VU Empirical Economics: Impact Evaluation

Introduction to policy evaluation

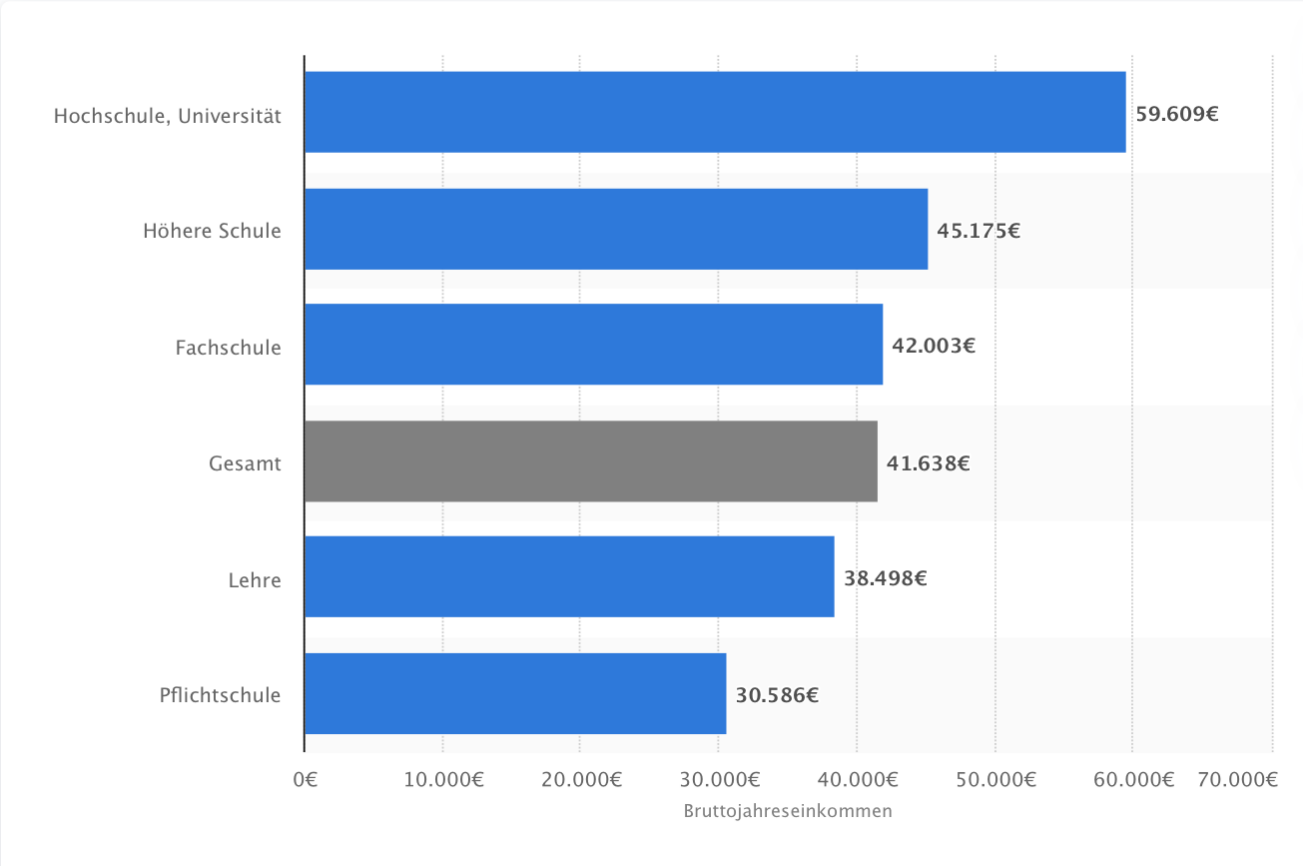
Andreas Steinmayr

University of Innsbruck

Fall 2023

Core reading: Gertler, chapters 1 and 2

Mean annual earnings of fulltime workers in Austria 2017



Source: Statistik Austria, Statista 2022

Purpose of this course

- Understand program evaluation as a *consumer*
  - Synthesize and understand evaluation results
  - Evaluate the quality of policy evaluations or choose between evaluation proposals
  - Make evidence-based decisions

