Essays in Empirical Economics using Microeconometric and Causal Machine Learning Methods

Michael Knaus

PEF Public Defense, 06.03.2018, St. Gallen

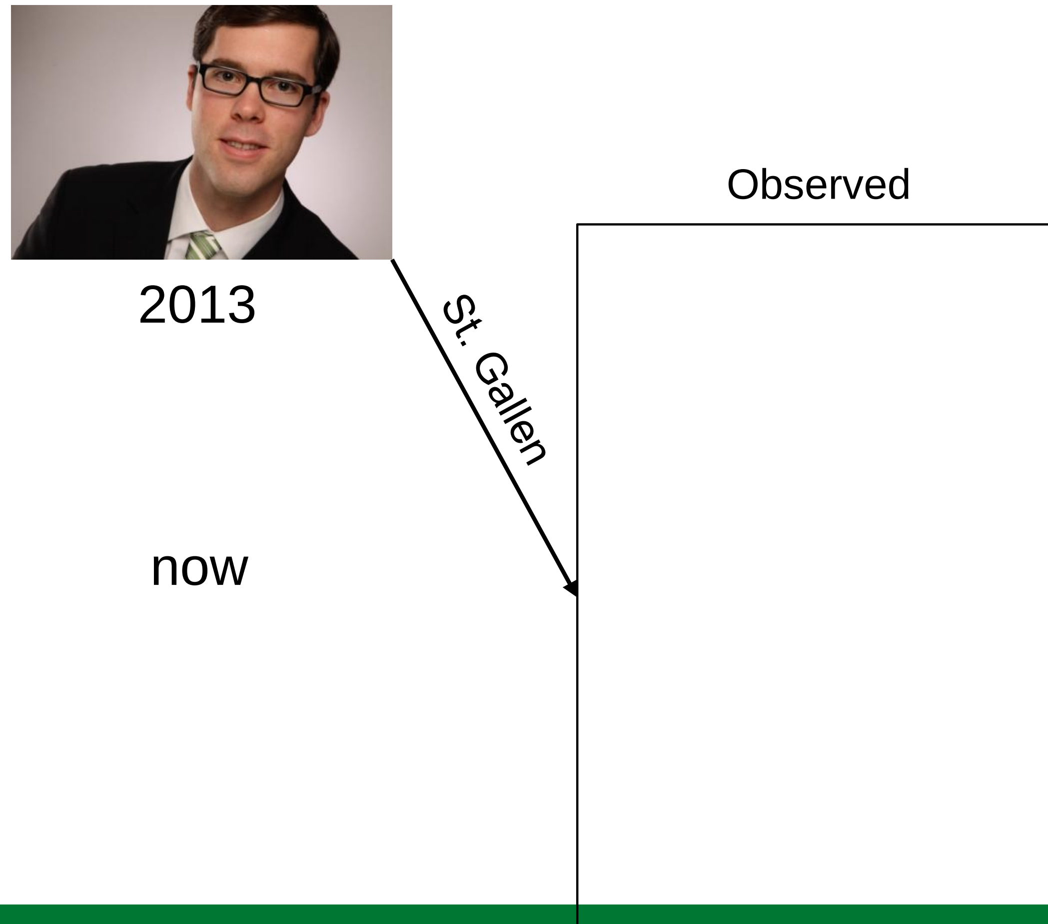
The «What if …?» questions

- What if this guy did something else?

