The Nature of Econometrics and Economic Data

Chapter 1

Wooldridge: Introductory Econometrics: A Modern Approach, 5e

What is econometrics?
- Econometrics = use of statistical methods to analyze economic data
- Econometricians typically analyze nonexperimental data

Typical goals of econometric analysis
- Estimating relationships between economic variables
- Testing economic theories and hypotheses
- Forecasting economic variables
- Evaluating and implementing government and business policy

Steps in econometric analysis
- 1) Economic model (this step is often skipped)
- 2) Econometric model

Economic models
- Maybe micro- or macromodels
- Often use optimizing behaviour, equilibrium modeling, ...
- Establish relationships between economic variables
- Examples: demand equations, pricing equations, ...

The Nature of Econometrics and Economic Data

Economic model of crime (Becker (1968))
- Derives equation for criminal activity based on utility maximization

\[ y = f(x_1, x_2, x_3, x_4, x_5) \]

"Wage" of criminal activities
- Wage for legal employment
- Other income
- Probability of getting caught
- Probability of conviction if caught
- Expected sentence

Steps in econometric analysis
- Functional form of relationship not specified
- Equation could have been postulated without economic modeling

The Nature of Econometrics and Economic Data

Model of job training and worker productivity
- What is effect of additional training on worker productivity?
- Formal economic theory not really needed to derive equation:

\[ \text{Hourly wage} = f(\text{educ}, \text{exp}, \text{training}) \]

Other factors may be relevant, but these are the most important (?)

The Nature of Econometrics and Economic Data

Econometric model of criminal activity
- The functional form has to be specified
- Variables may have to be approximated by other quantities

\[ \text{crime} = \beta_0 + \beta_1 \text{wage}_{edu} + \beta_2 \text{income} + \beta_3 \text{freq} + \beta_4 \text{age} + \beta_5 \text{years convicted} + \epsilon \]

Unobserved determinants of criminal activity
- e.g. moral character, family background
Econometric model of job training and worker productivity

\[ \text{wage} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{experience} + \beta_3 \text{training} + \varepsilon \]

Most of econometrics deals with the specification of the error term. Econometric models may be used for hypothesis testing. For example, the parameter $\beta_3$ represents effect of training on wage. How large is this effect? Is it different from zero?

Cross-sectional data set on wages and other characteristics

<table>
<thead>
<tr>
<th>observ</th>
<th>wage</th>
<th>educ</th>
<th>gender</th>
<th>married</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.50</td>
<td>11</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3.24</td>
<td>12</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3.00</td>
<td>11</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4.00</td>
<td>0</td>
<td>44</td>
<td>0</td>
</tr>
</tbody>
</table>

Observation number: Hourly wage

Cross-sectional data set on growth rates and country characteristics

<table>
<thead>
<tr>
<th>observ</th>
<th>country</th>
<th>growth</th>
<th>government consumption</th>
<th>Adult secondary education</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Argentina</td>
<td>0.09</td>
<td>165</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>Austria</td>
<td>3.32</td>
<td>165</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>Belgium</td>
<td>3.34</td>
<td>165</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>Belgium</td>
<td>3.34</td>
<td>155</td>
<td>100</td>
</tr>
</tbody>
</table>

Observation year: Growth rate of real per capita GDP, Government consumption as percentage of GDP, Adult secondary education rates

Cross-sectional data sets

- Sample of individuals, households, firms, cities, states, countries, or other units of interest at a given point in time/in a given period
- Cross-sectional observations are more or less independent
- For example, pure random sampling from a population
- Sometimes pure random sampling is violated, e.g. units refuse to respond in surveys, or if sampling is characterized by clustering
- Cross-sectional data typically encountered in applied microeconomics

Observations of a variable or several variables over time

- For example, stock prices, money supply, consumer price index, gross domestic product, annual homicide rates, automobile sales, ...
- Time series observations are typically serially correlated
- Ordering of observations conveys important information
- Data frequency: daily, weekly, monthly, quarterly, annually, ...
- Typical features of time series: trends and seasonality
- Typical applications: applied macroeconomics and finance

Econometric analysis requires data

- Different kinds of economic data sets
  - Cross-sectional data
  - Time series data
  - Pooled cross sections
  - Panel/Longitudinal data

Econometric methods depend on the nature of the data used

- Use of inappropriate methods may lead to misleading results

The Nature of Econometrics and Economic Data

- Cross-sectional data set on wages and other characteristics
- Time series data set on wages and other characteristics
- Cross-sectional data set on growth rates and country characteristics
- Cross-sectional data sets
The Nature of Econometrics and Economic Data