Purpose of this course

- Understand program evaluation as a *consumer*
  - Synthesize and understand evaluation results
  - Evaluate the quality of policy evaluations or choose between evaluation proposals
  - Make evidence-based decisions
- Provide skills to design and implement program evaluations
  - Many economists work on evaluations, not only in academia

Why do we evaluate programs?

- *Policy* and *program* will be used as synonyms

- Lots of money is spent every year to try to change things
  - government-provided job training programs
  - sugar-tax to reduce obesity
  - subsidies to increase R&D investment in firms
  - incentive schemes to increase worker productivity

- We want to know:
  - What difference have these programs made?
  - What outcomes have been affected? By how much and for whom?

Broader agenda: Evidence-based policy making

- Policy decisions informed by rigorously established evidence
  - In reality often: *Policy-based evidence making*
- Based on idea of evidence-based medicine
- Goal: Inform allocation of resources, guide policy decisions, enhance accountability
- Build general knowledge about the effectiveness of policies

Which policies should we evaluate

- Evaluations are costly, especially data collection
- What are the stakes of this policy?
  - Budget
  - Size of target population
  - *Potential* effect sizes
- Evaluate if the policy is
  - Innovative: new and promising
  - Replicable: can be scaled up or be applied in a different setting
  - Strategically relevant: flagship initiative, requires substantial resources, potentially covers a large number of people, could have large (side) effects or generate substantial savings
  - Untested: Little is known about the effectiveness of this type of policy
  - Influential: Results will be used to inform key policy decisions

Common errors

- People without knowledge in program evaluation tend to confuse
  - monitoring and evaluation
  - correlation and impacts
- Examples
  - The program was successful: 72% of participants find a job after job training
  - I feel better today, because I took Globuli yesterday
  - She has a high income, because she studied economics
  - Aztecs: without human sacrifices of children, rain would not come and crops would not flourish
- This can be extremely misleading

Simpson's (1951) paradox

- Event C increases probability of E in population, whereas it decreases probability of E in all sub-populations
- Example: Taking a particular pill is helpful for the population but harmful for men and women

Simpson's (1951) paradox

- Event C increases probability of E in population, whereas it decreases probability of E in all sub-populations

- Example: Taking a particular pill is helpful for the population but harmful for men and women

Combined	recovery (E)	not E	sum	recovery rate
Drug (C)	20	20	40	50%
No drug (not C)	16	24	40	40%

Simpson's (1951) paradox

Men	recovery (E)	not E	sum	recovery rate
Drug (C)	18	12	30	60%
No drug (not C)	7	3	10	70%

Women	recovery (E)	not E	sum	recovery rate
Drug (C)	2	8	10	20%
No drug (not C)	9	21	30	30%

What is going on?

Simpson's (1951) paradox

- Drug appears to be beneficial in for the population because
  - men recover more often than women with and without taking the drug &
  - men are more likely than women to take the drug
- In other words, sex is a confounder and needs to be controlled for the obtain valid causal conclusions
- Other example: UC-Berkeley investigation for sex-bias in graduate admissions (1975)

What is impact evaluation?

- **Monitoring** tracks what is happening with a program, looks at the program implementation
  - Is the money indeed spent the way it was supposed to be?
- **Impact evaluations** seek to answer a cause-and-effect question
  - Which changes are directly attributably to a program?
  - What is the effect of obtaining a university degree on earnings?

What is impact evaluation?

The central idea of any impact evaluation is to estimate the so-called **counterfactual**.

What the outcome would have been for program participants if they had not participated in the program.

