Principles of Econometrics

Gianfranco Piras

Outline

What is Econometrics? I

- **Econometrics** is the set of tools by which economists analyze data.
- We can use **econometrics** to:
  1. **Estimate** economic relationships
  2. **Test** economic theories
  3. **Evaluate** policy

Econometrics I

- **Econometrics** deals with the analysis of data generated by:
  - Individuals,
  - Firms,
  - Countries,
  - Municipalities
  - Schools,
  - and, more in general, entities that interact with one another.

- There are differences in the typology of data employed:
  1. **Experimental data**: Data from controlled experiments.
     - common in the natural sciences
     - harder to find in the social sciences
  2. **Nonexperimental data** (called **observational** or **retrospective**):
     - data sets are collected before
     - try to learn from what we observe

Steps in an Empirical Analysis I

- An **empirical analysis** uses data to test a theory, estimate an economic relationship, or determine the effects of a policy or intervention.
- There are some steps to follow to complete a **Successful Empirical Study**:
  1. Clear research question
  2. Specify an economic model
  3. From economics to econometrics
  4. Collect data and complete the analysis

Steps in an Empirical Analysis II

1. Be **very precise in asking the question** you hope to answer!
    - Does attending lectures in college lead to better grades on average?
    - If the severity of punishment for certain crimes increases, do crime rates fall on average?

Steps in an Empirical Analysis III

2. Specify an **economic model** (i.e. utility maximization):
    1. Becker (1968), wrote an influential article showing how **criminal behavior** can be modeled in a **utility maximizing framework.**
    2. To study the effects of **job training on worker productivity** we can start with an equation where wage is considered a function of training, education, experience, etc.

Steps in an Empirical Analysis IV

3. Turn the economic model into an **econometric model**. Again two examples:

   In a study of criminal behavior:

   - How should we measure the probability of being caught committing a crime?
   - What is the exact functional relationship among economic variables?
   - How do we account for unobserved factors that make relationships among variables inexact?

Steps in an Empirical Analysis V

- We can specify an econometric model for the wage/job training example as

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u$$

  - The constants β_0, β_1, β_2, and β_3 ("the betas") are the **parameters** of the model.
  - The last term in the equation is called the **error term** or **disturbance**.
  - Generally we want to test hypotheses about the parameters. For example, the hypothesis that job training has no effect on wage is $\beta_3 = 0$. The hypothesis that one year of experience is worth one year of education is $\beta_1 = \beta_2$.

Steps in an Empirical Analysis VI

4. **Collect data** on the variables and use statistical methods to estimate the parameters, construct confidence intervals for the parameters, and test hypotheses.

The Structure of Economic Data I

1. Cross-Sectional Data
   - Data are collected on individuals, families, firms, schools, or some other units at a given point in time.
   - In this course, we will assume that a cross-sectional data set represents a **random sample** (with replacement):
     a. Each unit in the population has the **same chance** of being in the sample
     b. The draws are **statistically independent** of one another.
   - It is useful to see how cross-sectional data sets are stored in common statistical packages.

The Structure of Economic Data II

```
R> library(wooldridge)
R> data(wage2)
```

The Structure of Economic Data III

```
R> str(wage2)
'data.frame':	935 obs. of  17 variables:
 $ wage   : int  769 808 825 650 562 1400 600 1081 1154 1000 ...
 $ hours  : int  40 50 40 40 40 40 40 40 45 40 ...
 $ IQ     : int  93 119 108 96 74 116 91 114 111 95 ...
 $ KWW    : int  35 41 46 32 27 43 24 50 37 44 ...
 $ educ   : int  12 18 14 12 11 16 10 18 15 12 ...
 $ exper  : int  11 11 11 13 14 14 13 8 13 16 ...
 $ tenure : int  2 16 9 7 5 2 0 14 1 16 ...
 $ age    : int  31 37 33 32 34 35 30 38 36 36 ...
 $ married: int  1 1 1 1 1 1 0 1 1 1 ...
 $ black  : int  0 0 0 0 0 1 0 0 0 0 ...
 $ south  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ urban  : int  1 1 1 1 1 1 1 1 0 1 ...
 $ sibs   : int  1 1 1 4 10 1 1 2 2 1 ...
 $ brthord: int  2 NA 2 3 6 2 2 3 3 1 ...
 $ meduc  : int  8 14 14 12 6 8 8 8 14 12 ...
 $ feduc  : int  8 14 14 12 11 NA 8 NA 5 11 ...
 $ lwage  : num  6.65 6.69 6.72 6.48 6.33 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

The Structure of Economic Data IV

