

Lecture 1: What is Econometrics

Zheng Tian

Outline

- 1 What is Econometrics
- 2 Economic Questions We Examine
- 3 Causal Effects and Idealized Experiments
- 4 Data Sources and Types

- 1 What is Econometrics
- 2 Economic Questions We Examine
- 3 Causal Effects and Idealized Experiments
- 4 Data Sources and Types

What do you think of Econometrics?

- Economics?
- Mathematics?
- Statistics?

Definition of Econometrics

Stock and Watson (2015) define Econometrics as

At a broad level, econometrics is the science and art of using economic theory and statistical techniques to analyze economic data.

Science

- The principle of **falsifiability** of scientific research, as Karl Popper defined.

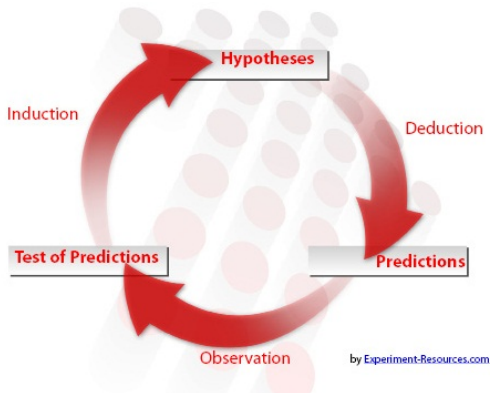


Figure: A reasoning cycle of scientific research

Art

Econometrics is an art because

- the data are usually incomplete and unobserved to validate a hypothesis
- we need to use human creativity to reach a balance between scientific rigor and realistic approximation.

A quote from Hansen

Econometrics is alchemy since econometricians can create nearly any result desired, but it is also science because econometricians also know how to reject and avoid spurious models. – Hansen (1996)

Economic theory, statistics, and data

- A complete process of econometric research consists of
 - Economic theory
 - Statistical techniques
 - Economic data

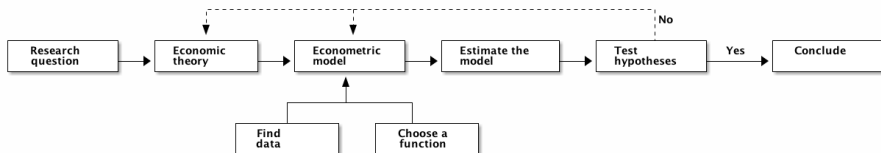


Figure: A workflow of econometric research

- 1 What is Econometrics
- 2 Economic Questions We Examine
- 3 Causal Effects and Idealized Experiments
- 4 Data Sources and Types

The argument for small class size goes like this

Small classes get more teacher-student interaction, fewer disruptions, and higher grades in test.



The question of interest

The research question

Is there any effect of reducing class size on improving students' grades in elementary schools?

Who cares such research?

- Teachers
- Parents
- School principles
- Superintendents of school districts

The research design

Qualitative research design A field investigation

Quantitative research design Randomized controlled experiments (RCE, or randomized controlled trial, RCT)

The sample and data

- Draw samples and collect data from 420 California school districts in 1999.
- Cross-sectional data. Each row represents a distinct unit of observation. All observations are collected in a single year.

Observation Number	School District	District Average Test Score	Expenditure per Pupil(\$)	Student-Teacher Ratio	District Average Family Income	Percentage of Students Learning English
1	Sunol Glen Unified	690.80	6384.91	17.89	22.69	0.00
2	Manzanita Elementa	661.20	5099.38	21.52	9.82	4.58
3	Thermalito Union Ele	643.60	5501.95	18.70	8.98	30.00
4	Golden Feather Unio	647.70	7101.83	17.36	8.98	0.00
5	Palermo Union Eleme	640.85	5235.99	18.67	9.08	13.86
6	Burrel Union Elemen	605.55	5580.15	21.41	10.41	12.41
7	Holt Union Elementa	606.75	5253.33	19.50	6.58	68.72
8	Vineland Elementary	609.00	4565.75	20.89	8.17	46.96
9	Orange Center Eleme	612.50	5355.55	19.95	7.39	30.08
10	Del Paso Heights Ele	612.65	5036.21	20.81	11.61	40.28
11	Le Grand Union Eleme	615.75	4547.69	21.24	8.93	52.91
12	West Fresno Elemen	616.30	5447.35	21.00	7.39	54.61
13	Allensworth Element	616.30	6567.15	20.60	5.34	42.72
14	Sunnyside Union Ele	616.30	4818.61	20.01	8.28	20.53
15	Woodville Elementar	616.45	5621.46	18.03	9.63	80.12
16	Pixley Union Element	617.35	6026.36	20.25	7.45	49.41
17	Lost Hills Union Eleme	618.05	6723.24	16.98	6.22	85.54
18	Buttonwillow Union	618.30	5589.89	16.51	7.76	58.91
19	Lennox Elementary	619.80	5064.62	22.70	7.02	77.01
20	Lamont Elementary	620.30	5433.59	19.91	5.70	49.81
21	Westmorland Union	620.50	5725.56	18.33	7.94	40.68
22	Pleasant View Eleme	621.40	4542.10	22.62	9.63	16.21
23	Wasco Union Elemer	621.75	5107.09	19.45	7.41	45.07
24	Alta Vista Elementary	622.05	4659.66	25.05	9.63	39.08
25	Livingston Union Ele	622.60	4555.46	20.68	8.02	76.67

