Applied Quantitative Methods II.

Lecture 2. Treatment effects, OLS, and randomization

Klára Kalíšková

University of Economics, Prague

CAUSAL INFERENCE AND TREATMENT EFFECTS

Research Questions – Examples?

- Does an additional year of schooling increase life time earnings?
- Do active labor market policies improve individual employment prospective?
- Does better neighborhood improve individual wages?
- Does classroom attendance improve score from final exam?

Stages of Empirical Research

1. Finding Research questions
2. Formulation of a model or intuition
3. Formulation of an econometric specification
4. Data collection
5. Estimation
6. Interpretation of results
7. Conclusions & Policy Recommendation

Typical Structure of a Research Output - Empirical Paper

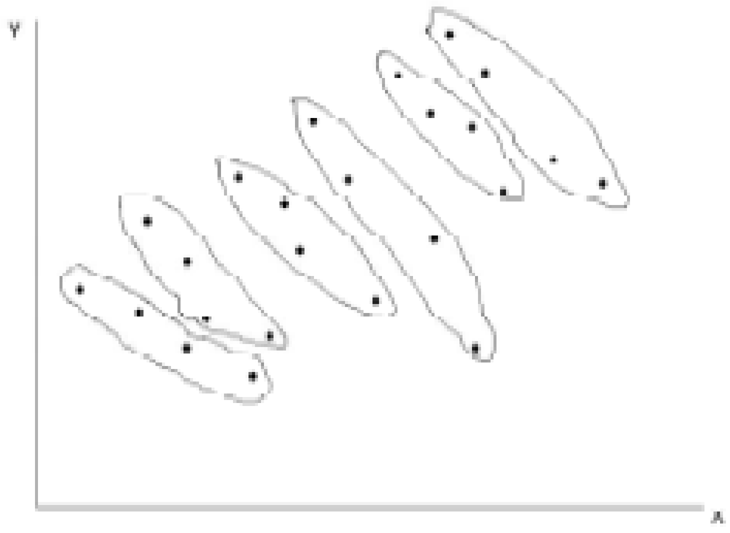
- Introduction (structure)
- Literature Review
- Methodology (empirical/identification strategy)
- Data Description
- Results
- Conclusion (structure)
- How to write a paper? http://faculty.chicagobooth.edu/john.cochrane/research/Papers/phd_paper_writing.pdf

Basic Problem in Microeconometrics

- We want to test whether *d* affects *y*
- Observed correlation(association) does not imply existence of causation
- Reasons:
  - Confounding (unobserved) factors that drive the correlation
- Observed correlation does not (usually) answer the important question:
  - What would happen if an individual/firm was (not) under particular treatment (i.e. obtaining education)?
- True impact of a treatment (d) on idividual outcome (y) – TREATMENT EFFECT
- Example: OLS framework

Association vs Causation: Example

Positive linear relationship?



NO, different intercept for individuals and negative relationship

Causal Inference – Treatment Effects

- Measuring the impact of treatment on particular outcome – two scenarios, treated or non-treated individual

$$(1)\quad \begin{aligned} y_i^1 &= \beta + \alpha_i + u_i \\ y_i^0 &= \beta + u_i \end{aligned}$$

potential outcomes
for the same individual *i*

Observable outcome after realization of treatment (d – 0,1)

$$(2)\quad y_i = d_i y_i^1 + (1 - d_i) y_i^0$$

$$y_i = \beta + \alpha_i d_i + u_i \,.$$

What is the selection rule?

- d – can depand on observed (Z) and unobsered factors (v)
- Examples?

Treatment effects

- Averega treatment effect (ATE)

$$(6)\quad \alpha^{ATE} = E(\alpha_i)$$

- Average treatment on treated (ATT)

$$(7)\quad \begin{aligned} \alpha^{ATT} &= E(\alpha_i | d_i = 1) \\ &= E(\alpha_i | g(Z_i, v_i) \geq 0) \end{aligned}$$

- Average treatment effect on non-treated (ATNT)

$$(8)\quad \begin{aligned} \alpha^{ATNT} &= E(\alpha_i | d_i = 0) \\ &= E(\alpha_i | g(Z_i, v_i) < 0). \end{aligned}$$

Treatment effects

- ATE – effect on randomly chosen individual from population
- ATT – effect on the participant of treatment (program)
- ATNT – hypothetical effect on those who did not participate
- LATE – local average treatment effect
- ITE – Intention to treat effect

Treatment Effects – Heterogenous in Population

- Local average treatment effect - LATE (instrumental variable, discontinuity design)
  - The effect is identified only for small subgroup that is participating
  - Usually this small subgroup is specific with respect to treatment

Example: we identify ATE only for small group around threshold

Intention to Treat Effect

- We do not know who would be participating in treatment
- We estimate effect on those who are eligible for certain treatment
- Example: joint taxation for couples
  - Kaliskova (2014) for Czech Rep.