Why is this difficult? For any given person/unit we observe the outcome only in the treated or untreated state, never in both!

Prospective versus retrospective evaluation

- **Prospective evaluation**
  - Set up at the same time as policy
  - Built into policy implementation / roll-out
- **Retrospective evaluation**
  - Policy evaluation after implementation
- PE more likely to produce credible evaluation results
  - (Baseline) data collection on treated and controls
  - Creates focus on policy objectives
  - Easier to construct credible counterfactual

Efficacy studies and effectiveness studies

- **Efficacy studies**
  - Carried out in a specific setting under closely controlled conditions
  - Small-scale pilot / proof of concept
- **Effectiveness studies**
  - Interventions that take place in normal circumstances, using regular implementation channels
  - Aim to produce findings that can be generalized to a large population

Internal versus external validity

- **Internal validity**
  - Evaluation identifies causal effect of program in a given setting
  - Varying degrees of credibility (RCT: Gold Standard)
- **External validity**
  - Generalizability of causal effect to other situation
  - Informative for larger or different population, different time

Internal versus external validity

- **Internal validity**
  - Evaluation identifies causal effect of program in a given setting
  - Varying degrees of credibility (RCT: Gold Standard)
- **External validity**
  - Generalizability of causal effect to other situation
  - Informative for larger or different population, different time
- Economists argue about their relative importance (Deaton 2010)
  - High degree of internal validity often easier in settings with low external validity (e.g. lab experiments with college students)

Cost-benefit and cost-effectiveness analysis

- **Cost-benefit analysis**
  - Estimates the total expected benefits of a program, compared to its total expected costs.
  - Seeks to quantify all of the costs and benefits of a program in monetary terms and assesses whether benefits outweigh costs.

- **Cost-effectiveness analysis**
  - Compares the relative cost of two or more programs or program alternatives in reaching a common outcome

Impact evaluation estimates the benefit side, and cost analysis provides the cost information! We focus on **impact evaluation**.

Preparing for an evaluation

- Specify the evaluation question
- Construct a theory of change
- Develop a results chain
- Select indicators to assess performance

Evaluation question

- First step: Formulate a clear study question
  - What is the impact of the policy on an outcome of interest?
  - Which changes are directly attributable to a program?

- Needs to be framed as a **well-defined, testable hypothesis**

Are these good evaluation questions?

