

Economics 326
Methods of Empirical Research in Economics

Lecture 1: Introduction

Vadim Marmer
University of British Columbia

February 3, 2009

What is Econometrics?

Econometrics is concerned with the development of statistical methods for:

- ▶ Estimation of economic relationships.
- ▶ Testing of economic theories.
- ▶ Forecasting of important economic variables.
- ▶ Evaluation of government and business policy.

Why statistics?

- ▶ 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.

Examples

- ▶ Estimation of demand and supply functions. Elasticity of demand/supply can be used to evaluate the effect of taxation
- ▶ Mincer, J., *Schooling, Experience, and Earnings*, 1974. Estimation of return to schooling and experience using individual census data.
 - Used to determine the optimal amount of schooling.
 - Study education in developing countries.
 - Study gender and race discrimination.
 - Study the impact of immigration on labour markets.
- ▶ Paarsch, H. J., *Journal of Econometrics*, 1997. Estimation of optimal reserve price for BC timber auctions.
- ▶ Chandra et al, *Pediatrics*, 2008. The effect of exposure to sexual content on television on likelihood of teen pregnancy.

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.
- ▶ It is usually natural to assume that the observations are statistically independent.

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).
- ▶ In natural sciences they use controlled experiments.
- ▶ Experiment are often impossible in economics (too costly and/or for ethical reasons).
- ▶ Observational data.

Examples

► **Education:**

$$\log(\text{Wage}) = \alpha + \beta \times \text{Years of Schooling} + U,$$

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.

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

$$\text{Number of Crimes} = \alpha + \beta \times \text{Size of the Police Force} + U.$$

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

Examples (continued)

► **Sex on TV and teen pregnancy:**

$$\begin{aligned} & \text{Prob of teen pregnancy=} \\ & = F(\beta_0 + \beta_1 \times \text{Exposure to Sex on TV} + \beta_2 \times \text{Total TV}). \end{aligned}$$

Self-selection : it is hard to control for interest in sex, no causal relationship.