Where is the problem in identifying the effects?

- Non-random assignment into treatment:
  - SELECTION into treatment group (treated are different from untreated – control group)
- What is the mechanism of the selection into treatment group?
- Remainder:

$$(4)\ d_i^* = g(Z_i, v_i).$$

Selection Mechanisms

- Selection on observables (corr of u and Z)
- Selection on unobservables (corr u and v)

- Selection on untreated outcome (corr u and d)
- Selection on expected gains (corr of α and d)

- Different estimation strategies use different assumptions about a selection mechanism

Example – Returns to Education

Decision about eductional attainment:

- Costs
- Expected gains

Problem: Are those who attain higher education the same as those who do not?

- What is the selection mechanism?
- What do we observe?
- What is unobservable?

Overview of Identification Strategies

- Controlled (social) experiment
  - Direct randomisation of treated and untreated
- Natural experiment
  - Finding naturally occuring treated and untreated group that are as similar as possible
- Instrumental variable
  - Finding variadle that assign into treatment but do not affect outcome
- Discontinuity design
  - Probability of treatment is changin discountinously with a characteristic (i.e. age)

What is the goal? Getting rid of selection mechanism

ESTIMATION – OLS METHOD

Stages of Empirical Research

1. Formulation of a model or intuition
2. Formulation of an econometric specification
3. Data collection
4. **Estimation**
5. Interpretation of results

OLS: Research question

- How quantity sold by a firm is related to the unit costs of production?

