



Stockholm
University

Pär Nyman

Methods in Political Science
27 April 2026

The Experimental Paradigm

An introduction to design-based causal inference

Background

- Strong trend towards more experiments.
- The development has been most evident among economists and in top journals, but also noticeable in investigative activities and political debates.
- What I have missed the most from my education.
- The lecture's purpose is to introduce experiments and lay a foundation for the rest of this bloc.
 - Familiarize yourselves with the trend, both to be inspired and to relate to it.
 - Eye-opener and source of inspiration. Maybe even plant a seed for the master thesis.

Disposition

1. Introduction
2. Logic of experiments
3. Typology
4. Controlled experiments
 - Laboratory experiments
 - Field experiments
 - Survey experiments
5. Natural experiments
 - Standard natural experiments
 - Regression discontinuity designs (RD)
 - Instrumental variables (IV)
6. Criticism of experiments
7. Lessons on design

Disposition

- The focus is on examples of conducted experiments.
- Examples from various fields: gender quotas, fairness norms, political campaigning, climate mitigation, economic development, and civil wars.
- Intuitive and non-technical walkthrough.
- After lunch: more technical but still general.
- Upcoming weeks: more specific techniques.

The logic of experiments

Correlation is not causation

- In quantitative studies, it is easy to establish *covariation*.

For the relationship to represent a causal effect of x on y , the following are also required (exogeneity assumption):

- No doubt about *causal direction*: y cannot affect x .
- *No OVB*: We must control for all confounders.

The logic of experiments

Causal direction

- Sometimes the causal direction is rather clear (biological sex affects earnings, distance to the equator affects undernourishment, etc.).
- Most problems have a less obvious place in the causal chain.
- A fundamental problem in many quantitative studies.

The logic of experiments

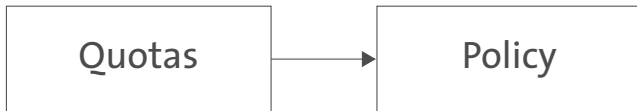
Eliminating OVB is almost impossible

- We do not know what all the confounders are.
- Even if we knew them, it is not obvious how we should measure them or if it is practically possible.
- Even if we knew and could measure all underlying variables, we do not know how to control for them. The linear and additive regression model is just one possibility.

The logic of experiments

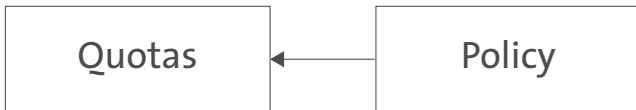
- Let's assume we want to examine the relationship between descriptive and substantive representation.
- Specifically, we wonder if gender quotas in parliaments affect implemented policy.
- We might do this by comparing policy in countries that have quotas with countries that do not have quotas.

The logic of experiments



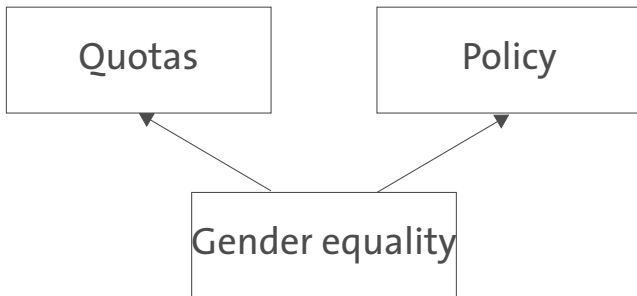
The logic of experiments

Risk of reverse causality!



The logic of experiments

Obvious risk of spurious relationships!



The logic of experiments

- Difficult to know what we need to control for. Gender equality? Women's political power? Voters' political preferences?
- Even if we knew, how can we *validly* measure something as abstract as women's political power?
- And if we could measure, how do we know what the relationship looks like? Can we really assume that all effects are *linear* and *additive* (i.e., no non-linearities or interactions)?

The logic of experiments

- The idea behind experiments is that if the values on x are completely random, then we can exclude the possibility that other variables have affected the independent variable.
- If no variable can affect the independent variable, there is no risk of reverse causality or spurious relationships.
- Researcher's dream: What if we could let a dice decide which of the world's countries apply gender quotas to political appointments!

The logic of experiments

- Since 1993, the chairperson of a third of the councils in parts of India (Rajasthan and West Bengal) must be a woman. Which villages are affected is randomly determined.
- Chattopadhyay and Duflo (2004) use this to study the effect of the reserved seats on the policy content.
- They find that villages required to have a female chair invest more in political projects that are popular among women, such as access to drinking water and various welfare programs (widow's pension, parental insurance, etc).