Figure: A screen shot of the dataset the California school districts in 1999

The econometric model

- Use common sense to build an econometric model in this case.
- Variables involved: the average test scores in a school district (*TestScore*) and the student-teacher ratio *STR*.
- For simplicity, we set up a **simple linear regression model** as follows,

$$TestScore = \beta_0 + \beta_1 STR + OtherFactors$$

- The hypothesis we make is that if *STR* has a non-zero effect on *TestScore*, that is, $\beta_1 \neq 0$.
- The model is then estimated using some estimation method, and we test the hypothesis with the estimation results using some test statistics.

The three other questions

Question 1 Does reducing class size improve elementary school education?

Question 2 Is there racial discrimination in the market for home loan?

Question 3 How much do cigarette taxes reduce smoking?

Question 4 What will the rate of inflation be next year?

A summary of data types

Table: Data types and econometric methods for all four questions

Questions	Data types	Econometric methods
#1	experimental, cross-sectional	multiple regression
#2	observational, cross-sectional	multiple regression with binary dependent variable
#3	observational, panel data	Panel data regression model
#4	observational, time series	multiple regression with lagged dependent variable

- 1 What is Econometrics
- 2 Economic Questions We Examine
- 3 Causal Effects and Idealized Experiments**
- 4 Data Sources and Types

Randomized controlled experiments (or trials, RCTs thereafter)

- Clinical trials to test the effectiveness of medical intervention.
- All participants are **randomly** assigned into two groups.
- The control group receives no treatment (or placebo)
- The treatment group receives the treatment.
- After a follow-up period, compare the two groups.

An illustration of RCTs

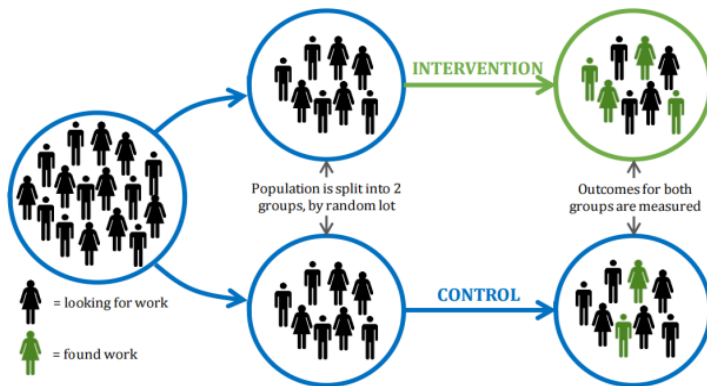


Figure: An illustration of a randomized controlled experiment

The advantage of RCTs

- Randomization minimizes selection bias.
- In the example of California school districts, randomized control experiments ensure that the only systematic difference between the classes in the control group and those in the treatment group is the treatment (reduced class size) itself, with the effects from other **confounding factors** eliminated.

The disadvantage of RCTs

Time and costs RCTs usually are expensive to undertake and take a long time to observe the effect of treatment.

Conflict of interest dangers RCTs may be funded by special interest groups so that its objectivity is doubtful.

Ethnics Especially in social science, we cannot impose some treatment due to ethnic concerns.

Causal effect

- **Causal effect** is defined to be the effect on an outcome of a given action or treatment as measured in an ideal RCT.
- The concept of the ideal randomized controlled experiment does provide a theoretical benchmark to define causal effects in research design.

- 1 What is Econometrics
- 2 Economic Questions We Examine
- 3 Causal Effects and Idealized Experiments
- 4 Data Sources and Types**

Experimental versus observational data

- **Experimental data** come from experiments designed to evaluate a treatment or policy or to investigate a causal effect.
- **Observational (or nonexperimental) data** are collected using surveys, and administrative records.
- The problem of using observational data to estimate causal effects is that the "treatment" is not randomly assigned.
- Much of econometric methods are developed to deal with causality using observational data.

Cross-sectional data

- Data on different entities for a single time period are called **cross-sectional data**.
- The sequence of each observation number is arbitrarily assigned.
- Cross-sectional data can be experimental data or observational data.

Time series data

- Time series data are data for a single entity collected at multiple time periods.
- The sequence of each record is based on the time period it happened.

Panel data

- Panel data, also called longitudinal data, are data for multiple entities in which each entity is observed at two or more time periods.
- Panel data are very useful for estimating causal effects.