Economic model

- Price of product: *p*
- Unit costs: *c*
- Demand for a product: *q(p)*

profit $$\pi(p) = q(p)(p - c)$$

demand $$q(p) = b - a.p$$

Result of maximization $$q = \frac{b}{2} - \frac{a}{2}.c$$

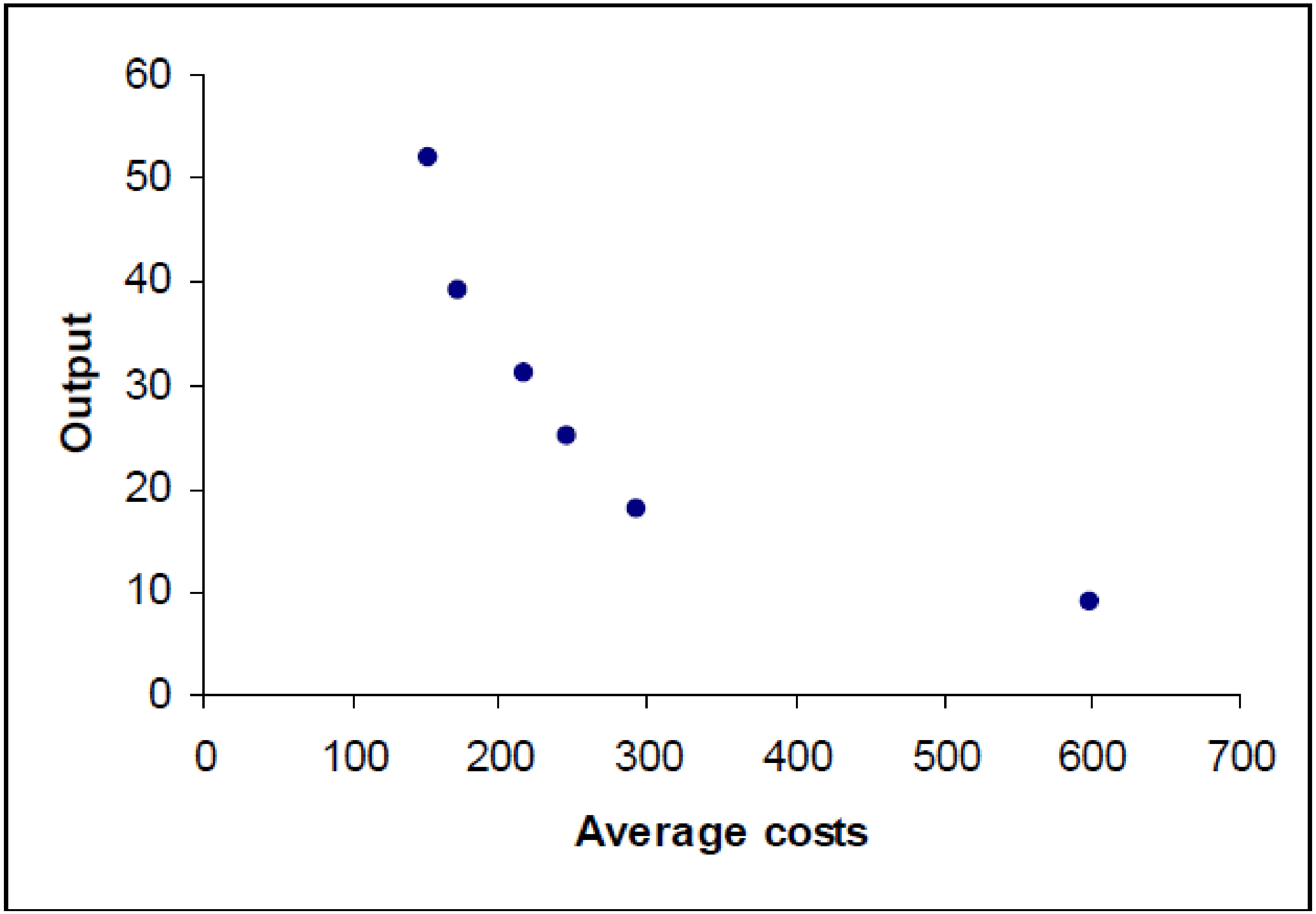
Econometric Specification

- Reformulated model $$q_i = \alpha_0 + \alpha_1 c_i$$

- Econometric specification $$q_i = \alpha_0 + \alpha_1 c_i + \varepsilon_i$$

- How to estimate coefficients?

Data



60
50
40
30
20
10
0
Output
0
100
200
300
400
500
600
700
Average costs

Looking for ideal linear model

- fitting our data points
- Minimize error
- Minimize vertical distance between data points and linear line
- Minimize sum of squared residuals

OLS

- We minimize residuals (distance from the regression line):

$$
\begin{aligned}
e_1 &= q_1 - \hat{\alpha}_0 - \hat{\alpha}_1 c_1 \\
e_2 &= q_2 - \hat{\alpha}_0 - \hat{\alpha}_1 c_2 \\
&\cdots\cdot \\
e_6 &= q_6 - \hat{\alpha}_0 - \hat{\alpha}_1 c_6
\end{aligned}
$$

OLS

$$\min_{\hat{\alpha}_0, \hat{\alpha}_1} \sum_{i=1}^{6} (q_i - (\hat{\alpha}_0 + \hat{\alpha}_1 c_i))^2$$