Short discussion: why focus on the women in reserved seats?

Typology

- **Controlled experiments** are designed by the researcher, who control the random assignment of subjects to a treatment or control group.
- **Natural experiments** are situations that arise outside the researcher's influence, but that resemble controlled experiments because the treatment allocation can be considered *as if* it were random.

Typology

Controlled experiments

- Laboratory experiments
- Field experiments
- Survey experiments

Natural experiments

- Standard natural experiments
- Regression discontinuity designs (RD)
- Instrumental variables (IV)

Typology

- When strict, I use *design-based causal inference* (causal inference from design, not from analysis) when referring to both kinds of experiments.
- Shorter concepts are more popular, but note:
 - Experiments sometimes refer to only controlled experiments.
 - Natural experiments are *observational*, but that term usually refers to studies without an experimental design.
 - Watch out for concept stretching (natural experiment is not a well-defined term).

Controlled experiments

- A controlled experiment is designed by the researcher.
- The researcher therefore has control over which units of analysis are exposed to a treatment and how the allocation (assignment) of this is done.
- Almost all natural science experiments are controlled experiments.
- For ethical and practical reasons, it is difficult to experiment with the world as a social scientist. To make it possible, we must limit the scope of our study (easier to study blood donation than revolutions).

Laboratory experiments

- Experiment participants appear at a location controlled by the researcher.
- Has long been popular among psychologists.
- The most common applications are testing assumptions about human behavior and hypotheses from game-theoretical reasoning.
- Strength: Complete control over the experimental situation.
- Weakness: Lack of generalizability.
 - The experimental situation is different from reality.
 - Participants know they are participating in a game.
 - Participants are often students.

Laboratory experiment: The Ultimatum Game

- Two players have to divide a sum of money between them.
- The first player proposes a distribution and it is then up to the second player to accept the distribution.
- If the second player accepts the distribution, both players receive the proposed amount. If the second player instead rejects the offer, neither player receives any money at all.
- The game is only played once, so accept is always the money-maximizing strategy.

Laboratory experiment: The Ultimatum Game

- Results: It has been found that the second player often rejects low bids, even though it means the player gets no money at all instead of at least a small amount.
- Interpretation: People follow norms of fairness and punish those who do not follow these norms, even if the punishment is costly to themselves.

Field experiments

- Field experiments are conducted in a natural environment, such as a school, a hospital or at people's homes.
- The treatment is often delivered by people who work in this environment (e.g., teachers, doctors, canvassers).
- Participants are usually unaware that they are taking part in an experiment.
- Better external validity than lab experiments, but the researcher tends to have less control over the setting.

Field experiment: Canvassing

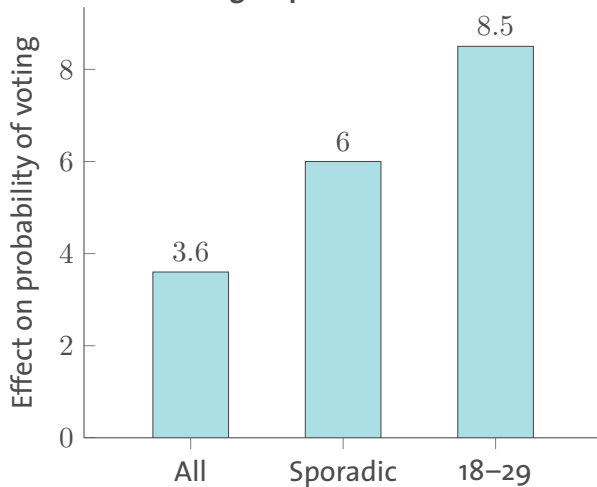
- Does door-to-door canvassing increase voter turnout in Swedish elections?
- We could ask the parties where they have knocked on doors and use that information as an independent variable in a regression analysis.
- But exogeneity assumption would be violated if they mobilize where turnout is low or where they lost many voters in the previous election.

Field experiment: Canvassing

- Before the 2014 European Parliament election, I cooperated with the Social Democrats in Södermanland.
- I randomly assigned 7,500 households to treatment and control groups. The Social Democrats knocked on doors in the treatment group.
- After digitizing the electoral rolls and comparing the two groups, we saw that door-knocking had increased voter turnout, and estimated the effect to 3.6 percentage points.