2013

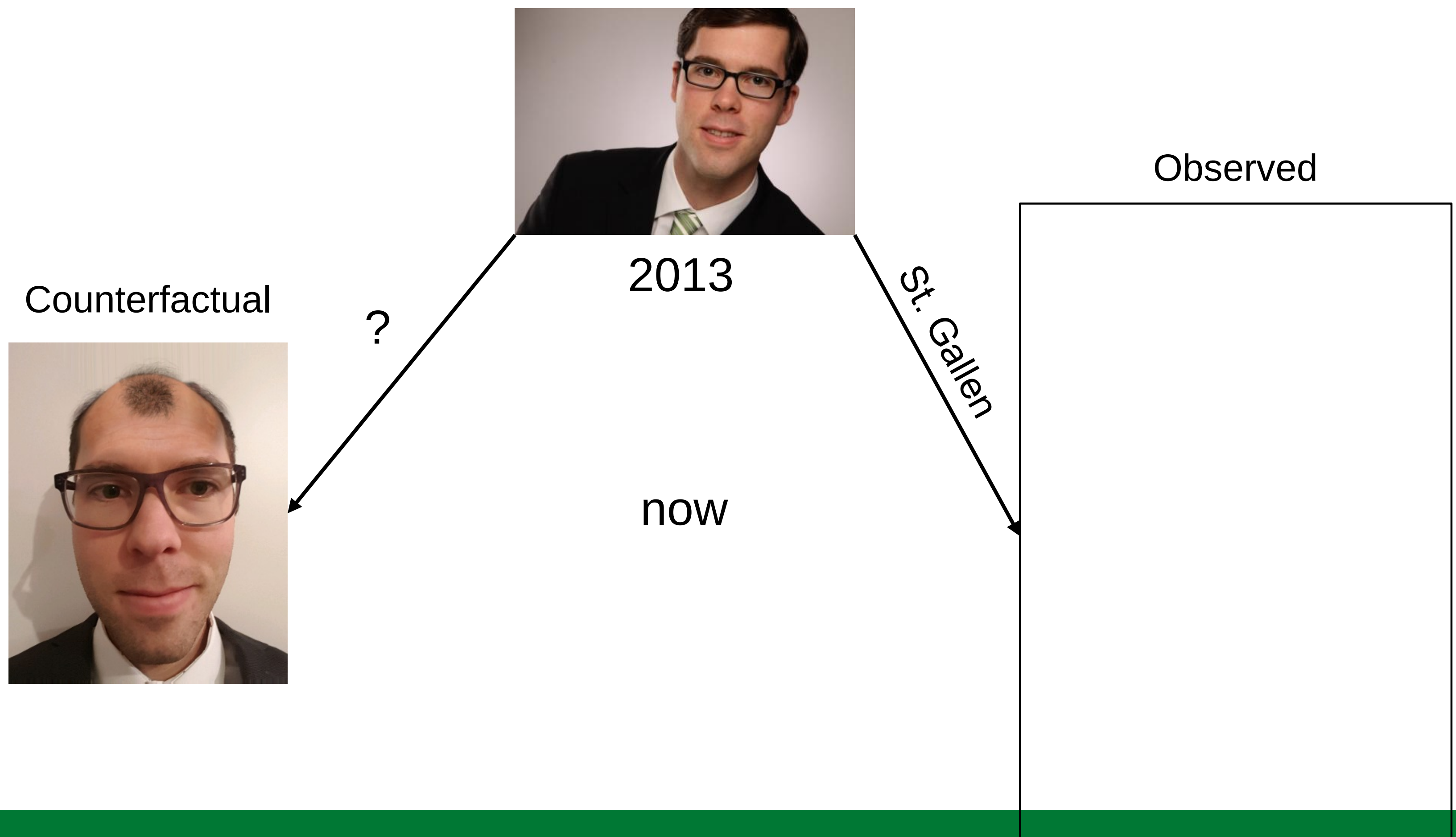
The «What if ...?» questions

- What if this guy did something else?