First order conditions:

$$\frac{\partial}{\partial \hat{\alpha}_0} = 0 \qquad 2\sum_{i=1}^{6} (q_i - (\hat{\alpha}_0 + \hat{\alpha}_1 c_i)) = 0$$

$$\frac{\partial}{\partial \hat{\alpha}_1} = 0 \qquad 2\sum_{i=1}^{6} c_i (q_i - (\hat{\alpha}_0 + \hat{\alpha}_1 c_i)) = 0$$

OLS

$$2\sum_{i=1}^{6}(q_i - (\hat{\alpha}_0 + \hat{\alpha}_1 c_i)) = 0$$

$$2\sum_{i=1}^{6}c_i(q_i - (\hat{\alpha}_0 + \hat{\alpha}_1 c_i)) = 0$$

$$\hat{\alpha}_1 = \frac{\sum c_i q_i - \frac{1}{n}\sum c_i \sum q_i}{\sum c_i^2 - \frac{1}{n}(\sum a_i)^2} = \frac{\sum (c_i - \bar{c})(q_i - \bar{q})}{\sum (c_i - \bar{c})^2}$$

$$\hat{\alpha}_0 = \bar{q} - \hat{\alpha}_1 \bar{c}$$

OLS – summary

- Model
  - Economic: q=f(c)
  - Econometric

$$q_i = \alpha_0 + \alpha_1 c_i + \varepsilon_i$$

- Data collection
- Fitting regression lines using minimization of squared residuals

$$\hat{\alpha}_1 = \frac{\sum (c_i - \bar{c})(q_i - \bar{q})}{\sum (c_i - \bar{c})^2} \qquad \hat{\alpha}_0 = \bar{q} - \hat{\alpha}_1 \bar{c}$$