Field experiment: Canvassing

Effect in different groups



Survey experiments

- In survey studies, it is easy to modify the wording of a question or the background information provided to respondents.
- If different wordings are randomly assigned, they cannot be influenced by underlying variables.
- Survey experiments are often used to improve survey methods. E.g., we can test how the order of questions or compensation to respondents affects the answers.
- They are also used when we do not want to ask directly. Let us look at three common types.

Survey experiments: Vignette

Vignette experiment about fuel taxes

- Respondents are presented with scenarios that vary randomly before answering questions about them.
- Factorial designs (allowing multiple aspects of the scenario to vary) are common.
- In a recent vignette experiment, we describe a hypothetical fuel tax increase and vary:
 - Who proposed it.
 - How the revenues are distributed.
 - Information about average mileage and the historical development of fuel costs per km.
- Responses are then linked to registry data on earnings, vehicle mileage, fuel consumption, etc.

Survey experiments: Vignette

Vignette experiment about fuel taxes

To reduce climate emissions from traffic, the use of fossil fuels needs to decrease. Parties on both the right and the left have therefore proposed increasing the tax on petrol and diesel by 2 SEK per liter. The revenue from the tax would be used to provide financial compensation to all adults living in rural areas, amounting to 8,000 SEK per year.

An average household in Sweden drives about 10,000 kilometers per year. Thanks to more fuel-efficient cars, fuel costs for a typical car have decreased by about 4 SEK per 10 kilometers since 2000.

Respondents were then asked to rate the proposal along several dimensions.

Survey experiments: Vignette

Vignette experiment about fuel taxes

- We are still processing the results, searching for an angle.
- But fairness appears more important than personal cost.
 - Strong correlation between perceived fairness of the reform and general opinion about it.
 - Large effects from revenue recycling perceived as fair (targeting deserving and affected groups).
 - Small effects from both subjective and objective economic consequences for the household.

Survey experiments: Conjoint

Conjoint experiments

- Respondents choose between two profiles.
- The characteristics of these profiles are randomly varied.
- In this experiment, we study how attractive competence is when parties choose cooperation partners.
- In particular, cognitive ability, which we operationalize as “well-read and adept”.

Survey experiments: Conjoint

How well a collaboration between two people will work out is hard to know before you have gotten to know each other. But if you follow your gut, which of the following two politicians do you think you would prefer to collaborate with on the local council?

Person 1:

Erik is a 41-year-old Green Party representative and has served on the local council for two terms. He works as a lawyer at a large company. Erik plays soccer in an amateur soccer team. He has been active in the Green Party since his twenties.

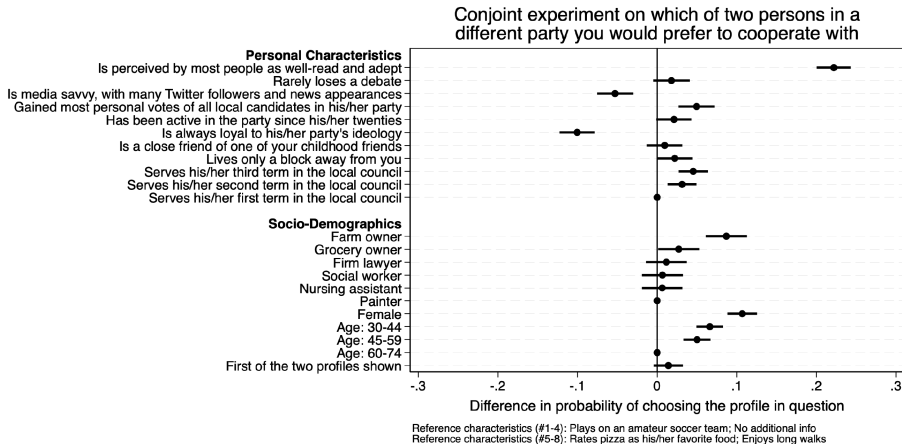
Person 2:

Karin is 60 years old and has served on the local council for three terms. She represents the Green Party and is a farmer with her own farm. Karin is perceived by most people as well-read and adept. She is always loyal to her own party's ideology.

I think I would prefer to collaborate with...

- ☐ Person 1
- ☐ Person 2

Survey experiments: Conjoint



Survey experiments: List

List experiments / unmatched count

- Most survey experiments encourage more truthful responses, but some do so more directly.
- One approach is to assign respondents a randomizing task (e.g., rolling a die) that determines which question they answer. Then one can be sensitive and one should have a known probability (e.g., is it Friday today?).
- A more common technique is to ask how many statements respondents agree with, adding one additional item for the treatment group, and interpret the difference as the estimated share who agrees with the extra statement.
- In a 2024 paper, we study why local governments do not invest more. Are electoral incentives an explanation?

