Economics 326
Methods of Empirical Research in Economics

Lecture 1: Introduction

Hiro Kasahara
University of British Columbia

January 4, 2010

What is Econometrics?

Econometrics is concerned with the development of statistical methods for:

- Estimating and Testing of economic theories.
- Forecasting of important economic variables.
- Evaluation of government and business policy. For example,
  - The effect of political campaign expenditures on voting outcomes.
  - The effect of school spending on student performance.
  - The effect of R&D subsidy on firm's productivity.
- Nonexperimental data vs. Experimantal data.

Empirical Economic Analysis

- Economic theory is used to construct models characterizing relationships between variables of interest.
- However, economic models are only approximations .
- A model can take into account a number of important factors, but there will be many factors left out that also affect outcomes.
- We therefore replace the exact (deterministic) model with a probabilistic model.

Example 1.1: Becker's (1968) economic model of crime

Economic Model: the reward from property crime vs. the cost

$$y = f(x_1, x_2, x_3, x_4, ...)$$

y = hours spent in criminal activities (crime)
x_1 = "wage" for an hour spent in criminal activities
x_2 = hourly wage in legal employment ($wage_m$)
x_3 = income other than from crime or employment (otherinc)
x_4 = probability of getting caught (freqarr)

Econometric Model:

$$\begin{aligned} crime &= \beta_0 + \beta_1 wage_m + \beta_2 othinc + \beta_3 freqarr + \\ &\quad \beta_4 freqconv + \beta_5 avgsen + \beta_6 age + u \end{aligned}$$

The term "u" captures unobserved factors:

(1) the reward for criminal activity, (2) family background, (3) measurement error

Example 1.2: Mincer's (1974) wage regression

Economic Model: wage depends on one's human capital

$$wage = f(educ, exper, training)$$

Econometric Model:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 (exper)^2 + \beta_4 training + u$$

The term "u" captures unobserved factors:

(1) innate ability, (2) family background, (3) quality of education

Hypothesis Testing: whether the training affects wage or not

$$H_0 : \beta_4 = 0$$

Types of data: cross-section

- A cross-sectional data set consists of observations on individuals such as workers or firms collected in a single period of time.
- Example: A cross-sectional data set on wages and other individual characteristics (Table 1.1, Page 7):

obs number	wage	education	experience	female	married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
⋮	⋮	⋮	⋮	⋮	⋮

- The order of observations is not important.
- Random Sampling: we assume that the observations are statistically independent. (See Appendix C.1)

Types of data: time series

- A time series data set consists of observation on several variables over time.
- Example: Minimum wage, unemployment, and related data for Puerto Rico (Table 1.3, Page 9):

obs number	year	minimum wage	unemployment	gnp
1	1950	0.20	15.4	878.7
2	1951	0.21	16.0	925.0
3	1952	0.23	14.8	1015.9
⋮	⋮	⋮	⋮	⋮

- The frequency at which the data is collected can be daily, weekly, monthly, quarterly, and annually. In Finance, high frequency trade data.
- The order of observations is important.
- Observations are often correlated; trends.

Types of data: panel

- A panel data set consists of a time series for each cross-sectional member.
- Example: A two-year panel data set on city crime statistics (Table 1.5, Page 11):

obs numb	city	year	murders	population	unempl	police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Causality

- While we are interested in causal relations, statistics allows us to establish correlations (associations) in the data.
- In order to say that one variable has a causal effect on another, other factors affecting the outcome must be held fixed (controlled for).
- The notion of cetris paribus— "other (relevant) factors being equal"
- In natural sciences they use controlled experiments.
- Experiment are often impossible in economics (too costly and/or for ethical reasons).
- We do not observe everything.

Examples

- **Effects of Fertilizer on Crop Yield (Griliches, 1957):**

$$CropYeild = \beta_0 + \beta_1 \times Fertilizer + u$$

How are fertilizer amounts are chosen? Is it possible that the better the land quality, the more fertilizers are chosen? Maybe, random experiment might be possible.

Examples

- **Effects of Fertilizer on Crop Yield (Griliches, 1957):**

$$CropYeild = \beta_0 + \beta_1 \times Fertilizer + u$$

How are fertilizer amounts are chosen? Is it possible that the better the land quality, the more fertilizers are chosen? Maybe, random experiment might be possible.

- **Education:**

$$\log(wage) = \beta_0 + \beta_1 \times \text{Years of Schooling} + u,$$

u = other factors, for example, ability. Since it is very hard to control for ability, one can overestimate the return to education by relying on usual correlations.

Examples (continued)

- **Size of the police force and crime:**

$$\#Crime = \beta_0 + \beta_1 \times \text{Size of the Police Force} + u.$$

Usually, cities with a lot of criminal activity have a bigger police force. Simple correlations can spuriously indicate a positive effect of police force on the crime.

Examples (continued)

- **Size of the police force and crime:**

$$\#Crime = \beta_0 + \beta_1 \times \text{Size of the Police Force} + u.$$

Usually, cities with a lot of criminal activity have a bigger police force. Simple correlations can spuriously indicate a positive effect of police force on the crime.

- **The Effect of the Minimum Wage on Unemployment:**

$$Unemployment = \beta_0 + \beta_1 \times \text{Minium Wage} + u.$$

Reverse causality: High employment may lead to political pressure for higher minimum wage.