Terminology

y – dependent v., endogenous v., explained v., LHS

x – independent, exogenous, explanatory, RHS

ε – disturbance term

e – residual

intercept, constant

α_0 coefficients, slope paramete

α_1

$$\hat{y}_i = \hat{\alpha}_0 + \hat{\alpha}_1 x_i$$ Regression line

Disturbance term(ε) vs. residual

- Theoretical specification

$$q_i = \alpha_0 + \alpha_1 c_i + \varepsilon_i$$

- Estimated model

$$q_i = \hat{\alpha}_0 + \hat{\alpha}_1 c_i + e_i$$

$$\hat{q}_i = \hat{\alpha}_0 + \hat{\alpha}_1 c_i$$

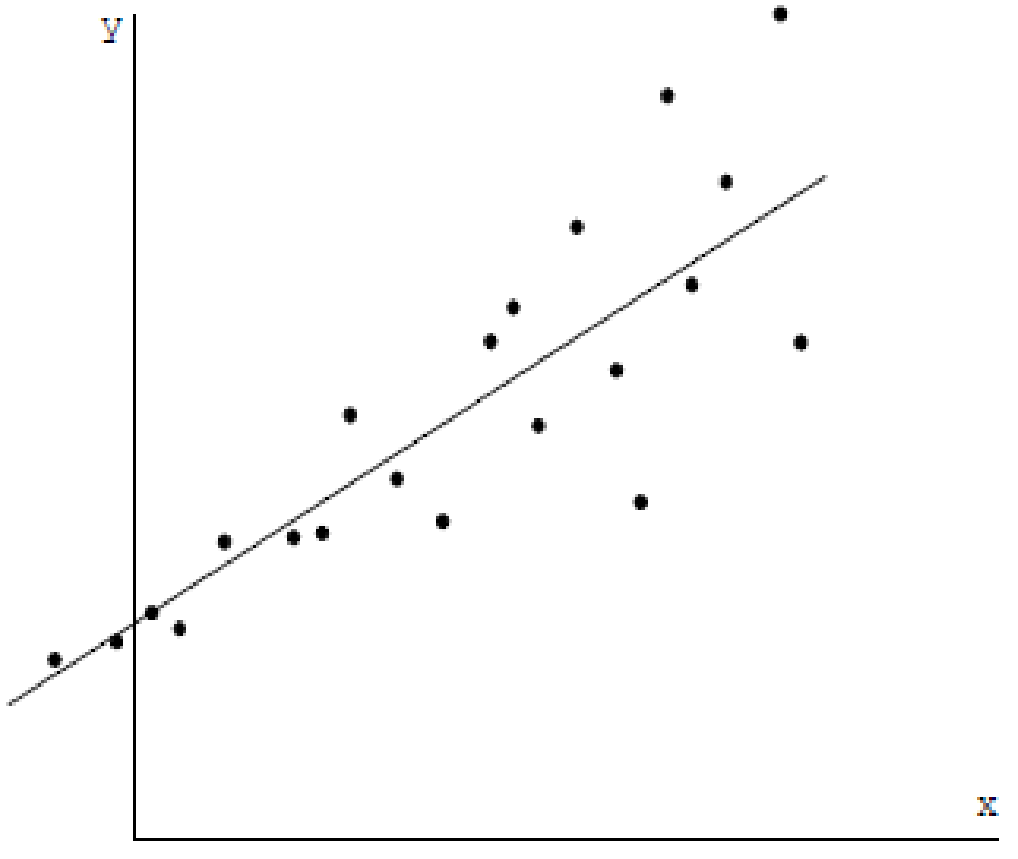
$$q_i = e_i + \hat{q}_i$$

$$q_i - \hat{q}_i = e_i$$

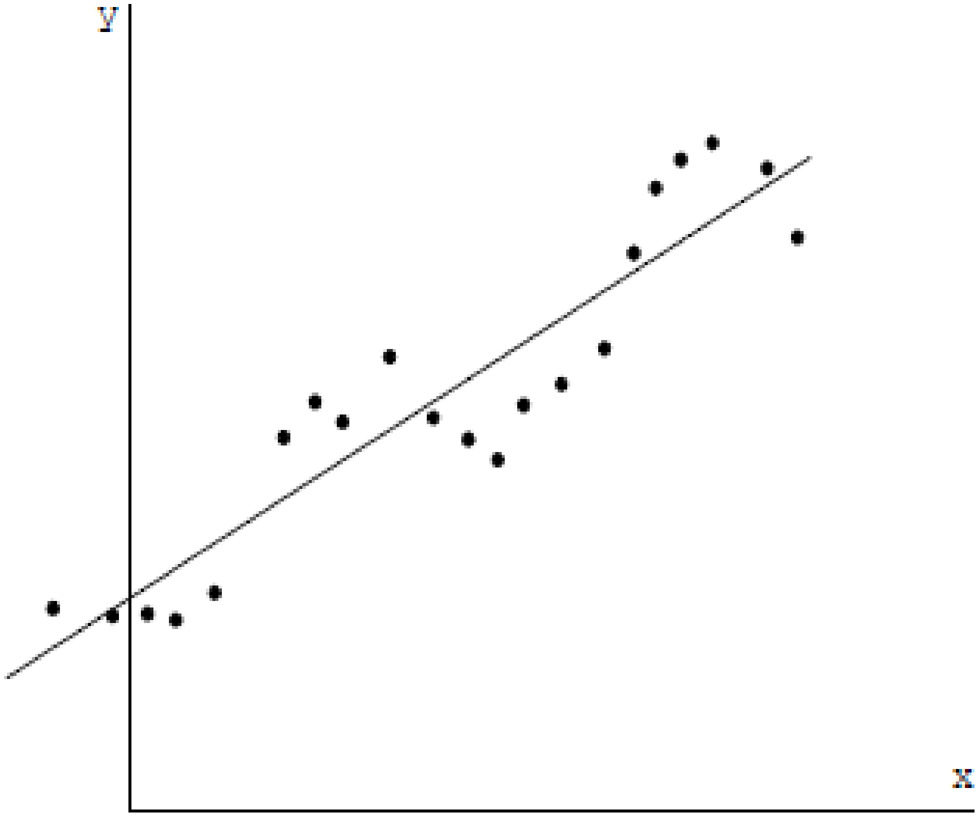
Basic assumptions of OLS

- Disturbance term should be:
  - From one distribution (normální)
  - This distribution has $E(\varepsilon_i)=0$
  - Variance is constant over *i*: $Var(\varepsilon_i)=\sigma^2$ (homoscedasticity)
  - No autocorrelation
    - $Cov(\varepsilon_i, \varepsilon_j)=0$
  - There is not correlation with explanatory variable
    - $Cov(x_i,\varepsilon_i)=0$