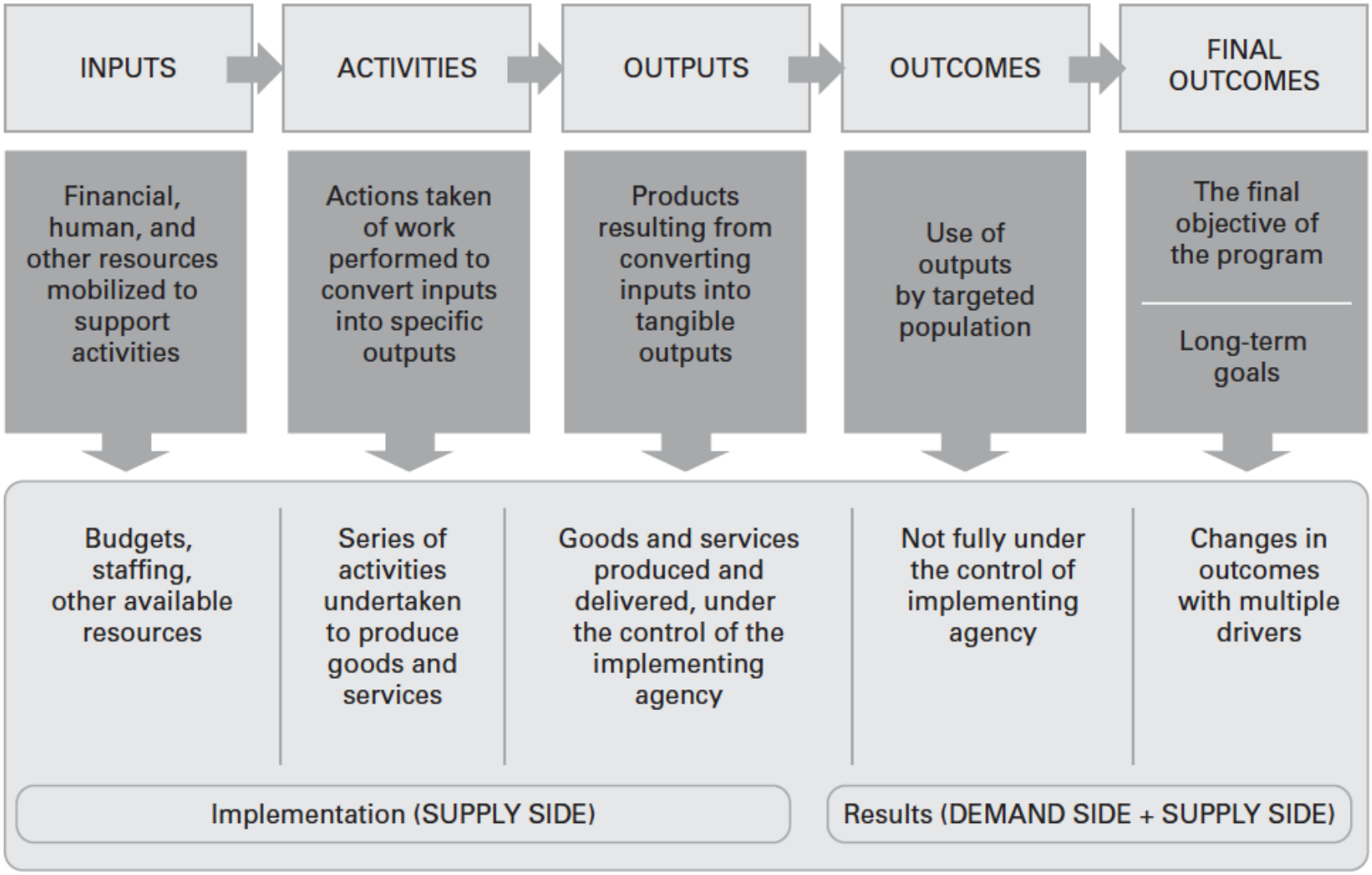
- What is the effect of studying economics on later earnings?
- What is the effect of being a woman on the likelihood of becoming a politician?
- What is the effect of reducing the speed limit on highways from 130 to 100 km/h on the number of traffic deaths?

Theory of change

- Describe the causal pathway (sequence of events from policy to final outcomes)
- Formulate necessary assumptions and enabling conditions
- Include all stakeholders
- Consult existing literature

Can be based on formal model or simply as a results chain.

Results chain



Results chain - example

- Government agency provides trainings about settlement to immigrants
  - **Inputs**: Staff from government agency, trainers, facilities, government budget, contact information of immigrants
  - **Activities**: Design curriculum of training, select and prepare trainers, design, prepare, and distribute written material
  - **Outputs**: Number of immigrants trained, number of written materials provided
  - **Outcomes**: Immigrants behave differently (smarter) based on information provided
  - **Final outcomes**: Immigrants are (economically) more successful, higher well-being, less welfare spending for immigrants, reduced social tensions

Performance indicators (Outcomes)

- Ideally use indicators along the whole results chain
  - Not only final outcomes
  - Understand mechanisms
  - Especially important if evaluation finds no effect on final outcome
- Selection of indicators should involve stakeholders
- SMART indicators
  - Specific: translates required information into operational measure
  - Measurable: information can be measured and obtained
  - Attributable: indicator is linked to the policy's efforts
  - Realistic: information can be obtained timely, with reasonable frequency, at reasonable cost
  - Targeted: to the objective population