Applied Microeconometrics (L1): Introduction

Nicholas Giannakopoulos

University of Patras
Department of Economics

ngias@upatras.gr

October 7, 2022

Overview

Doing empirical analysis

Treatment effects approach

Structural approach

Types of “research” questions

Correlation does not imply causation

Process of project development

Example of empirical analysis

- What?
  - Estimate social returns to college education
- Why?
  - The magnitude of the social return to education is important for assessing the efficiency of public investment in education
- How?
  - Compare average wages across cities with different share of college graduates (Morretti, 2004)

Example of empirical analysis

- What about the validity of this approach?
  - in the case of zero social returns, there is a wage premium for college graduates (i.e., human capital theory: Becker, 1993)
  - thus, we need to disentangle the private returns to schooling from the social returns to schooling (Morretti, 2004)

Identification issues

- Which group to analyze?
  - college graduates vs non-college graduates
- Economic rationale?
  - an increase in the share of college graduates (workers) will raise productivity of non-college graduates (workers)
  - an increase in the share of college graduates (workers) will negatively affects its own marginal product
  - wages are equal to workers' marginal productivity
  - more college graduates imply lower wages
  - human capital spillovers
- Identification when individuals and cities are heterogeneous
  - comparison of wages for otherwise similar individuals who work in cities with different shares of college graduates in the labor force

Basic Model

$$\text{wages}_{ic} = (\text{education}_{ic}, \text{environment}_{c}, \text{else}_{ic})$$

$$\text{where, i (workers)} = 1, ..., N \text{ and c (cities)} = 1, ..., C$$

i	c	wage	education	environment	else
1	1	980	12	10.0	male
2	1	580	9	10.0	female
3	1	689	9	10.0	male
4	2	900	12	12.3	male
5	2	1200	16	12.3	female
6	2	650	9	12.3	male
7	3	750	9	10.3	male
8	3	250	6	10.3	female
9	3	1024	16	10.3	male
10	4	589	6	9.0	male
11	4	888	9	9.0	female
12	4	1011	12	9.0	male

Basic Model

$$lnW_{ic} = \beta_0 + X_{ic}\beta_1 + Z_c\beta_2 + Y_{ic}\beta_3$$

- Using only the sample of non-college graduates
  - estimate a positive β_2, even if social returns to education are zero.
- Using only the sample of college graduates
  - estimate a positive β_2, only if social returns to education are positive.

Additional Issues

$$lnW_{ic} = \beta_0 + X_{ic}\beta_1 + Z_c\beta_2 + Y_{ic}\beta_3$$

- Other factors that may be important
  - Individuals' ability
  - City-specific labor demand and/or supply shocks
- Overcoming endogeneity problems
  - Longitudinal data (individual level)
  - Proxies for local demand conditions (city level)
  - Using instruments for the share of college graduates (city level)