Survey experiments: List

Everyone received these four statements:

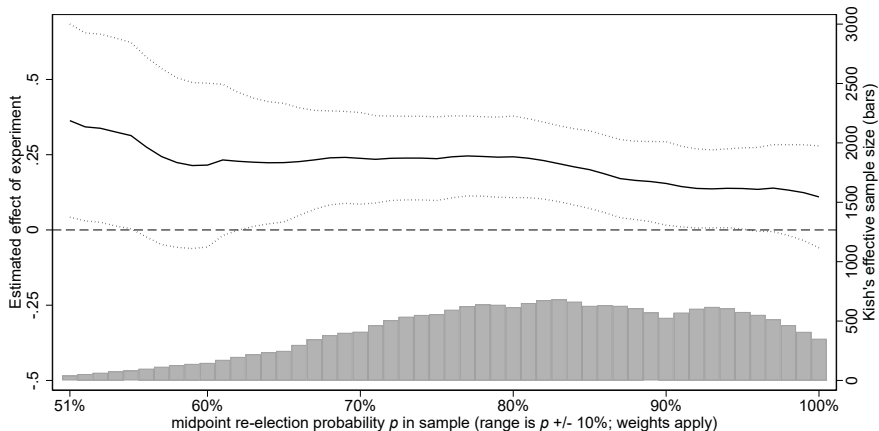
- No cost–benefit analysis of the bike path has been done.
- The bike path would pass through a protected biotope.
- The proposal was not originally raised by your party.
- A petition indicates that motorists are critical of the project.

The treatment group also received a fifth:

- Construction will not begin until after the next election.

The average was 0.17 higher in the treatment group.

Survey experiments: List



Survey experiments

Easy to implement, harder to sell

- Stated responses in surveys do not necessarily translate into real-world behaviour, which makes effect magnitudes harder to interpret and external validity low.
- Their low cost is probably a double-edged sword that further reduces the enthusiasm for them.
- I primarily use them to complement studies based on other designs (partly because my survey experiments lack an idea).
- The low cost makes them well suited for smaller projects, such as master's theses.

Natural experiments

- For practical or ethical reasons, it is often not desirable to carry out a controlled experiment.
- In a natural experiment, neither the treatment nor the instruction is under the control of the researcher. Instead, we search for situations that resemble controlled experiments because the allocation of a treatment is random or can be considered random for various reasons.
- Three examples: standard (for lack of a better word), regression discontinuity designs, and instrumental variables.

Standard natural experiments

- Standard natural experiments refer to situations where there has been a deliberate random allocation of a treatment or where the allocation for some other reason can be considered random.
- Examples of this are lotteries and certain political reforms.
- Speculation: Politicians are overly reluctant to lotteries (housing, schools, etc.), but sometimes use them to avoid political responsibility for difficult decisions (e.g., the Vietnam draft lottery).

Standard: The effect of income

- Income is something that social scientists use to explain political attitudes, health, well-being, political participation, and many other things.
- These relationships often appear strong, but it is difficult to know if they are causal or spurious.
- It would be absurd if we started handing out money to randomly selected people to see how this income affects various outcomes.
- Or would it?

Standard: The effect of income

- Doherty, Green, and Gerber (2006) compare people who have won large sums in lotteries with people who have played just as much but without winning.
- They find that these people are more negatively inclined towards property tax and redistribution.
- Other studies show that lottery winnings affect health (Lindahl 2005) and happiness (Gardner and Oswald, 2001).

Regression discontinuity designs (RD)

- Based on the presence of a sharp threshold, such that analysis units on one side of the border are subjected to a treatment that the analysis units on the other side of the threshold do not receive.
- If it is not possible for the analysis units to influence whether they are placed just above or below the threshold value, the allocation around the threshold – and thus the allocation of the treatment – can be considered random.

Regression discontinuity designs (RD)

- Typical thresholds are the votes required for being elected to an assembly (compare just ticked off with just outside), the age to vote (compare people born the day before with people born after election day), and many kinds of tests (but only valid if no adjustment is made for those close to the threshold).
- Regression discontinuity design (RD or RDD).

RD: Does SD affect asylum policy?

Classic question: the effect of political representation

- In a Bachelor's thesis I supervised, the author investigates whether the number of municipal mandates for the Sweden Democrats affects the municipality's refugee reception.

