Identification (1)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute

1. Introduction

- Topics:
  1. Discrete choice
     - Random utility framework, estimating binary and multinomial regression models
  2. Spatial econometrics
     - Spatial data, autocorrelation, spatial regressions
  3. Identification
     - Research design, IV, OLS, RDD, quasi-experiments, standard errors
  4. Hedonic pricing
     - Theory and estimation
  5. Quantitative spatial economics
     - General equilibrium models in spatial economics

1. Introduction

Wednesday

09:30-10:30	Lecture 1	Discrete Choice I (The random utility framework)
10:45-11:45	Lecture 2	Discrete Choice II (Estimating discrete choice models)
12:00-13:00	Lecture 3	Spatial Econometrics I (Spatial data)
14:00-15:30	Tutorial 1	Assignment 1

Thursday

09:30-10:30	Lecture 4	Spatial Econometrics II (Spatial autocorrelation)
10:45-11:45	Lecture 5	Spatial Econometrics III (Spatial regressions)
12:00-12:30	Lecture 6	Identification I (Research design)
13:30-14:00	Tutorial 2	Discussion of Assignment 1
14:00-15:00	Tutorial 3	Assignment 2

Friday

09:30-10:00	Lecture 7	Identification II (RCTs, OLS, IV, quasi-experiments)
10:00-10:30	Lecture 8	Hedonic pricing I (Theory)
10:45-11:45	Lecture 9	Hedonic pricing II (Estimation)
12:00-12:30	Tutorial 4	Discussion of Assignment 2