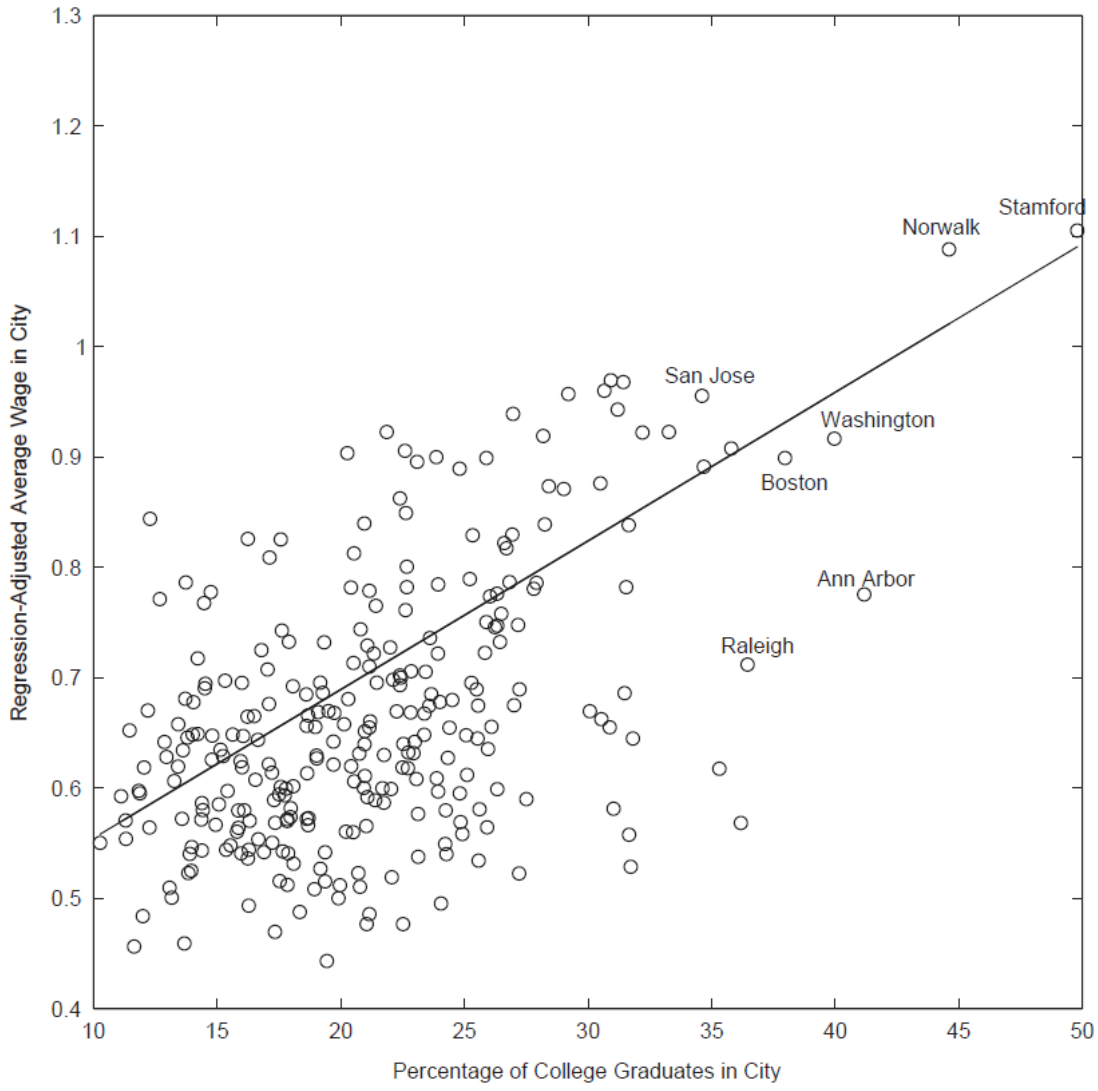
Summary Statistics (example)

Table 1
Summary statistics

	NLSY 1979–94	Census	
		1990	1980
Individual-level variables			
Log hourly wage	1.89	2.30	2.26
	(0.51)	(0.70)	(0.67)
Years of education	13.10	13.17	12.86
	(2.36)	(2.78)	(3.00)
Years of experience	9.06	19.25	19.14
	(3.96)	(12.7)	(13.83)
Female	0.48	0.47	0.44
Black	0.29	0.10	0.11
Hispanic	0.21	0.09	0.06
US Citizen	0.95	0.94	0.96
Work in manufacturing	0.18	0.18	0.24
City-level variables			
Share of college graduates	0.24	0.25	0.21
	(0.05)	(0.06)	(0.05)
Unemployment rate	6.62	5.17	5.97
	(1.62)	(1.05)	(1.07)
Log monthly rent	5.83	6.11	5.85
	(0.17)	(0.18)	(0.17)
Cities	201	282	282
Individuals (N)	6791	1,981,985	1,694,678
$N \times T$	44891		

Notes: Standard deviations of continuous variables are in parenthesis.