```
R> head(wage2)
  wage hours  IQ KWW educ exper tenure age married black south urban
1  769    40  93  35   12    11      2  31       1     0     0     1
2  808    50 119  41   18    11     16  37       1     0     0     1
3  825    40 108  46   14    11      9  33       1     0     0     1
4  650    40  96  32   12    13      7  32       1     0     0     1
5  562    40  74  27   11    14      5  34       1     0     0     1
6 1400    40 116  43   16    14      2  35       1     1     0     1
  sibs brthord meduc feduc    lwage
1    1       2     8     8 6.645091
2    1      NA    14    14 6.694562
3    1       2    14    14 6.715384
4    4       3    12    12 6.476973
5   10       6     6    11 6.331502
6    1       2     8    NA 7.244227
```

The Structure of Economic Data V

2. Time Series Data
   - Consists of observations on the **same "unit" over multiple time periods** (e.g, interest rates, unemployment rates, crime rates, etc...).
   - A key feature of time series data is that the **order is relevant**!
   - Another important difference with cross-sectional data is that we cannot assume outcomes are independent across observation.
   - When we apply econometric methods to time series data, we will have to recognize that the observations are **correlated across time**.

The Structure of Economic Data VI

- Many time series exhibit clear **trends**. While real GDP sometimes rises and sometimes falls, on average it has grown over time. The notion of **trend** is not relevant for cross-sectional data.
- The **frequency** with which time series data are recorded can also be important.
- For time series that are not measured annually, **seasonality** can be important.

The Structure of Economic Data VII

```
R> data(phillips)
R> head(phillips)
  year unem  inf inf_1 unem_1 cinf      cunem
1 1948  3.8  8.1    NA     NA   NA         NA
2 1949  5.9 -1.2   8.1    3.8 -9.3  2.1000001
3 1950  5.3  1.3  -1.2    5.9  2.5 -0.5999999
4 1951  3.3  7.9   1.3    5.3  6.6 -2.0000002
5 1952  3.0  1.9   7.9    3.3 -6.0 -0.3000000
6 1953  2.9  0.8   1.9    3.0 -1.1 -0.0999999
```

The Structure of Economic Data VIII

3. Panel Data
   - The key feature is that with a panel data set the **same units** (people, houses, schools, and so on) are **followed over time**.
   - For example, we can collect information on test pass rates, spending, and some socioeconomic variables annually over, say, the last 10 years for the school in DC!
   - Following the same units over time has **advantages** when trying to infer causality.
   - Panel data analysis is a more advanced topic.

The Structure of Economic Data IX

```
R> data(crime4)
R> str(crime4)
'data.frame':       630 obs. of  59 variables:
 $ county  : int  1 1 1 1 1 1 1 3 3 3 ...
 $ year    : int  81 82 83 84 85 86 87 81 82 83 ...
 $ crmrte  : num  0.0399 0.0383 0.0303 0.0347 0.0366 ...
 $ prbarr  : num  0.29 0.338 0.33 0.363 0.325 ...
 $ prbconv : num  0.402 0.433 0.526 0.605 0.579 ...
 $ prbpris : num  0.472 0.507 0.48 0.52 0.497 ...
 $ avgsen  : num  5.61 5.59 5.8 6.89 6.55 ...
 $ polpc   : num  0.00179 0.00177 0.00184 0.00189 0.00192 ...
 $ density : num  2.31 2.33 2.34 2.35 2.36 ...
 $ taxpc   : num  25.7 24.9 26.5 26.8 28.1 ...
 $ west    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ central : int  1 1 1 1 1 1 1 1 1 1 ...
 $ urban   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ pctmin80: num  20.2 20.2 20.2 20.2 20.2 ...
 $ wcon    : num  206 213 220 223 244 ...
 $ wtuc    : num  334 369 1395 399 359 ...
 $ wtrd    : num  182 190 197 201 207 ...
 $ wfir    : num  272 301 310 350 383 ...
```

The Structure of Economic Data X