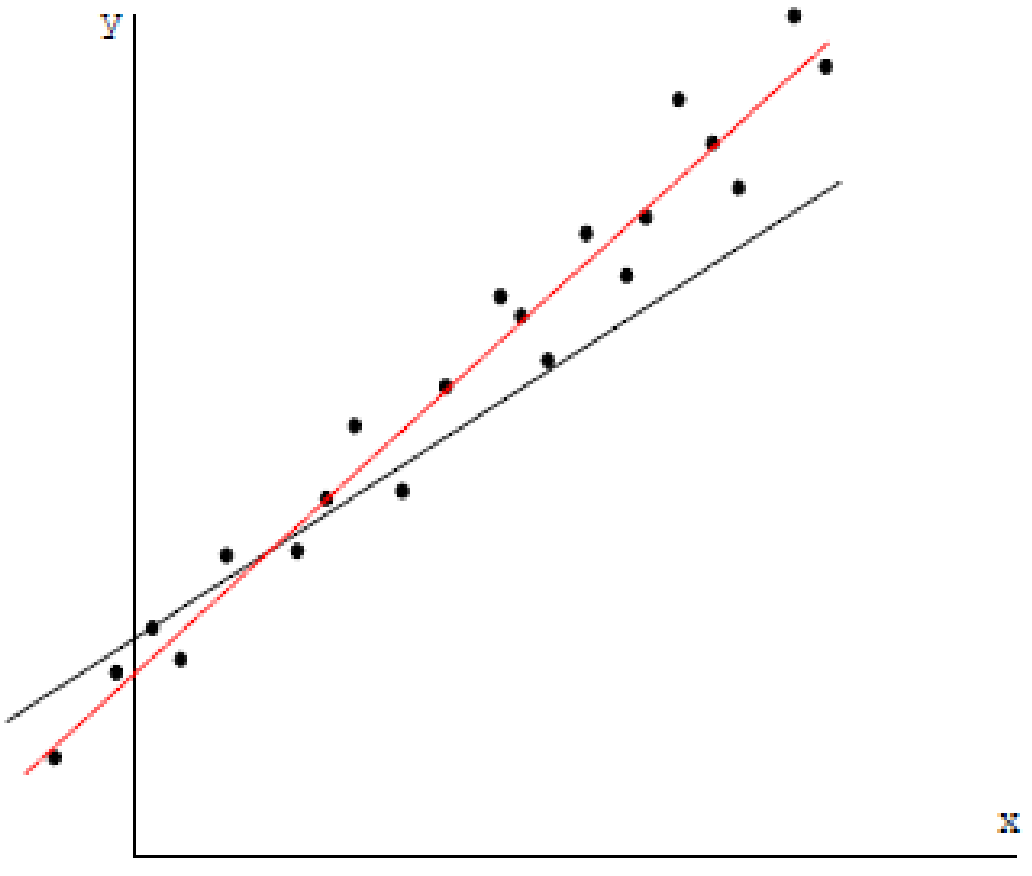
Heteroscedasticity



Autocorrelation



Correlation of disturbance term with explanatory variable



Regression specifications – examples

$$\ln w_{itg} = \alpha_0 + \alpha_1 Edu_{itg} + \alpha_2 Female_{itg} + \alpha_3 Edu_{itg} * Female_{itg} + \mathbf{X}'_{itg}\boldsymbol{\beta} + \gamma T_t + \delta G_g + \varepsilon_{igt}$$

$$y_{itg} = \alpha + \mathbf{X}'_{itg}\boldsymbol{\beta} + \gamma T_t + \delta G_g + \varepsilon_{igt}$$

$$y_{it} = \alpha + \eta d2_t + \tau w_{it} + c_i + u_{it}, t = 1, 2,$$

CONTROLLED (SOCIAL) EXPERIMENTS IN ECONOMICS

Reminder: Overview of Identification Strategies

- Controlled (social) experiment
  - Direct randomisation of treated and untreated
- Natural experiment
  - Finding naturally occuring treated and untreated group that are as similar as possible
- Instrumental variable
  - Finding variadle that assign into treatment but do not affect outcome
- Discontinuity design
  - Probability of treatment is changin discountinously with a characteristic (i.e. age)

What is the goal? Getting rid of selection mechanism

Motivation - remainder

- Effect of a treatment is difficult to measure in the presence of selection
- Selection problem biases the estimate of treatment on outcome
  - Example: How educational program affects wages?
  - Types of selection
- Treated and untreated are usually not the same in unobservables

Introduction

$$y_i = \beta + \alpha_i d_i + u_i \,.$$

- What do we want to achieve?

$$\textbf{R1:}\ \mathrm{E}\left[u_i \middle| d_i = 1\right] = \left[u_i \middle| d_i = 0\right] = \left[u_i\right]$$

$$\textbf{R2:}\ \mathrm{E}\left[\alpha_i \middle| d_i = 1\right] = \left[\alpha_i \middle| d_i = 0\right] = \left[\alpha_i\right]$$

- Treated and untreated after randomization
  - Same in unobserved characteristics
  - Same in their potential gains from a treatment