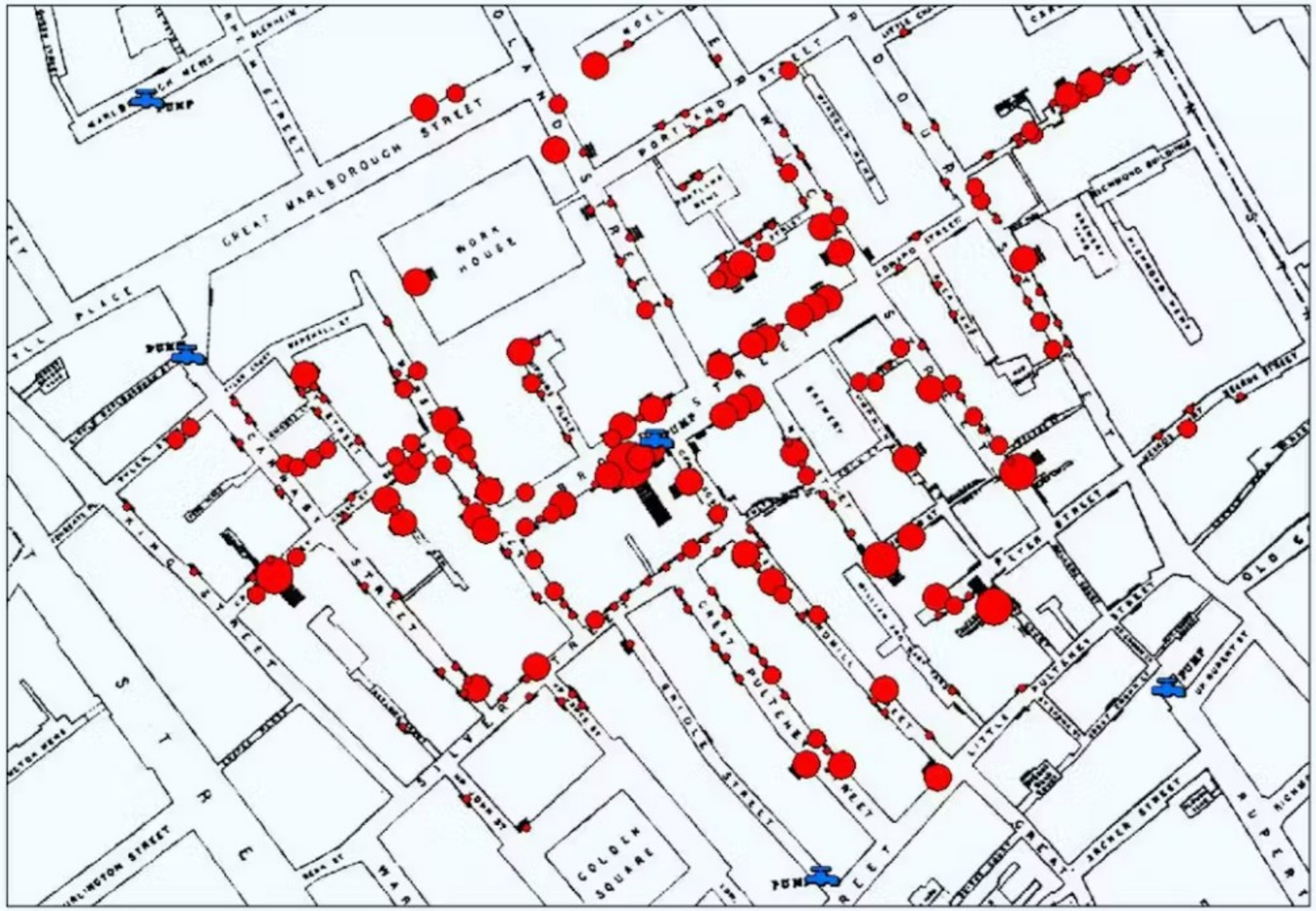
1. Introduction

- Academics usually aim to identify *causal* effects

- Causal effects: one process, *a cause*, contributes to the production of another process
  - the effect of a 'treatment' variable x on an outcome variable y

1. Introduction

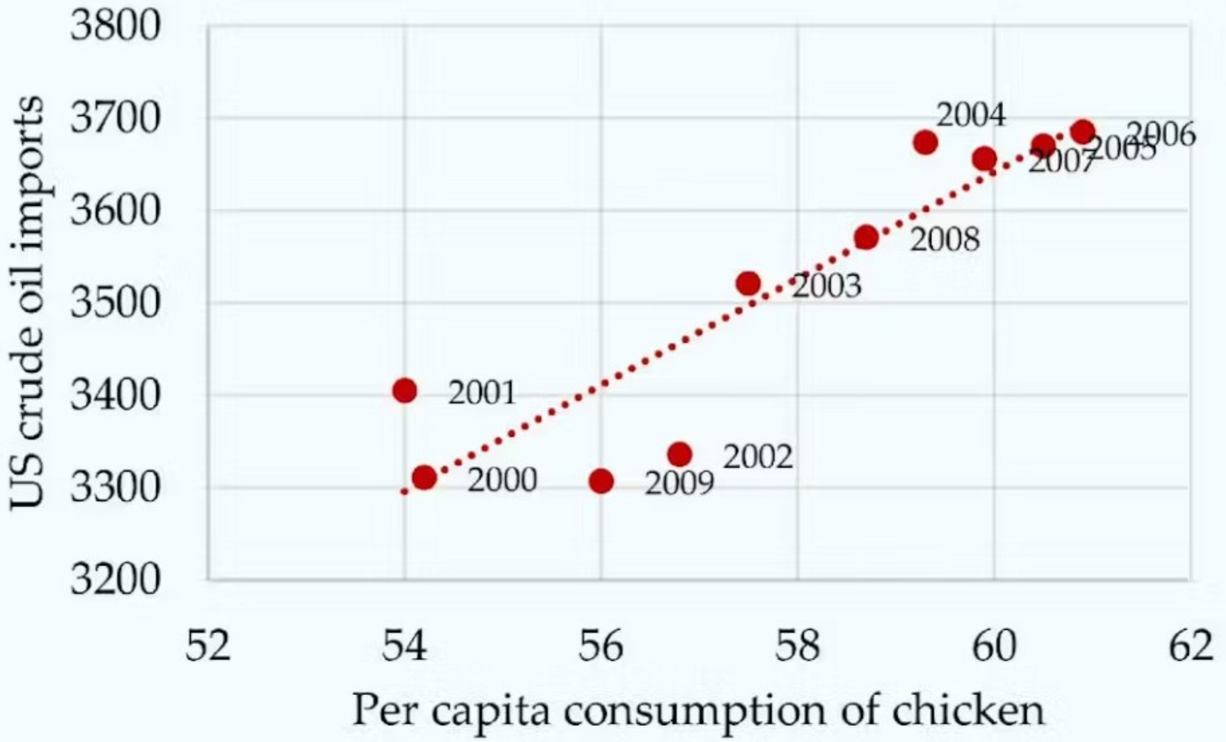
- Interesting correlations



- Spatial correlation of cholera deaths in 1854
  - John Snow
  - Contaminated water…

1. Introduction

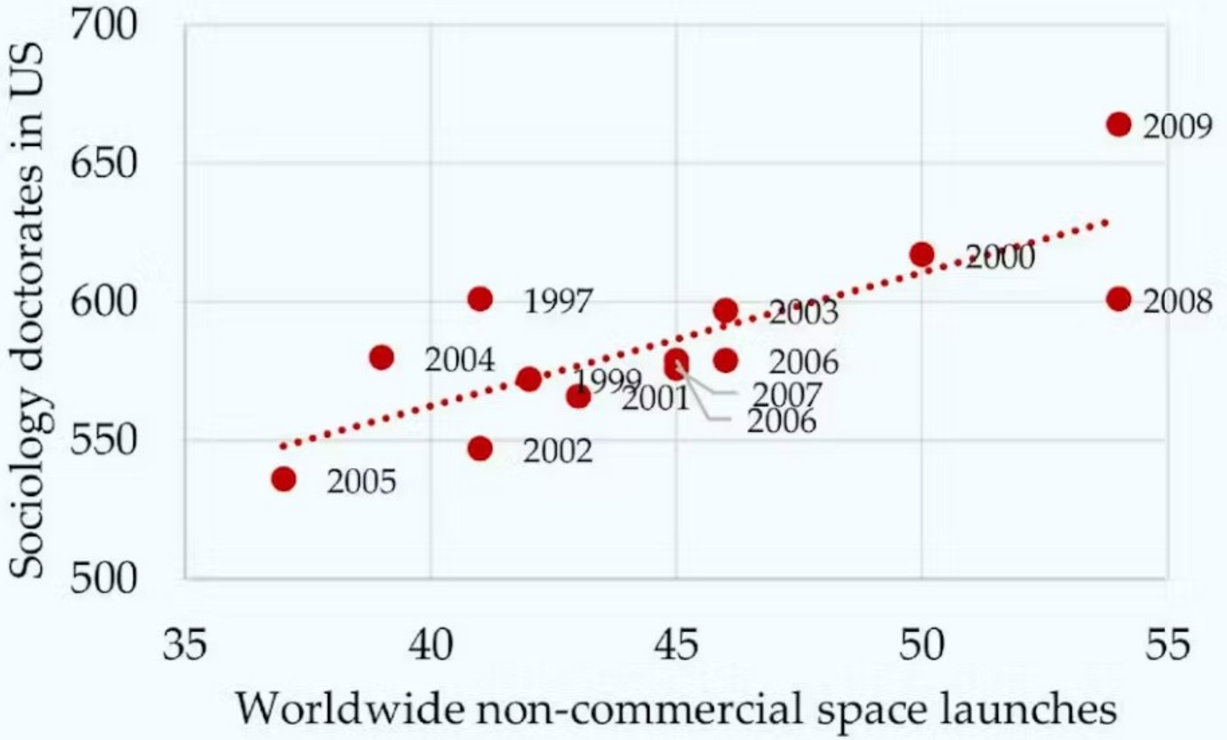
Spurious correlations



$\rho = 0.90$

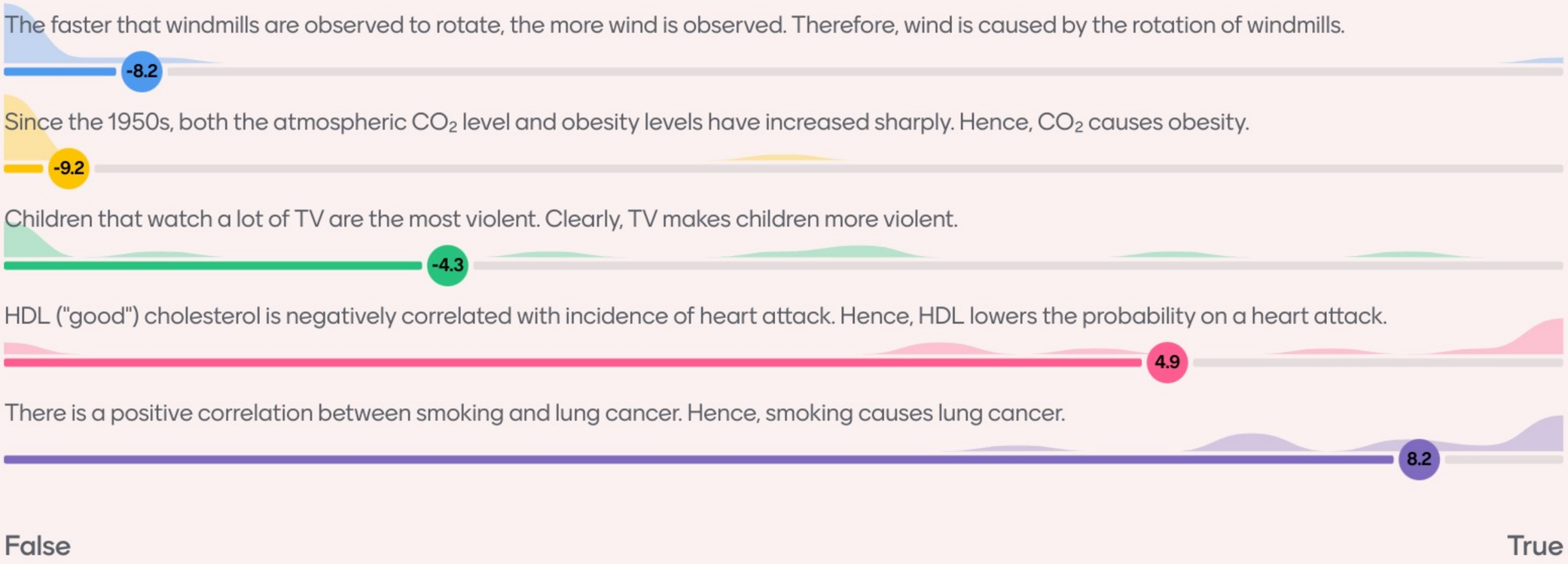
1. Introduction

Spurious correlations



$\rho = 0.79$

Do you think the below relationships may be causal or are spurious correlations? Please judge the following statements.



2. Research design

- 8 steps when undertaking research

1. Formulate your hypotheses
2. Determine the 'treatment' variable(s) and the 'outcome' variable(s)
3. Think of an identification strategy to identify causal effects
4. Select samples, discuss measurement error and provide descriptives
5. Determine functional form of variables of interest
6. Think of different issues in estimating standard errors
7. Estimate model and interpret the results
8. Provide robustness checks of the results

2. Research design

1. Formulate your hypotheses

- Economic hypotheses

- Based on economic theory

- Humans often use *reverse* causal reasoning
  - *"House prices have gone down the last years, but why?"*
  - *Forward* causal inference supplies answers
  - *Reverse* causal inference supplies questions

2. Research design

2. Determine the '<u>treatment</u>' variable(s) and the '<u>outcome</u>' variable(s)

- Define what variables are available in your data
- Focus on one (or a few) x variable(s) and one (or a few y variables
- Think about expected order of magnitude

3. Think of an identification strategy to identify causal effects

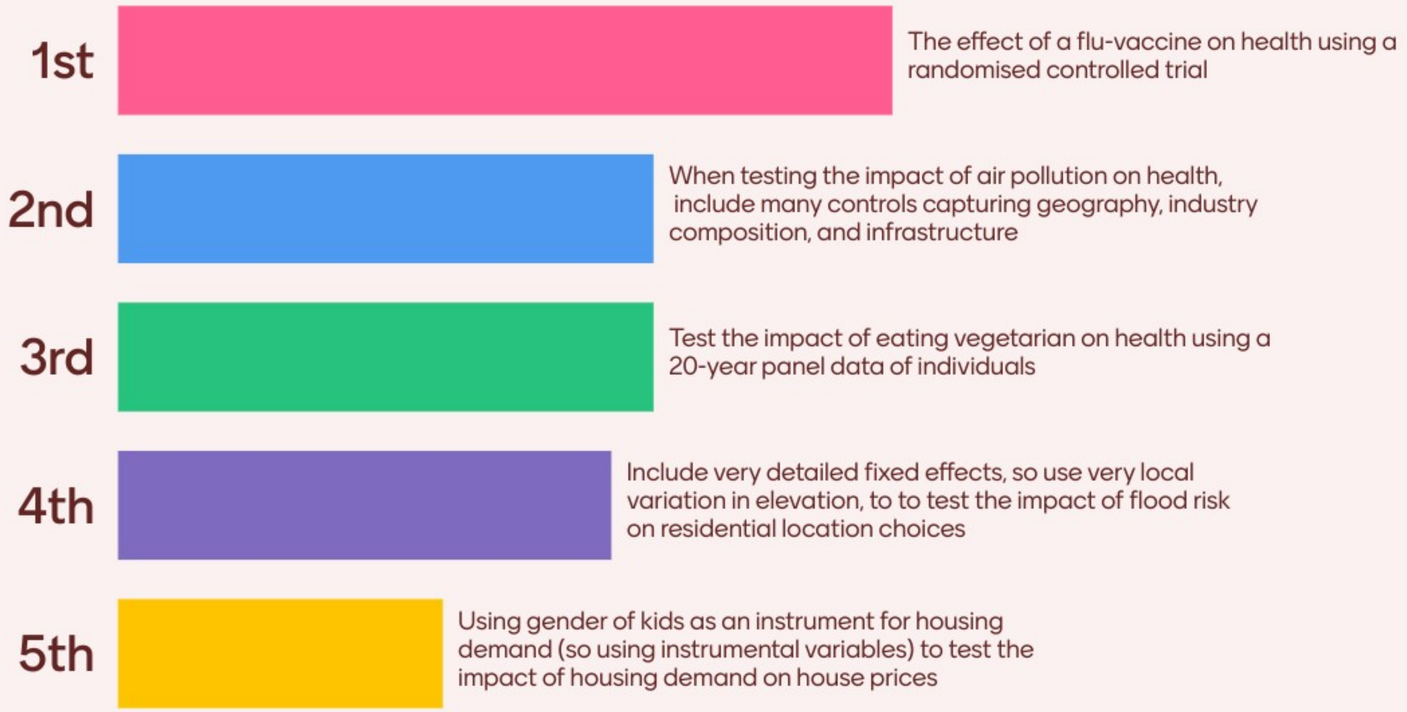
- What is your 'treatment' group and what is your 'control' group?

- Discuss endogeneity issues
  - Might there be a selection effect?
  - What are potential unobserved factors? Are these correlated with the treatment status?
  - Reverse causality?
  - *(Measurement error?)*

3. Think of an identification strategy to identify causal effects

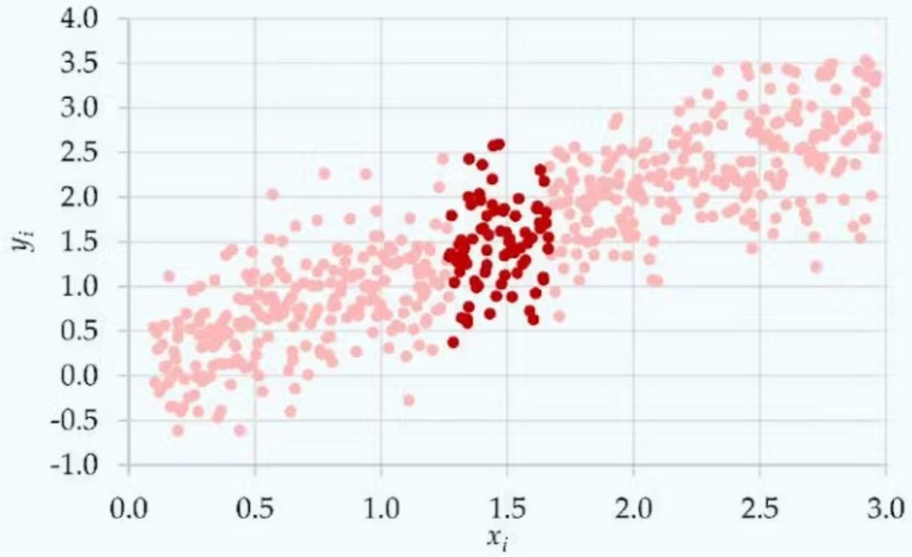
- Define the appropriate econometric methods
  - Discuss the identifying assumptions at length!

Please rank the identification strategies in how convincing these are in your opinion



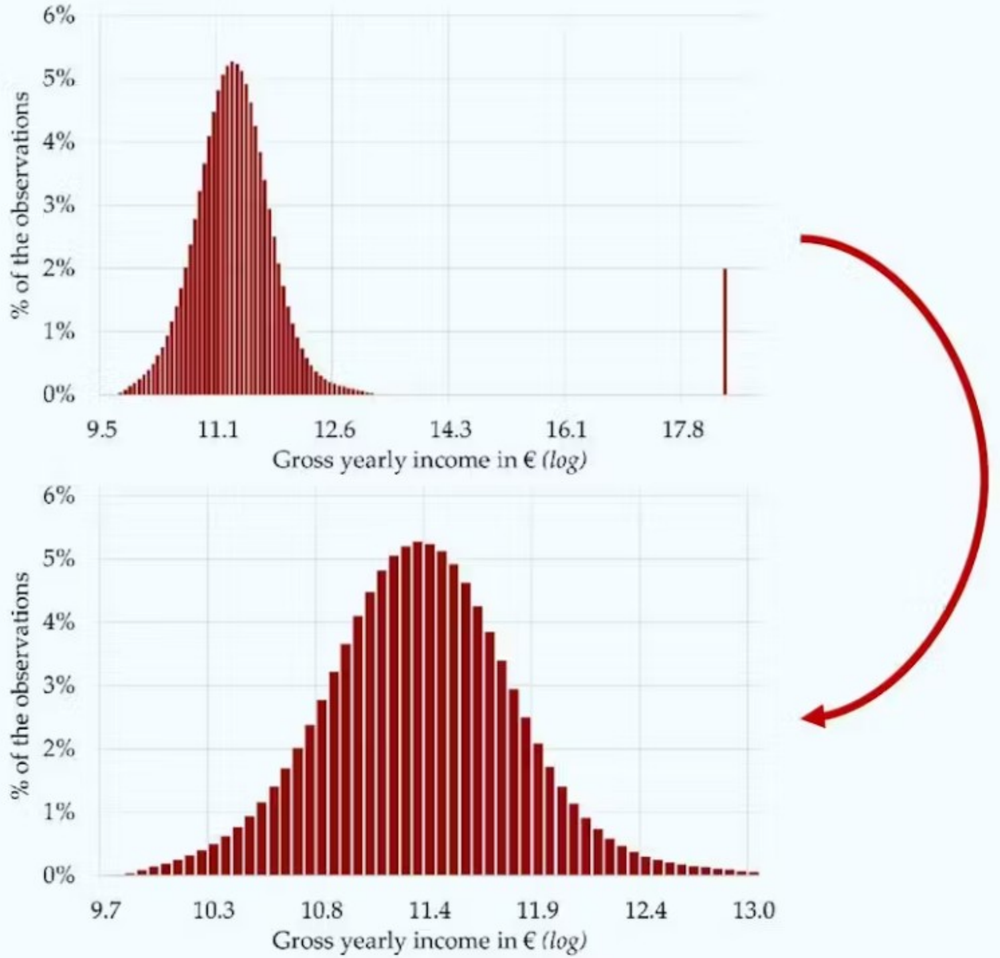
4. Select samples, discuss measurement error and provide descriptives

- Should you use the full dataset?
- Variance in x is necessary!



4. Select samples, discuss measurement error and provide descriptives

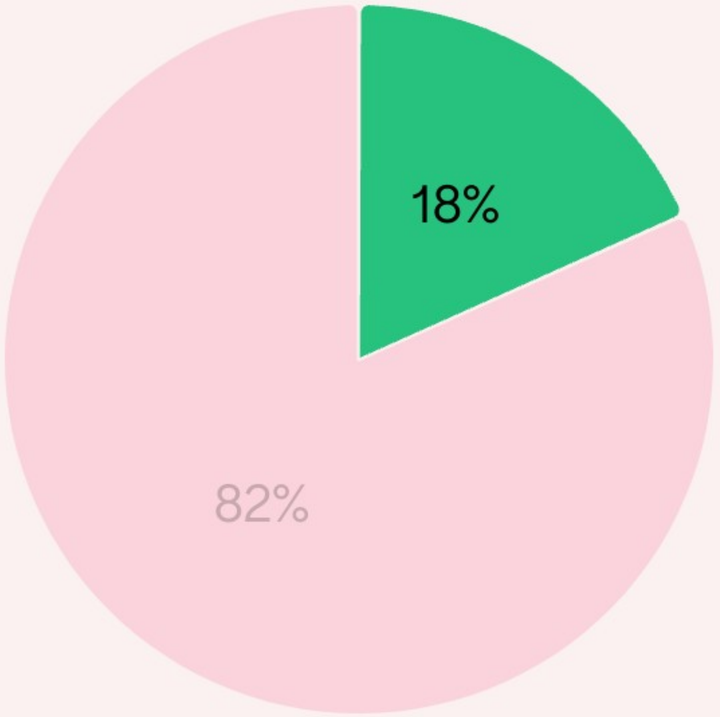
- **Data cleaning**



4. Select samples, discuss measurement error and provide descriptives

- Measurement error is present in many datasets

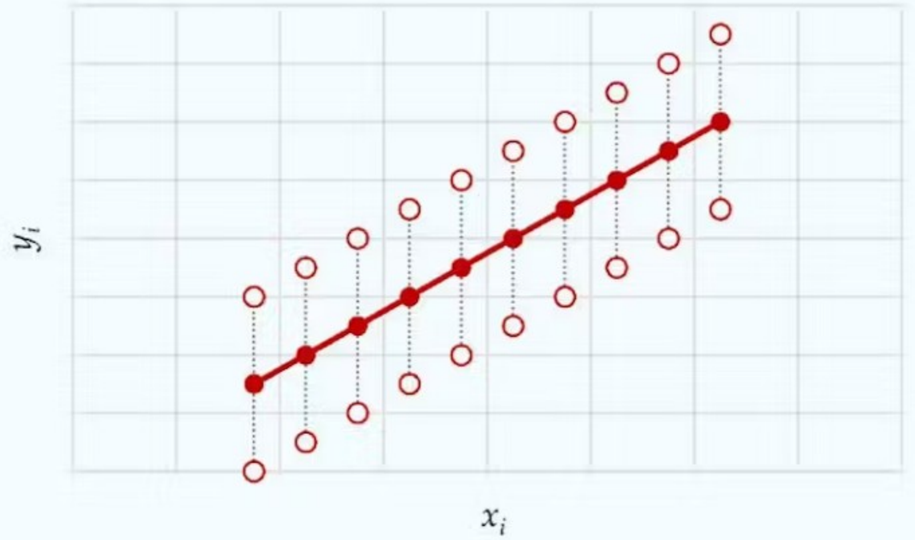
What statement is true with respect to measurement error?



0%	Random measurement error in both the outcome variable and explanatory variables are no big issues	×
0%	Random measurement error in the outcome variable leads to estimates that are biased towards zero	×
18%	Random measurement error in the explanatory variable leads to estimates that are biased towards zero	✓
82%	Both random measurement error in the outcome and explanatory variable leads to biased estimates	×

4. Select samples, discuss measurement error and provide descriptives

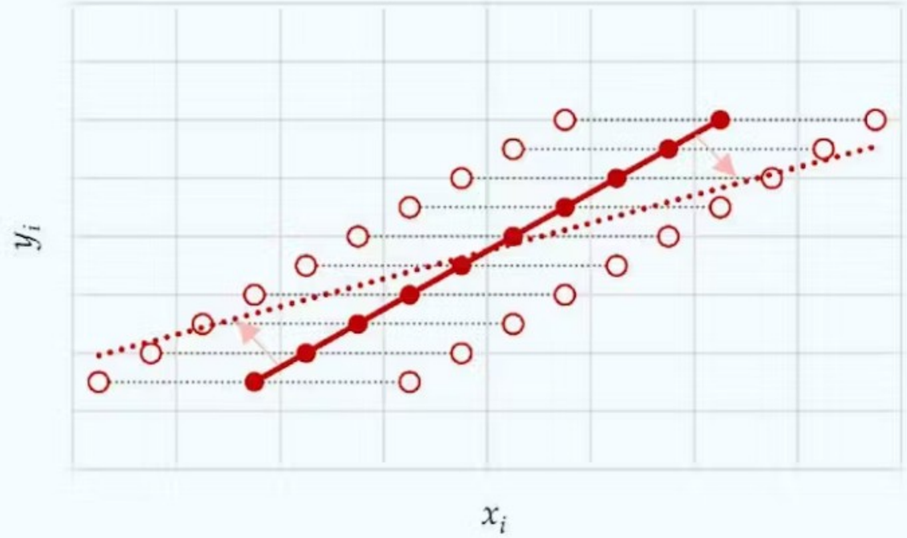
- Random measurement error in y is not so much of a problem



- $y_i^* - u_i = \beta x_i + \epsilon_i \rightarrow y_i^* = \beta x_i + (\epsilon_i + u_i)$

4. Select samples, discuss measurement error and provide descriptives

- Random measurement error in x biases the effect towards zero



- $y_i = \beta(x_i + u_i) + \epsilon_i$ ➔ $\beta \to 0$ if u_i is large

5. Determine functional form of variables of interest

- The specification of $f(\cdot)$ is referred to as the functional form

$$y_i = f(x_i, c_i, \epsilon_i)$$

- Often a linear functional form is assumed:

$$y_i = \beta x_i + \gamma c_i + \epsilon_i$$

Can you allow for non-linear effects in Ordinary Least Squares?

Yes

0

No

5. Determine functional form of variables of interest

- Economists are often interested in elasticities
  - Elasticity is percentage change of y in response to a change in x
  - $\frac{\partial y}{\partial x}\frac{x}{y}$

- They therefore often estimate log-linear regressions:

$$\log y_i = \beta \log x_i + \gamma c_i + \epsilon_i$$

  - because $\beta = \frac{\partial \log y}{\partial \log x} = \frac{\partial y}{\partial x}\frac{x}{y}$

Show that $\beta = \frac{\partial \log y}{\partial \log x} = \frac{\partial y}{\partial x}\frac{x}{y}$, so β is an elasticity in a log-linear regression

I am ready!

5. Determine functional form of variables of interest

- **Elasticities**
  - **Let's assume** $\log y_i = \beta \log x_i + \gamma c_i + \epsilon_i$
  - $y_i = e^{\beta \log x_i + \gamma c_i + \epsilon_i}$
  - $\frac{\partial y_i}{\partial x_i} = e^{\beta \log x_i + \gamma c_i + \epsilon_i} \frac{1}{x_i} \beta$
  - $\frac{\partial y_i}{\partial x_i} = \frac{y_i}{x_i} \beta$
  - $\beta = \frac{\partial y_i}{\partial x_i} \frac{x_i}{y_i}$, **which is an elasticity**

5. Determine functional form of variables of interest

- When use logs?
  - Economic theory
  - Residuals have a skewed distribution
  - Heteroscedasticity
  - Different unit sizes

6. Think of different issues in estimating standard errors

- Whether β is statistically significant depends on standard error
  - The smaller the standard error, the more precise your conclusions are

- Issues to bear in mind…
  - Should you cluster your standard errors?
  - Is heteroscedasticity a problem?
  - Is there serial/spatial autocorrelation?

7. Estimate model and interpret the results

- Use statistical software to estimate your model

- Usually we are interested in marginal effects
  - How much does y change (in units or %) when x change with one unit (or %)
  - $\frac{\partial y}{\partial x}$ (in levels) or $\frac{\partial y}{\partial x}\frac{x}{y}$ (in %)

7. Estimate model and interpret the results

- Properly interpret β and its statistical significance
  - "When x increases by 1 *(units)* y increases by .. *(units)*. This effect is statistically significant at the ...% level."
  - Specify the units!

7. Estimate model and interpret the results

- Statistical hypothesis testing is dependent on *statistical significance*

- Economic significance ≠ statistical significance
  - A large effect may be imprecise
  - A small, but stat. sign. effect may be irrelevant

- Always discuss both economic and statistical significance

- See McCloskey and Ziliak (1996)

7. Estimate model and interpret the results

- Make sure how your variables are measured
  - logs, dummies, etc.

Specifications:	x	$\log x$
y	$\boldsymbol{y = \beta x + \eta}$ $\hat{\beta} = \frac{\partial y}{\partial x}$ $x \uparrow 1 \rightarrow y \uparrow \hat{\beta}$	$\boldsymbol{y = \beta \log x + \eta}$ $\hat{\beta} = \frac{\partial y}{\partial \log x}$ $x \uparrow 1\% \rightarrow y \uparrow \hat{\beta}/100$
$\log y$	$\boldsymbol{\log y = \beta x + \eta}$ $\hat{\beta} = \frac{\partial \log y}{\partial x}$ $x \uparrow 1 \rightsquigarrow y \uparrow (\hat{\beta} * 100)\%$ *(for marginal changes in x)*	$\boldsymbol{\log y = \beta \log x + \eta}$ $\hat{\beta} = \frac{\partial \log y}{\partial \log x}$ $x \uparrow 1\% \rightarrow y \uparrow \hat{\beta}\%$

7. Estimate model and interpret the results

- **Note on larger changes in x in log-linear regressions.**

- **Let's assume the model** $\log y = \beta x + \epsilon$**, with** $x \in 0.1$**.**
  - **Example: dummy variables**
  - **Halvorsen & Palmquist:** $x \uparrow 1 \rightarrow y \uparrow ((e^{\widehat{\beta}} - 1) * 100)\%$

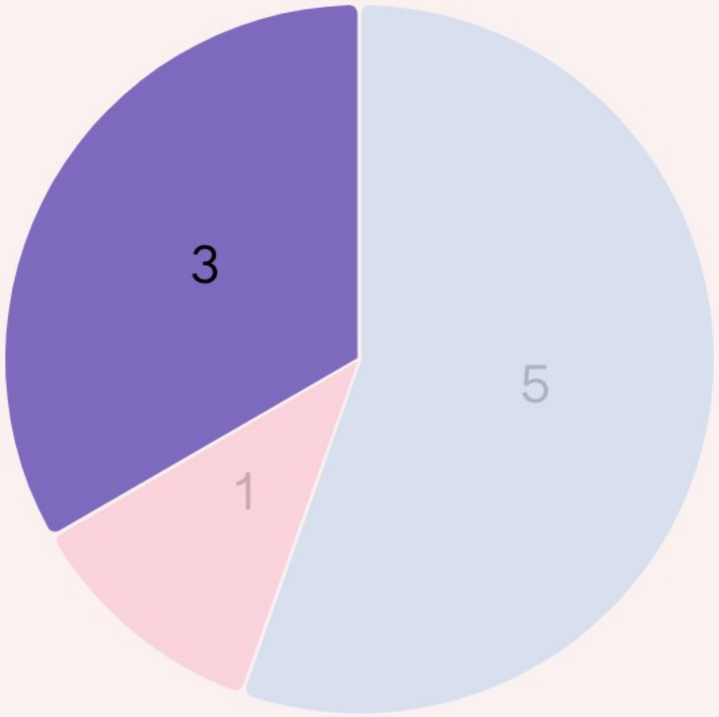
7. Estimate model and interpret the results

- **Note on larger changes in x in log-linear regressions.**

- **Let's assume the model** $\log y = \beta x + \epsilon$**, with** $x \in 0.1$**.**
  - **Example: dummy variables**
  - **Halvorsen & Palmquist:** $x \uparrow 1 \rightarrow y \uparrow ((e^{\hat{\beta}} - 1) * 100)\%$

→ **How much does y increase when $\hat{\beta} = 0.01$ and $\hat{\beta} = 1$?**

You estimate with $x \in \{0, 1\}$. What is the change in y_i when $\beta = 0.01$ and when $\beta = 1$?



5	0.01 unit and 1 unit	×
0	0.01 unit and 1.72 units	×
0	1% and 100%	×
1	1.72% and 172%	×
3	1% and 172%	✓

8. Provide robustness checks of the results

- You make many somewhat arbitrary choices

- Test for sensitivity of your results with respect to these choices
  - … sensitivity analysis

3. Summary

Today:

- Economists are generally interested in *causal* effects

- 8 steps when undertaking research
  1. Formulate your hypotheses
  2. Determine the 'treatment' variable(s) and the 'outcome' variable(s)
  3. Think of an identification strategy to identify causal effects
  4. Select samples, discuss measurement error and provide descriptives
  5. Determine functional form of variables of interest
  6. Think of different issues in estimating standard errors
  7. Estimate model and interpret the results
  8. Provide robustness checks of the results

Identification (1)

Applied Econometrics for Spatial Economics

Hans Koster

Professor of Urban Economics and Real Estate

VU VRIJE UNIVERSITEIT AMSTERDAM

tinbergen institute