Time series data on minimum wages and related variables

Table 1.3: Minimum Wages, Unemployment, and Related Data for Puerto Rico

<table>
<thead>
<tr>
<th>year</th>
<th>wages</th>
<th>coverage</th>
<th>coverage rate</th>
<th>GNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>0.09</td>
<td>20.3</td>
<td>15.4</td>
<td>876.7</td>
</tr>
<tr>
<td>1951</td>
<td>0.13</td>
<td>29.7</td>
<td>16.0</td>
<td>905.6</td>
</tr>
<tr>
<td>1952</td>
<td>0.21</td>
<td>28.7</td>
<td>16.3</td>
<td>879.7</td>
</tr>
<tr>
<td>1966</td>
<td>0.35</td>
<td>58.1</td>
<td>16.9</td>
<td>4298.8</td>
</tr>
<tr>
<td>1967</td>
<td>0.32</td>
<td>58.2</td>
<td>16.8</td>
<td>4496.7</td>
</tr>
</tbody>
</table>

Average minimum wage for given year
Average coverage rate
Unemployment rate
Gross national product

Panel or longitudinal data

- The same cross-sectional units are followed over time
- Panel data have a cross-sectional and a time series dimension
- Panel data can be used to account for time-invariant unobservables
- Panel data can be used to model lagged responses
- Example:
  - City crime statistics; each city is observed in two years
  - Time-invariant unobserved city characteristics may be modeled
  - Effect of policy changes may exhibit time lag

Pooled cross sections

- Two or more cross sections are combined in one data set
- Cross sections are drawn independently of each other
- Pooled cross sections are often used to evaluate policy changes
- Example:
  - Evaluate effect of change in property taxes on house prices
  - Random sample of house prices for the year 1993
  - A new random sample of house prices for the year 1995

Pooled cross sections on housing prices

Table 1.4: Pooled Cross Sections: Two Years of Housing Prices

<table>
<thead>
<tr>
<th>year</th>
<th>price</th>
<th>property</th>
<th>police</th>
<th>city</th>
<th>area in square feet</th>
<th>number of bathrooms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>0.95</td>
<td>1.2</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1990</td>
<td>0.96</td>
<td>1.3</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1986</td>
<td>0.95</td>
<td>1.2</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1990</td>
<td>0.96</td>
<td>1.3</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1986</td>
<td>0.95</td>
<td>1.2</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1990</td>
<td>0.96</td>
<td>1.3</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1986</td>
<td>0.95</td>
<td>1.2</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
<tr>
<td>1990</td>
<td>0.96</td>
<td>1.3</td>
<td>1.5</td>
<td>1.8</td>
<td>125</td>
<td>3</td>
</tr>
</tbody>
</table>

Causality and the notion of ceteris paribus

Definition of causal effect of $x$ on $y$:

"How does variable $y$ change if variable $x$ is changed but all other relevant factors are held constant?"

- Most economic questions are ceteris paribus questions
- It is important to define which causal effect one is interested in
- It is useful to describe how an experiment would have to be designed to infer the causal effect in question
The Nature of Econometrics and Economic Data

- **Causal effect of fertilizer on crop yield**
  - “By how much will the production of soybeans increase if one increases the amount of fertilizer applied to the ground?”
  - Implicit assumption: all other factors that influence crop yield such as quality of land, rainfall, presence of parasites etc. are held fixed

- **Experiment:**
  - Choose several one-acre plots of land; randomly assign different amounts of fertilizer to the different plots; compare yields
  - Experiment works because amount of fertilizer applied is unrelated to other factors influencing crop yields

The Nature of Econometrics and Economic Data

- **Effect of the minimum wage on unemployment**
  - “By how much (if at all) will unemployment increase if the minimum wage is increased by a certain amount (holding other things fixed)?”

- **Experiment:**
  - Government randomly chooses minimum wage each year and observes unemployment outcomes
  - Experiment will work because level of minimum wage is unrelated to other factors determining unemployment
  - In reality, the level of the minimum wage will depend on political and economic factors that also influence unemployment

The Nature of Econometrics and Economic Data

- **Measuring the return to education**
  - “If a person is chosen from the population and given another year of education, by how much will his or her wage increase?”
  - Implicit assumption: all other factors that influence wages such as experience, family background, intelligence etc. are held fixed

- **Experiment:**
  - Choose a group of people; randomly assign different amounts of education to them (infeasable!); compare wage outcomes
  - Problem without random assignment: amount of education is related to other factors that influence wages (e.g., intelligence)

The Nature of Econometrics and Economic Data

- **Testing predictions of economic theories**
  - Economic theories are not always stated in terms of causal effects
  - For example, the expectations hypothesis states that long term interest rates equal compounded expected short term interest rates

\[
(1+r_{long}) = (1+r_{short1})(1+r_{short2})\ldots(1+r_{shortm})
\]

  - An implication is that the interest rate of a three-months T-bill should be equal to the expected interest rate for the first three months of a six-months T-bill; this can be tested using econometric methods

The Nature of Econometrics and Economic Data

- **Effect of law enforcement on city crime level**
  - “If a city is randomly chosen and given ten additional police officers, by how much would its crime rate fall?”
  - Alternatively: “If two cities are the same in all respects, except that city A has ten more police officers, by how much would the two cities crime rates differ?”

- **Experiment:**
  - Randomly assign number of police officers to a large number of cities
  - In reality, number of police officers will be determined by crime rate (simultaneous determination of crime and number of police)