Scatterplot (example)



Results (example): OLS and IV

Table 5
The effect of changes in share of college graduates on wage of education groups: 1980 and 1990 Census data

	1980–1990				1990		1980	
	OLS (1)	OLS (2)	2SLS (3)	2SLS (4)	OLS (5)	2SLS (6)	OLS (7)	2SLS (8)
First stage								
Age structure			3.36	3.52				
			(0.67)	(0.67)				
Land Grant						0.05		0.05
						(0.01)		(0.01)
Second stage								
Less than high-school	1.47	1.44	2.22	1.91	0.75	0.77	0.67	0.58
	(0.15)	(0.15)	(0.51)	(0.52)	(0.06)	(0.20)	(0.07)	(0.17)
High-school	1.38	1.34	2.08	1.67	0.85	0.84	0.74	0.74
	(0.13)	(0.13)	(0.45)	(0.45)	(0.06)	(0.18)	(0.06)	(0.14)
Some college	1.29	1.25	1.66	1.24	0.86	0.94	0.70	0.63
	(0.12)	(0.12)	(0.42)	(0.42)	(0.06)	(0.18)	(0.06)	(0.14)
College +	0.87	0.83	0.86	0.47	0.74	0.55	0.70	0.45
	(0.11)	(0.10)	(0.35)	(0.37)	(0.06)	(0.19)	(0.07)	(0.17)
City effects	Yes	Yes	Yes	Yes				
Region effects					Yes	Yes	Yes	Yes
Unempl. and other city controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Katz and Murphy Index		Yes		Yes				

Results (example): Panel Data

Table 2
The effect of increases in percentage of college graduates on wages: NLSY data

	(1)	City effects (2)	City, indiv. effects (3)	City×indiv. effects (4)	City×indiv. effects (5)	City×indiv. effects (6)	City×indiv. effects (7)	City×indiv. eff. only manuf. (8)
(B) *Allowing the private return to education to vary by year and city*								
College share	1.42 (0.38)	1.16 (0.36)	1.17 (0.31)	1.02 (0.33)	1.02 (0.33)	1.23 (0.33)	1.23 (0.33)	1.10 (0.91)
Experience	0.04 (0.00)	0.04 (0.00)	0.02 (0.02)	0.03 (0.02)	0.03 (0.02)	0.01 (0.02)	0.01 (0.02)	0.09 (0.03)
Experience sq./100	−0.04 (0.02)	−0.07 (0.01)	−0.13 (0.01)	−0.14 (0.01)	−0.14 (0.01)	−0.14 (0.01)	−0.14 (0.01)	−0.17 (0.03)
Female	−0.18 (0.01)	−0.19 (0.00)	—	—	—	—	—	—
Black	−0.20 (0.01)	−0.20 (0.01)	—	—	—	—	—	—
Hispanic	−0.07 (0.02)	−0.07 (0.01)	—	—	—	—	—	—
R^2	0.28	0.29	0.72	0.75	0.75	0.76	0.76	0.88

Example: Effect of Spillovers on Firm Output

Manufacturing firms operating during a 30-year period: Production Function (Empirical Model):

- $log(Q_t) = \delta_t + \beta_1 log(L_t) + \beta_2 log(K_t) + \beta_3 S_t + QL + u_t,\ t = 1, ..., 30$
  - Q_t is the monetary value of the total production in year t
  - L_t is the total number of labor units in year t
  - K_t is the monetary value of the capital utilization in year t
  - S_t is a measure of foreign firm concentration in the sector
  - QL contains time-invariant unobserved productivity factors (e.g., unobserved managerial or worker quality)
  - u_t represents unobserved shocks in each time period
  - δ_t different intercepts in each year, allows for aggregate productivity to change over time
- $log(Q_{it}) = \delta_t + \beta_1 log(L_{it}) + \beta_2 log(K_{it}) + \beta_3 S_{it} + QL_i + u_{it}$
  $t = 1, ..., 30,\ i = 1, ..., N$

Take home message (!)

- precise research question
- analyze in details the investigated relationship
- understand what the econometric model describes

Treatment effects (TE) approach

- TE is context-specific and addresses policy-related questions (and to assess future policies)
- TE evaluates the causal impact of an existing policy for individuals affected by the policy (treatment group)
- TE emphasizes on identifying the causal effect of the treatment (policy), i.e. by randomization of treatment status

Structural approach

- Traditional approach
- The structural approach aims to use data from a particular context to identify (using the theory) behavioral rules behavior that can be extrapolated to other contexts
- The structural approach aims to choose a theory-based model that best fits the data
- The structural approach is used for ex-post or ex-ante policy simulation

The Modelling Process

1. Statement of theory/hypothesis
2. Specification of mathematical model
3. Specification of the econometric model
4. Obtaining the data / conduct preliminary data analysis
5. Estimation of the econometric model and interpretation of results
6. Diagnostic Analysis
7. Hypothesis testing
8. Prediction/forecasting

Types of "research" questions

1. Descriptive
2. Forecasting
3. Causal

Descriptive

Evidence on correlations, not causal mechanisms

- How much do men and women earn annually on average?
- Has inequality increased overtime?
- Intergenerational income mobility
- How are modern managerial strategies associated with firm's profitability and performance?

Forecasting

Use current data to predict future events, not causal mechanisms

- What will the global temperature be in 2020?
- How large should be GDP growth in order to generate a decrease in unemployment rate?
- Is it possible to predict if a loan is going to be reimbursed?

Causal

Questions of type: What-if...?

- How much higher grade you get from your master's thesis as a result of taking this course (relative to your best outside option)?
- Does smoking cause lung-cancer?
- Does death penalty decrease crime rates?
- Would it be profitable for a firm to allow employees to work from home?

What do we know?

- Correlation does not imply causation
- y (rain) can cause x (people with umbrellas) even if x takes place before y
  - ...subsidizing umbrellas is not a great policy to increase rain (!)

Correlation does not imply causation

- $Corr(x, y) \neq 0$
  1. x affects y
  2. y affects x
  3. but... z affects x and y
- Examples
  1. Red cars are more likely to get involved in accidents
  2. People that sleep less tend to live longer
  3. Countries that eat more chocolate receive more Nobel prizes
- Other cases
  1. Students in households with more books tend to perform better in PISA
  2. GDP growth and public debt
  3. R&D expenditure and firms' profits

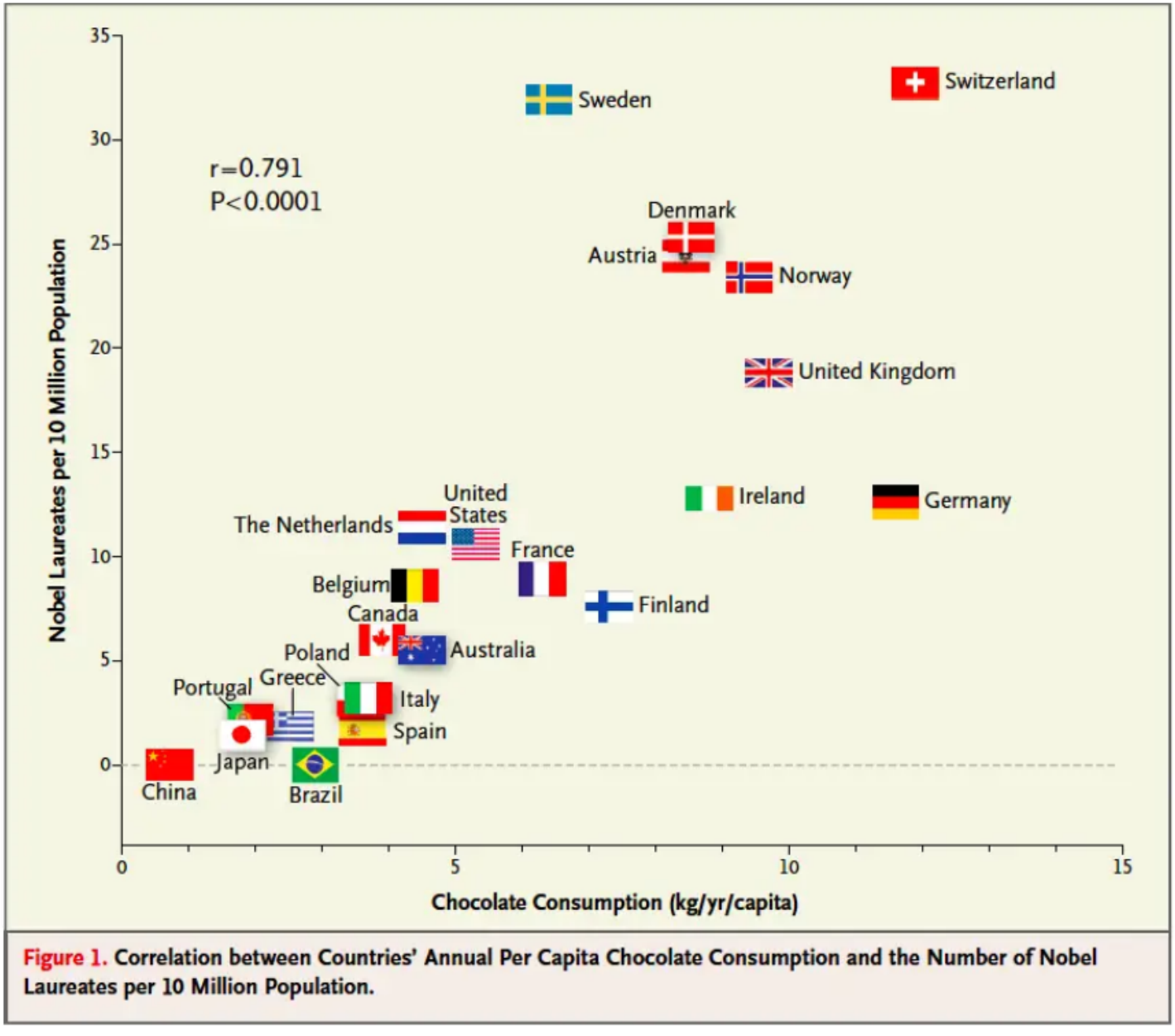


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

Source: https://www.businessinsider.com/chocolate-consumption-vs-nobel-prizes-2014-4

FAQs (Angrist and Pischke, 2009)

- Relationship of interest (level of analysis: individual, firm, country)
  - the causal effect of schooling on wages: the increment to wages an individual would receive if he or she got more schooling (Card, 1999)
  - effect of colonial institutions on economic growth: countries with more democratic institutions from their colonial rulers later enjoyed higher economic growth as a consequence (Acemoglu, Johnson, and Robinson, 2001)

FAQs (Angrist and Pischke, 2009)

- Ideal experiment (hypothetical scenario)
  - offer to potential dropouts a reward for finishing school, and then studying the consequences (Angrist and Lavy, 2007)
  - go back in time and randomly assign different government structures to former colonies on their Independence Days
- Fundamentally Unidentified Questions (FUQ'd): research questions that cannot be answered by any experiment
  - what is the causal effect of race or gender? ..."imagine your chromosomes were switched at birth"
  - but...in Labor Economics we care about "labor market discrimination" i.e., on whether someone treats you differently because they believe you to be black or white, male or female.
  - "counterfactual world": men are perceived as women

FAQs (Angrist and Pischke, 2009)

- Identification strategy
  - when the researcher uses observational data (i.e., data not generated by a randomized trial) to approximate a real experiment (Angrist and Krueger, 1999)
  - Question: estimate the effects of finishing high school on wages. Natural experiment: interaction between compulsory attendance laws in American schools and students' season of birth (Angrist and Krueger, 1991)
    - Rationale: Season of birth affects the degree to which high school students are constrained by laws allowing them to drop out on their birthdays
- Haavelmo (1944): two classes of experiments
  1. experiments that we should like to make to see if certain real economic phenomena when artificially isolated from other influences would verify certain hypotheses
  2. experiments based on Nature: we are passive observers

FAQs (Angrist and Pischke, 2009)

- Mode of inference
  - Q: what is your mode of statistical inference? Rubin (1991)
  - A: depends on the population to be studied, the sample to be used, and the assumptions made when constructing standard errors
  - inference is more complex, especially with data that are clustered or grouped
  - t-stat looks too good...
    - -use robust standard errors...
    - -significance gone

Clustered data