Covariation is not enough

- Reverse causation: the number of refugees can affect support for the Sweden Democrats.
- Omitted variable bias: both variables can be influenced by political preferences, racism, etc.

RD: Does SD affect asylum policy?

- Szulkin compares municipalities where the Sweden Democrats were lucky and won the last mandate with municipalities where they were unlucky with the mandate distribution.
- No difference in voting share – random whether they got that mandate or not.
- Result: More mandates to the Sweden Democrats reduce refugee reception.

Instrumental variables (IV)

- A slightly trickier method involves using a third variable to find variation in our explanatory variable (x) that cannot be affected by the dependent variable (y) or other variables that affect the dependent variable.
- We then choose a third variable (z) – an instrument – which affects x but not y and is not affected by other relevant variables.

Instrumental variables (IV)

- The instrument is used to calculate the expected values of x . That part of the variation in x has only been affected by the instrument.
- These predictions of x are then used to estimate the causal relationship with y .

$$\hat{y} = a_1 + b_1 \hat{x}$$

$$\hat{x} = a_2 + b_2 z$$

- Can be used for experiments with non-perfect compliance.

IV: Weather and growth

- Let's assume that we want to study whether economic growth affects the risk of civil war.
- An obvious problem is that civil war probably affects economic growth.
- There is also a risk of spurious relationships.

IV: Weather and growth

- Miguel, Satyanath, and Sergenti 2004 use the variation in rainfall as an instrument to predict growth.
- They study 41 African countries between 1981 and 1999, where rain is crucial for agriculture.
- They find that a reduction in economic growth (caused by reduced rain) of 5 percentage points increases the likelihood of internal conflict by 50 percent.

IV: Weather and growth

- The risk with their design is that rainfall might affect the likelihood of civil war without going through economic growth (this would violate the exclusion restriction).
- The authors examine some possible mechanisms but find no support for them:
 - It is not flooded roads that hinder troop movements.
 - There is no strong relationship between rainfall and heat.
- We will get back to RD and IV during Week 7 and 8.

Critique: Relevance

- It can be difficult to generalize the results.
- Limits the research question to narrow phenomena and questions of dubious relevance.
- Often studies only a limited part of a relationship.
- Studies effects, but tells nothing about the mechanisms. If the effects are uninteresting, the studies therefore lack theoretical relevance.

Critique: Ethical problems

- Experimental-like design of political reforms conflicts with principles of equal treatment.
- These problems are often more serious in poor countries.
- Experiments can harm third parties.
 - And they can harm the researcher in return!
- Survey experiments often provide false information.
- Time-consuming ethical review.

Lessons about design

- Design is the *structure of or plan for* a study.
- A poor design can rarely be saved by a very advanced statistical model.
- A good design, on the other hand, goes far, even with such simple quantitative methods as comparing the mean value in two groups (e.g., treatment and control group).

Lessons about design

Good natural experiments often involve a combination of quantitative and qualitative methods.

- Experiments are usually analyzed with quantitative methods.
- Finding and evaluating natural experiments requires the kind of detailed knowledge we often associate with – or obtain through – qualitative research. How do we know that the quota system in India was completely random? By reading the law text and interviewing decision-makers.

Lessons about design

Where are we headed?

- The academic world will tire of theoretically irrelevant experiments, but they are important in evaluating reforms.
- The understanding of experimental design also outside academia will increase, hopefully leading to more experiments run by governments and civil society.
- As with quantitative and qualitative methods, experiments and classic observational studies should be combined to take advantage of both methods' respective strengths.
- The logic and terminology of experiments will likely keep influencing other research.

Wrapping up

The rest of this week

- A closer look at the core assumptions after lunch.
- Experiments are (relatively) easy to analyze. No new Stata commands required for the seminar assignment.
- But unlike last week, you will still need Stata for part of the assignment.
- The workshop introduction will focus on loops and simulations.

Wrapping up

Don't miss Valborg and May Day!

Idag 27 april 8° 0° min  0,1 mm ↓ 6 m/s  ↑ 05:00 ↓ 20:35	Tisdag 28 april 8° 0° min  0 mm ↓ 6 m/s  ↑ 04:57 ↓ 20:37	Onsdag 29 april 13° -2° min  0 mm ↗ 3 m/s  ↑ 04:55 ↓ 20:40	Torsdag 30 april 20° 4° min  0 mm ↓ 4 m/s  ↑ 04:52 ↓ 20:42	Fredag 1 maj 22° 7° min  0 mm ↓ 3 m/s  ↑ 04:49 ↓ 20:45
--	--	--	--	--