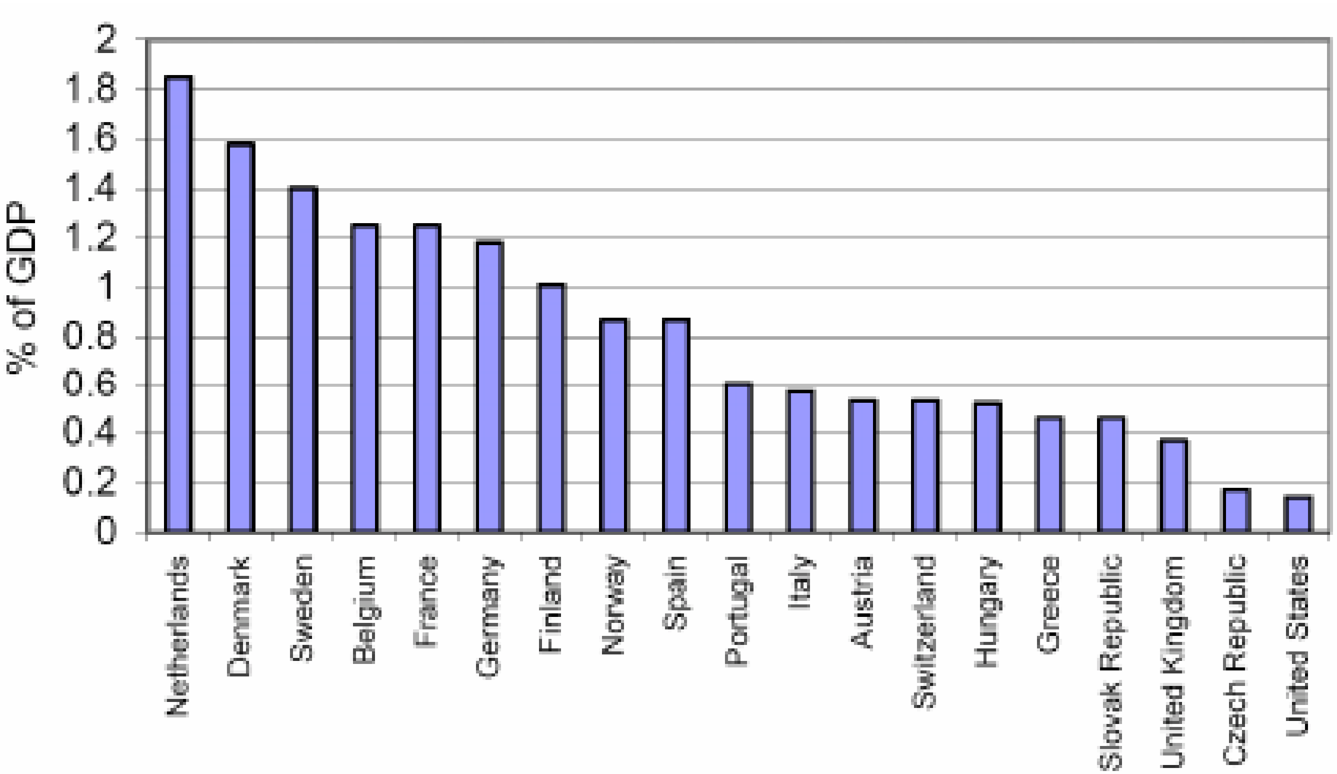
Examples of Implementation

- Active labor market policy programs
  - Evaluation of impact
  - James Heckman (Nobel Prize Laureates – 2000)
- Educational and health economics
  - Peer effects, class size effects
  - Bruce Sacerdote
- Development economics
  - Policy programs
  - Ester Duflo

Active Labor Market Policies (ALMP)

- Main goal: Evaluate the impact of participants on their employment prospective
  - Probability of being employed
  - Wage
- Average treatment effect on treated
  - Only applicants are randomized

Fig. 1.— **Spending in ALMP as percentage of GDP 2002**



Source: OECD (2004).

ALMP – Main Problems with Randomisation

- Attrition – individuals are dropping out from the program not randomly
  - Additional statistical methods and assumptions required
  - It is becoming non-experimental method
- Ethical issues (as opposed to biostatistics)
- Too few individuals willing to participate
- Randomization can change behavior of people – untreated could be subject to different treatment
- Demanding on the organization
  - Job centers might not be willing to organize it
  - Servants want to be involved in the choise of the program for particular unemployed
- Used only on pivotal programs and non-experimental evaluation methods in lager scale programs

Development Economics

- How to improve underperforming educational system?
  - Duflo, Dupaas, Kremer (2009)
  - Reduced teacher-pupil ratio to randomly chosen set of schools in Kenya
  - More resources do not lead to better students performence, parental involvement does
- Microfinance – problem of asymmetric information
  - Karlan and Zinman (2008)
  - Example of field experiment – randomized 58 000 of loan offers to former clients to South African lender
  - Three stages of randomization – different information about the loan are revealed
  - 13-21% of default is due to moral hazard

(Randomization of "offers" also used in labor economics: discrimination on labor market etc…)

Randomization - overview

\+ Significantly reduces differences between treatment and control group

\+ solve problem of selection on unobservables

- Only few people can be randomized
- Problem of attrition
- Limited number of research questions could answered (?corporate sector?)
- May be we should be interested more about selection per se

EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY