990	8	7	3
⋮	⋮	⋮	⋮	⋮
500	1990	8.5	9	1
501	2000	12	6	0.5
⋮	⋮	⋮	⋮	⋮
1000	2000	10	6.5	1.5
1001	2000	9	9	2
⋮	⋮	⋮	⋮	⋮
1500	2010	8.5	9.5	4

Data Type 4: Panel (Longitudinal) Data

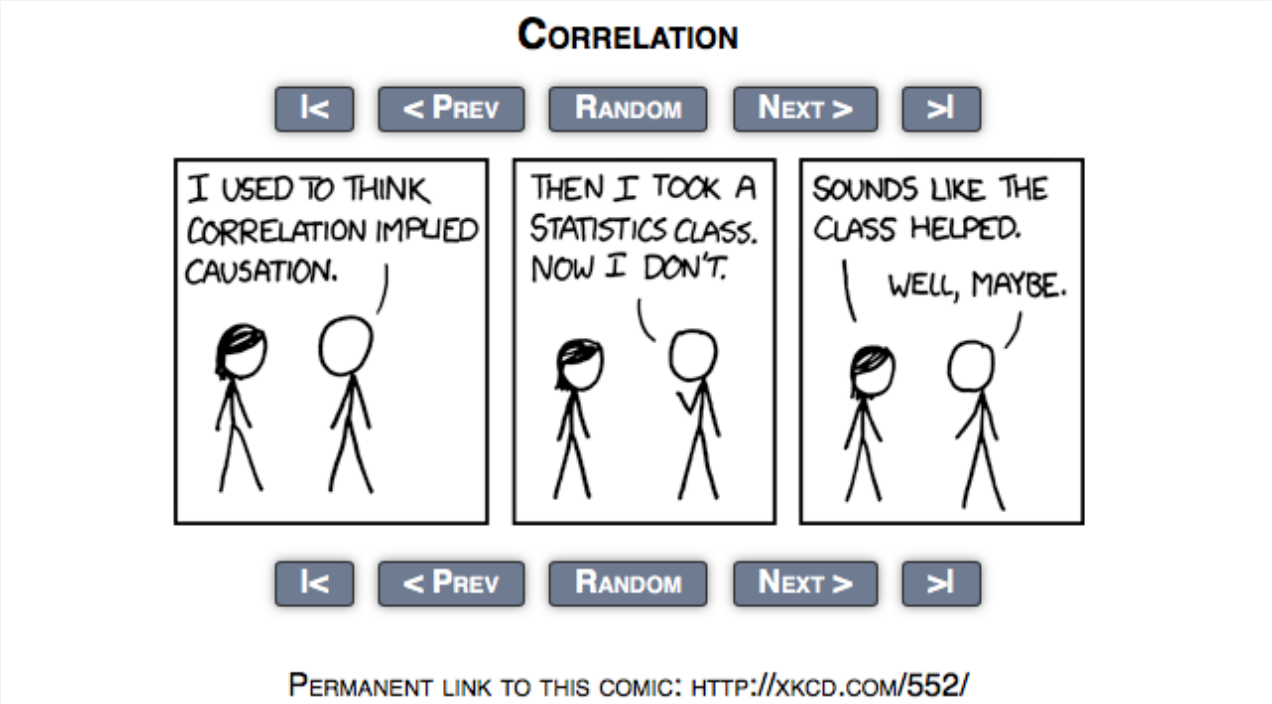
- Data available for a cross-section of many agents across multiple years.
- It is essentially a time series of cross sections.
- Unlike pooled cross-sections, panel data allows us to track how an individual agent's behavior changes over time.
- Panel data is more costly to collect and can suffer from attrition over time
  - making the sample no longer random and
  - complicating the analysis.

Panel Data Example

A panel rural household survey data (demographic portion)

Household no.	Year	Head's age	Head's educ.	Head's main occupation	Household size
1	1995	24	6	1	4
1	2000	29	6	1	5
1	2005	34	6	2	5
2	1995	53	1	2	7
2	2000	58	1	2	8
2	2005	63	1	2	7
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
1000	1995	36	4	2	4
1000	2000	41	4	2	4
1000	2005	46	4	1	6

Causality versus Correlation



Causality versus Correlation

- We will typically be interested in measuring the *causal impact* of one variable (e.g., a job training program) on an outcome variable of interest (e.g., wages);
- It is important to distinguish this from simply finding a relationship (correlation) between two variables;
- What we want to measure is the impact of one variable on another, *ceteris paribus*; i.e., holding everything else constant.
  - It only makes sense when we say X cause Y when everything else is held constant.
  - For example, the Engel's law which states that the proportion of income spent on food consumption declines as income increases will hold only if everything else is held constant.

Examples

- Effects of fertilizer on crop yield;
- The returns to education;
- The effect of a change in minimum wages on employment;
- The effect of law enforcement on city crime levels;
- The impact of hospital visits on health outcomes;
- The impact of a voluntary time-of-use rate on a household's pattern of electricity usage.

Defining Causality

- We can define what we mean by causality by using Rubin's notion of potential outcomes;
- Suppose we are interested in the *causal impact* of a change (e.g., an additional year of education) on an outcome of interest (e.g., the individual's wages);
- For an individual, let
  - Y_{i1} denote the outcome with the change and
  - Y_{i0} denote the outcome without the change.
- The causal impact of the change for that individual is then $Y_{i1} - Y_{i0}$.
- The problem is that we only observe one of these *potential outcomes*;
- The practical problem is finding a way to *fill-in* the missing information.
- Randomized experiments provide one approach.
- Natural experiments provide another.