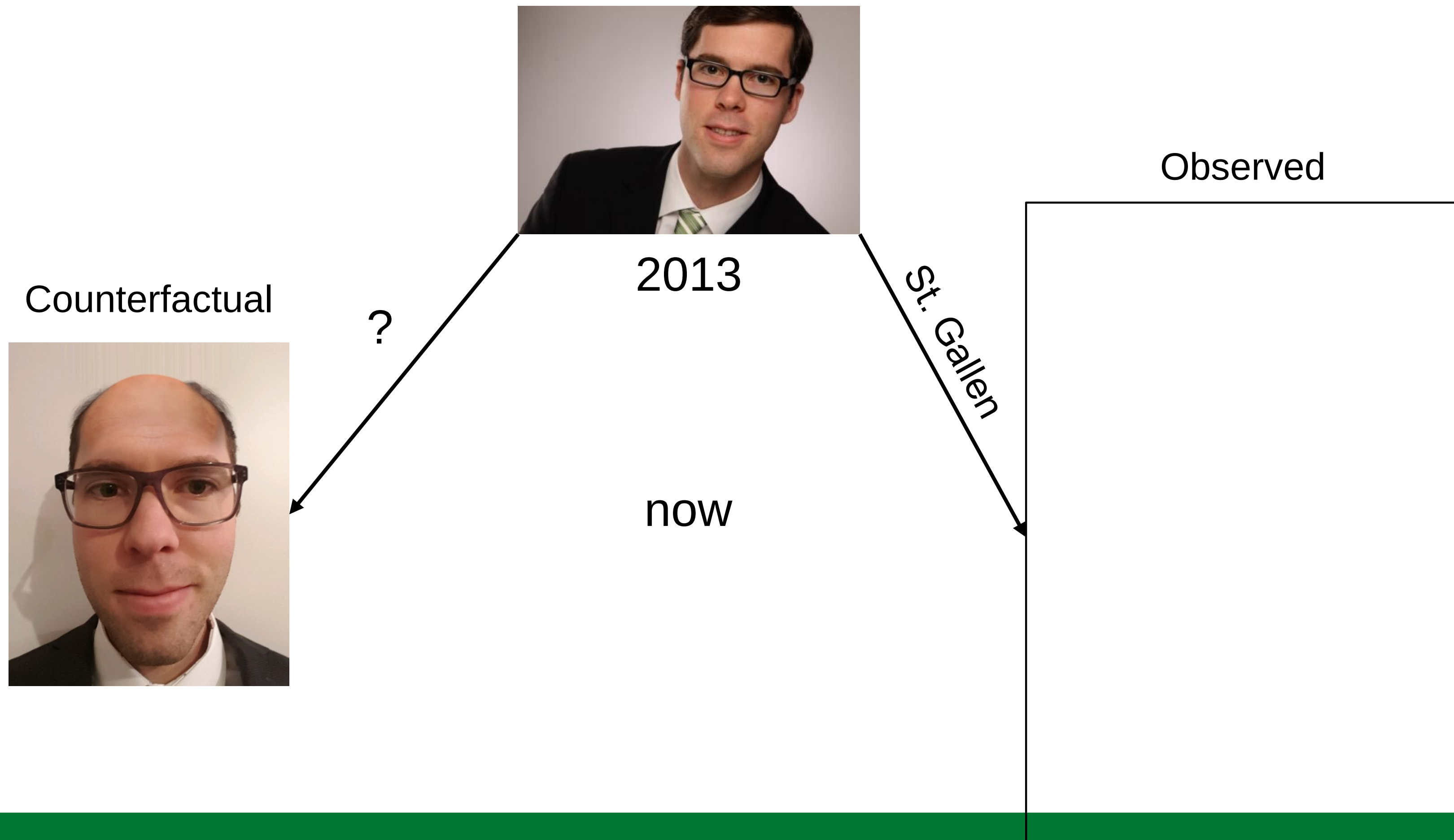
The «What if …?» questions

- What if this guy did something else?



The «What if …?» questions

- What if this guy did something else?



The «What if …?» questions

- It is impossible to observe the counterfactual on the individual level

$\Rightarrow$ The *causal* effect of different choices is not observed

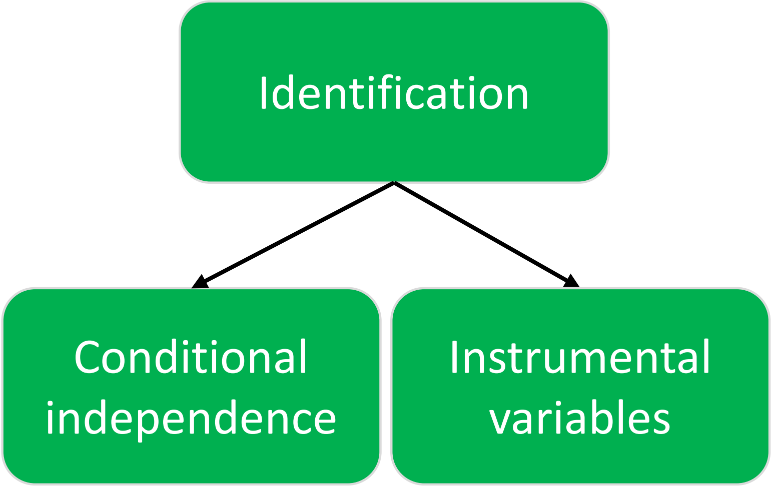
- *Goal* of the analyses in this *dissertation* is to estimate at least *average causal effects* of
  - Playing music (Ch. 1)
  - Physical education (Ch. 2)
  - Active labor market policy (Ch. 3)
  - Job changes (Ch. 4)

Ingredients of causal analyses

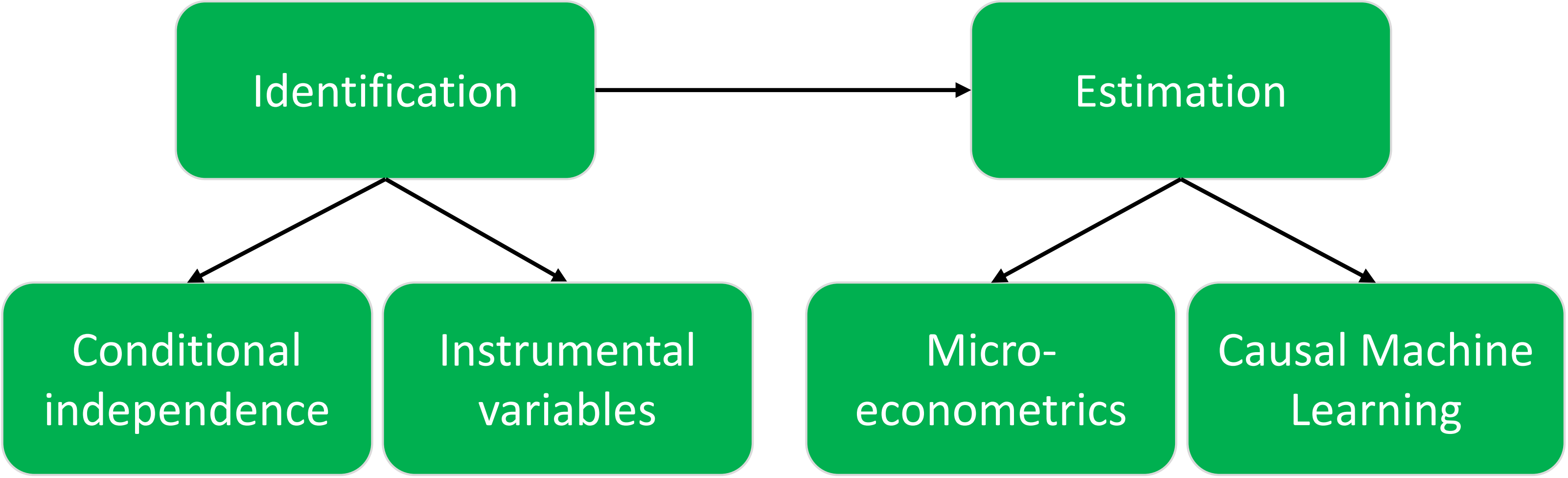
- Identification: Arguments that the *research design* allows to identify the *causal* effect of interest
  - Requires institutional and contextual knowledge of the problem
  - Involves untestable assumptions

- Estimation: *Statistical techniques* to estimate the causal effect of interest
  - Large and steadily increasing toolbox of methods
  - Might require additional assumptions

Ingredients of the analyses in the thesis



Ingredients of the analyses in the thesis

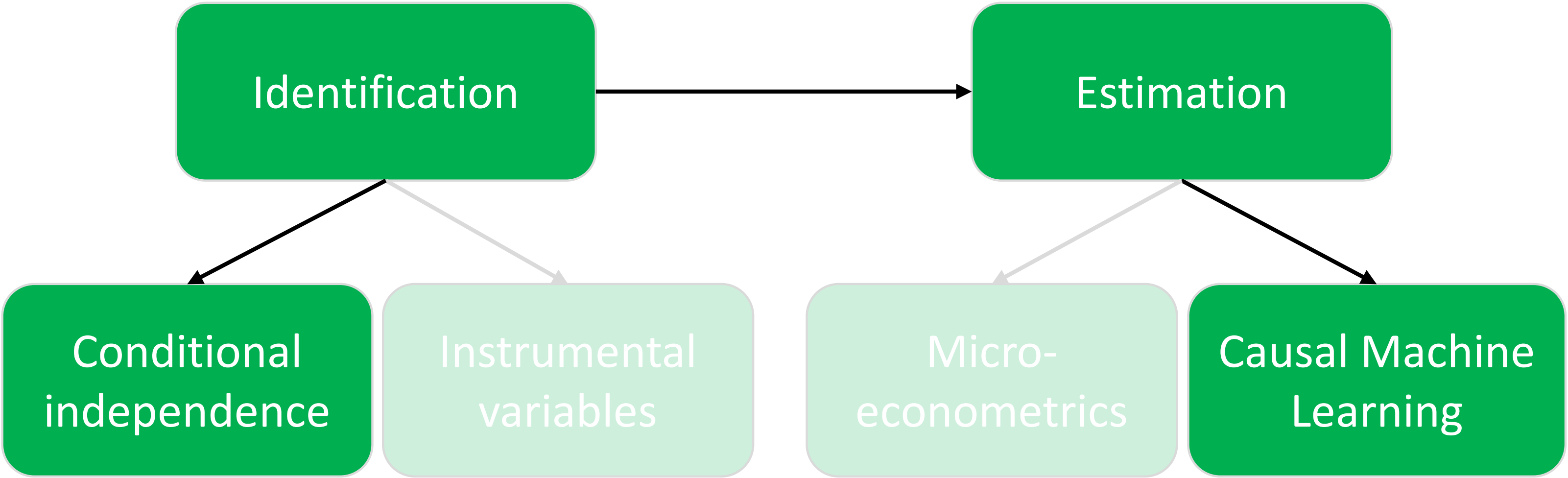


Microeconometrics vs. Causal Machine Learning

	Microeconometrics	Causal Machine Learning
Identification	Same	
Model selection	Manual (at least to some degree)	Data-driven
Statistical inference	God gave us the final model	Accounts for the model selection step

- Causal machine learning *splits* the *estimation* of causal effects into several *prediction problems*
- The predictions problems are then solved by standard or adapted machine learning tools

Chapter 1: Intensity of Musical Practice and Youth Development



Chapter 1: Intensity of Musical Practice and Youth Development

- Motivation: Understand the impact of *extracurricular activities* on cognitive and non-cognitive *skills*

- Contribution:
  - *Dose-response* relation between musical practice and skill development
  - Observed parental tastes increase *credibility* of *identification*

- Data: *German National Educational Panel Study* with ~7,000 9th graders

Chapter 1: Intensity of Musical Practice and Youth Development

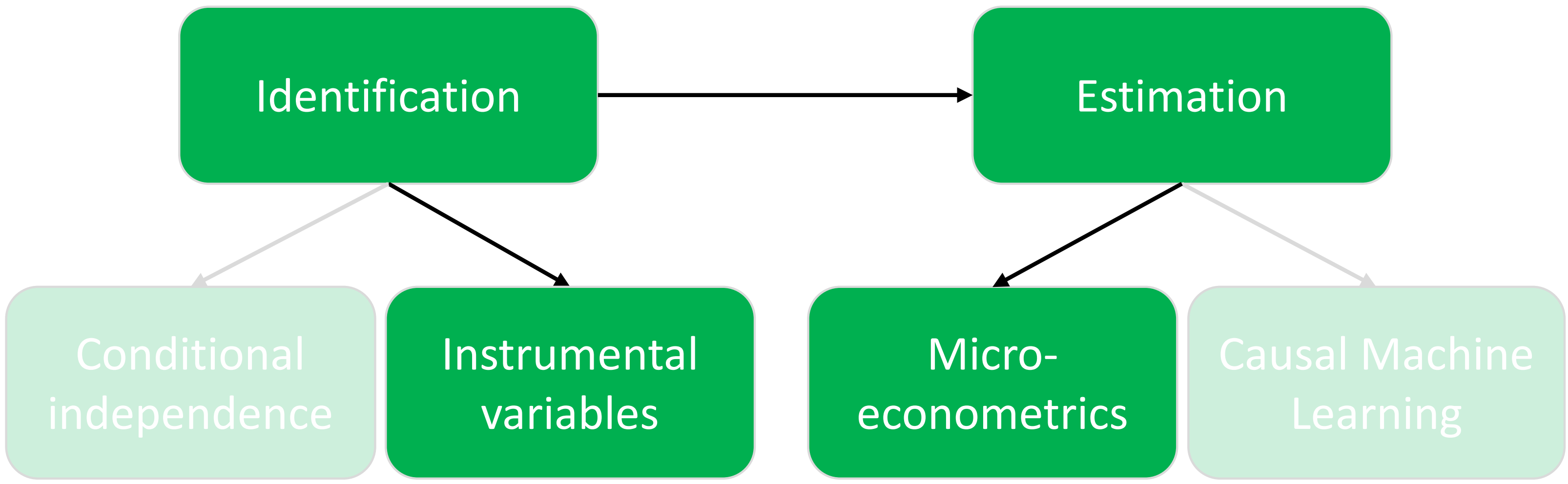
- Identification: *Conditional independence assumption*
  - Detailed parental background
  - Individual characteristics

- Estimation: *Double Machine Learning* (Farrell, 2015)

- Results:
  - *Improved cognitive skills* for medium and high intensity
  - *Improved grades* already for low intensity, mainly for German
  - Agreeableness and openness in *Big Five signficantly improved*

Chapter 2: For better or worse? - The Effects of Physical Education on Child Development

with Michael Lechner and Anne Reimers



Chapter 2: For better or worse? - The Effects of Physical Education on Child Development

- Motivation:
  - *Discussions* about *increasing physical education* (PE)
  - *Scarce evidence* regarding effect of *regular PE* on child development

- Contribution:
  - Comprehensive analysis of *all five domains* of intended PE effects:
    - Cognitive / non-cognitive / motor skills
    - Extracurricular physical activity
    - Health and fitness

- Data: *German Motorik-Modul* with ~5,500 observations

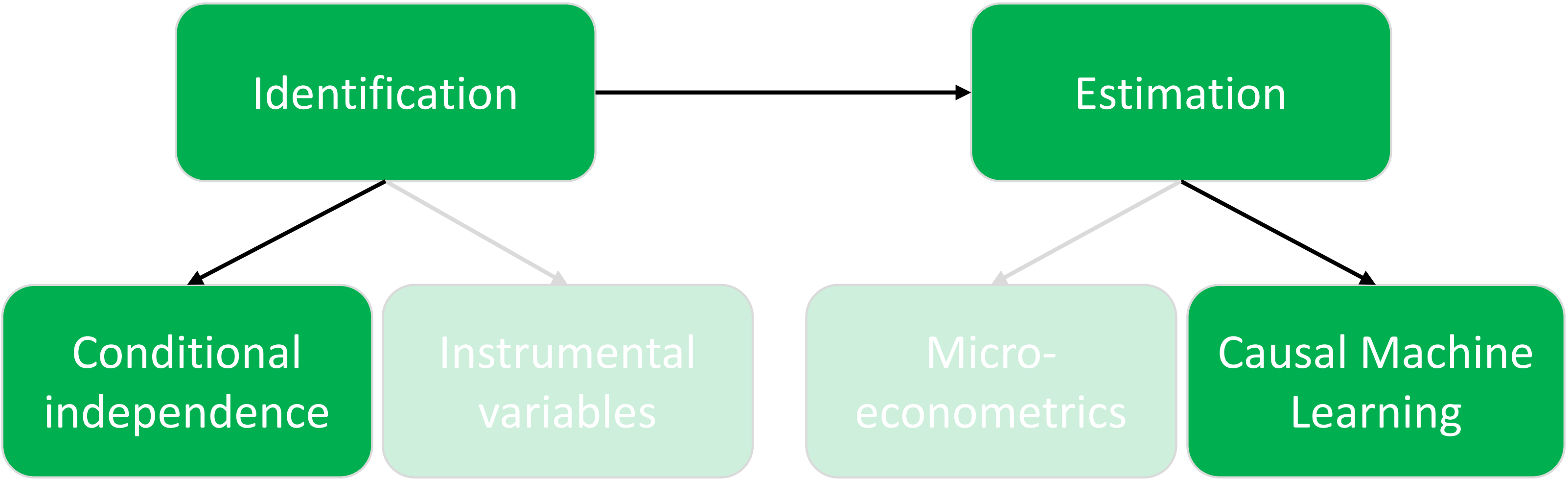
Chapter 2: For better or worse? - The Effects of Physical Education on Child Development

- Identification: *Instrumental variable*
  - Exploit variation in PE requirements across federal states
- Estimation: Semi-parametric IV using Inverse Probability Tilting
- Results:

Outcome group	**All**	**Boys**	**Girls**
Cognitive skills	+	+	+
Non-cognitive skills	-	-	+
Motor skills	+	∘	+
Physical activity	∘	∘	+
Health	∘	∘	∘

Chapter 3: Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach

with Michael Lechner and Tony Strittmatter



Chapter 3: Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach

- Motivation: Estimate *conditional average treatment effects* (CATEs) of job search programmes

- Research questions:
  - Do "*causal machine learning*" methods provide *useful* tools to uncover *heterogeneous causal effects* in policy evaluations?
  - Did *Swiss* job search programmes have *differential effects* for different groups of unemployed?

- Data: Administrative data of Swiss unemployed and caseworkers

Chapter 3: Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach

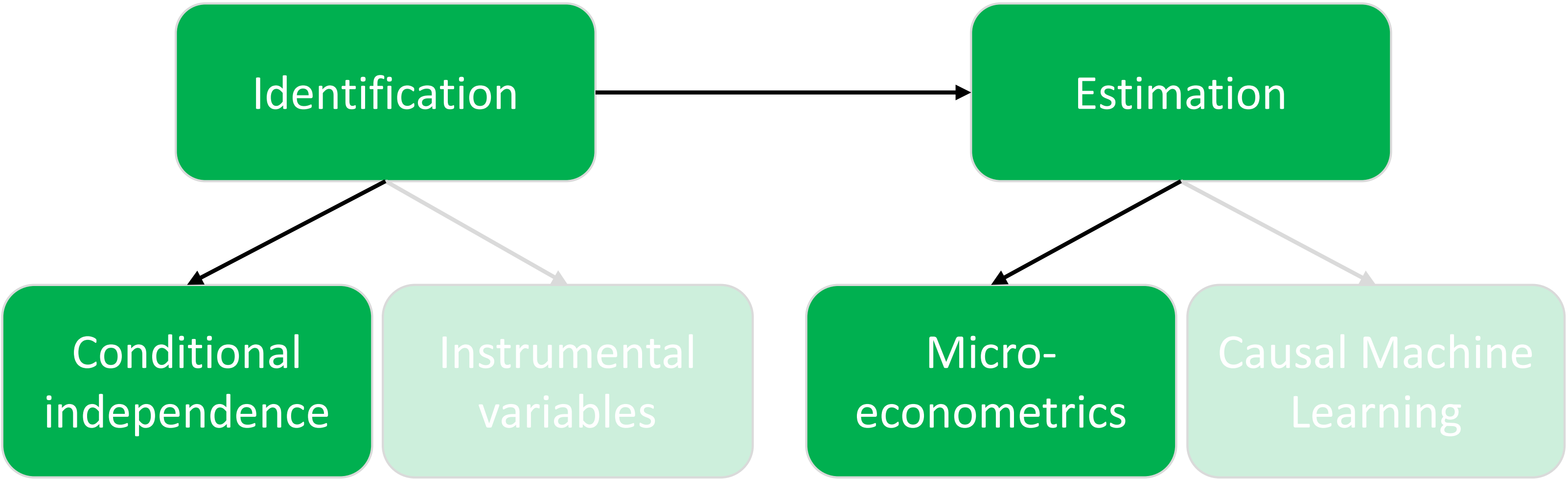
- Identification: *Conditional independence assumption*
    - Detailed unemployed and caseworker characteristics

- Estimation: *Modified Covariate Method* (Tian et al., 2014) and several alternatives

- Results:
    - On average *substantial lock-in effects* of the programme
    - Substantial *heterogeneity detected* behind standard average effects
    - *Assignment mechanisms fail* to identify individuals with *gains*

Chapter 4: Work Hour Mismatch and Job Mobility: Adjustment Channels and Resolution Rates

with Steffen Otterbach



Chapter 4: Work Hour Mismatch and Job Mobility: Adjustment Channels and Resolution Rates

- Motivation:
  - Existing literature interprets *increased work hour flexibility* of *job movers* as *evidence* for *free hour choice* across jobs
  - Suggests that job movers can resolve work hour mismatches

- Contribution:
  - Comprehensive analysis including actual resolution of work hour mismatch in the analysis

- Data: *German Socio-Economic Panel* providing rare measure of desired work hours in numbers

Chapter 4: Work Hour Mismatch and Job Mobility: Adjustment Channels and Resolution Rates

- Identification: *Conditional independence assumption*
  - Detailed socio-demographic, job-related, and regional information

- Estimation: *Propensity score matching* with bias adjustment

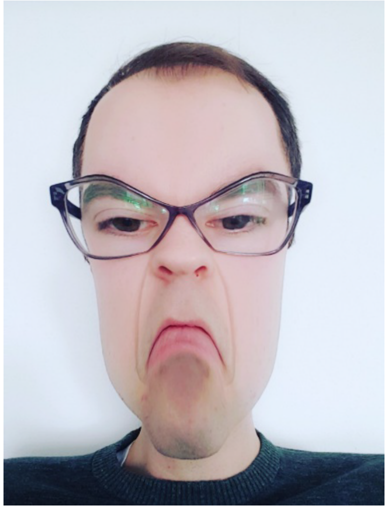
- Results:
  - *Confirm previous finding* of larger adjustments in actual work hours
  - Job mobility *only moderately increases* the probability to *resolve* work hour *mismatches*

Concluding remarks

- *Machine learning* adds *new interesting options* to the causal analysis toolbox
- However, it does *not solve* the fundamental *identification* problem that counterfactuals are unobserved
- Still, ML can be used to create hypothetical counterfactuals

Concluding remarks

- *Machine learning* adds *new interesting options* to the causal analysis toolbox
- However, it does *not solve* the fundamental *identification* problem that counterfactuals are unobserved
- Still, ML can be used to create hypothetical counterfactuals



Thank you for your attention!