```
$ wser    : num  216 232 240 252 261 ...
$ wmfg    : num  229 240 270 282 299 ...
$ wfed    : num  409 420 439 459 490 ...
$ wsta    : num  236 254 250 262 281 ...
$ wloc    : num  231 237 249 264 289 ...
$ mix     : num  0.0999 0.103 0.0807 0.0785 0.0932 ...
$ pctymle : num  0.0877 0.0864 0.0851 0.0838 0.0823 ...
$ d82     : int  0 1 0 0 0 0 0 0 1 0 ...
$ d83     : int  0 0 1 0 0 0 0 0 0 1 ...
$ d84     : int  0 0 0 1 0 0 0 0 0 0 ...
$ d85     : int  0 0 0 0 1 0 0 0 0 0 ...
$ d86     : int  0 0 0 0 0 1 0 0 0 0 ...
$ d87     : int  0 0 0 0 0 0 1 0 0 0 ...
$ lcrmrte : num  -3.22 -3.26 -3.5 -3.36 -3.31 ...
$ lprbarr : num  -1.24 -1.08 -1.11 -1.01 -1.12 ...
$ lprbconv: num  -0.911 -0.837 -0.643 -0.503 -0.547 ...
$ lprbpris: num  -0.75 -0.679 -0.735 -0.654 -0.699 ...
$ lavgsen : num  1.72 1.72 1.76 1.93 1.88 ...
$ lpolpc  : num  -6.33 -6.34 -6.3 -6.27 -6.25 ...
$ ldensity: num  0.836 0.846 0.851 0.853 0.861 ...
$ ltaxpc  : num  3.25 3.21 3.28 3.29 3.34 ...
```

The Structure of Economic Data XI

```
 $ lwcon   : num  5.33 5.36 5.39 5.41 5.5 ...
 $ lwtuc   : num  5.81 5.91 7.24 5.99 5.88 ...
 $ lwtrd   : num  5.21 5.24 5.28 5.3 5.33 ...
 $ lwfir   : num  5.61 5.71 5.74 5.86 5.95 ...
 $ lwser   : num  5.37 5.44 5.48 5.53 5.56 ...
 $ lwmfg   : num  5.43 5.48 5.6 5.64 5.7 ...
 $ lwfed   : num  6.01 6.04 6.08 6.13 6.2 ...
 $ lwsta   : num  5.46 5.54 5.52 5.57 5.64 ...
 $ lwloc   : num  5.44 5.47 5.52 5.58 5.66 ...
 $ lmix    : num  -2.3 -2.27 -2.52 -2.54 -2.37 ...
 $ lpctymle: num  -2.43 -2.45 -2.46 -2.48 -2.5 ...
 $ lpctmin : num  3.01 3.01 3.01 3.01 3.01 ...
 $ clcrmrte: num  NA -0.0394 -0.2353 0.1362 0.0518 ...
 $ clprbarr: num  NA 0.1545 -0.0229 0.0926 -0.1081 ...
 $ clprbcon: num  NA 0.0741 0.194 0.14 -0.0439 ...
 $ clprbpri: num  NA 0.071 -0.0553 0.0809 -0.0453 ...
 $ clavgsen: num  NA -0.00357 0.03688 0.17221 -0.05061 ...
 $ clpolpc : num  NA -0.0114 0.0384 0.0269 0.0202 ...
 $ cltaxpc : num  NA -0.0326 0.0615 0.0147 0.0472 ...
 $ clmix   : num  NA 0.0309 -0.2447 -0.0273 0.1721 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

The Structure of Economic Data XII

```
R> head(crime4[,1:6], 14)
   county year    crmrte   prbarr  prbconv  prbpris
1       1   81 0.0398849 0.289696 0.402062 0.472222
2       1   82 0.0383449 0.338111 0.433005 0.506993
3       1   83 0.0303048 0.330449 0.525703 0.479705
4       1   84 0.0347259 0.362525 0.604706 0.520104
5       1   85 0.0365730 0.325395 0.578723 0.497059
6       1   86 0.0347524 0.326062 0.512324 0.439863
7       1   87 0.0356036 0.298270 0.527596 0.436170
8       3   81 0.0163921 0.202899 0.869048 0.465753
9       3   82 0.0190651 0.162218 0.772152 0.377049
10      3   83 0.0151492 0.181586 1.028170 0.438356
11      3   84 0.0136621 0.194986 0.885714 0.500000
12      3   85 0.0120346 0.206897 0.909091 0.366667
13      3   86 0.0129982 0.156069 1.037040 0.392857
14      3   87 0.0152532 0.132029 1.481480 0.450000
```

Causality and the Notion of Ceteris Paribus I

- The concept of **causality** is key in econometrics.
  1. How can we know that more spending *causes* better student performance (on average)?
  2. How can we know another year of education *causes* an increase in wages (on average)?
- Crucial to establishing causality is the notion of **ceteris paribus**: "all (relevant) factors equal." If we succeed in "holding fixed" other relevant factors, then sometimes we can establish that changes in one variable (say, education) in fact "cause" changes in another variable (wage).