THE VALUE ADDED OF CAUSAL MACHINE LEARNING FOR PUBLIC POLICY

OECD, Paris, October 2023

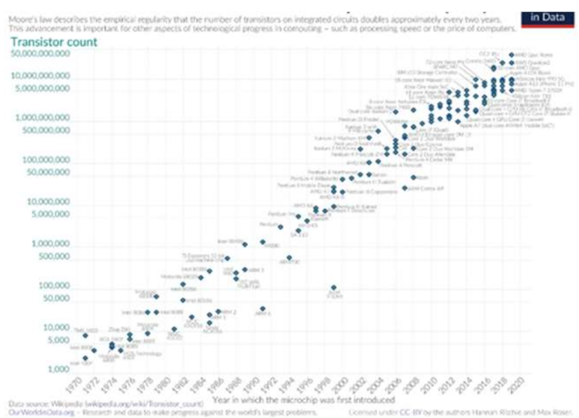
Michael Lechner
Swiss Institute for Empirical Economic Research (SEW)
University of St. Gallen | Switzerland

University of St.Gallen
Swiss Institute for Empirical Economic Research

Econometrics developed methods to credibly estimate effects of active labour market policies

Better & cheaper computers

- Moore's law: # of transistors on micro chips 2x every 2 years



More & better data

- Cheaper to collect & store
- Individuals more readily accept that their data is used by others
- Easier to merge different administrative data sets (even outside Scandinavia)

Better algorithms

- Machine Learning (ML)

Data Science & Machine Learning **became basis of many successful business models**

Combining these developments

Causal **M**achine **L**earning

- ML inspired estimators applied to *well-identified* causal questions with rich data

Considerable interest in CML

- Method developments in statistics, computer science & some applied fields
  - Many researchers work on similar questions ➔ fast progress
- Field specific versions of CML are spreading (needs of fields differ)
  - Econometrics (active labour market programme evaluation, …)
  - Epidemiology (personalized medicine, …)
  - Marketing (targeted marketing & political campaigns, …)
  - …

University of St.Gallen
Swiss Institute for
Empirical Economic Research

The promise of Causal Machine Learning for public policy

More robust & precise estimation of average population effects

➔ Better understanding of effects of policies at large

Better estimation of heterogeneity of effects

➔ Better understanding of the implications of policies for specific groups

➔ Better targeting of policies to specific groups

Better decision making

➔ Improving decisions by *algorithmic* / *algorithm-assisted* decision rules

➔ Better targeting of policies to particular firms, individuals, etc.

1 | Introduction

2 | Methodology for a special case

3 | An example: Active labour market policies in Flanders

4 | Conclusions & road ahead

Overview of CML methodology for a special case

Policies with individual variation

- Some units are affected
- Some units are not affected

Policy variable (=*treatment*) is binary

- For simplification of notation only

Research design: Selection-on-observables / unconfoundedness / conditional independence

- Includes experiments

University of St.Gallen
Swiss Institute for
Empirical Economic Research

Effects & aggregation levels

D: Treatment (0 or 1)

Y^1: Outcome when $D=1$

Y^0: Outcome when $D=0$

X: Confounder & heterogeneity variables

Z: Specific heterogeneity variables (low dim.)

Observable: $X,\ H,\ Y = DY^1 + (1-D)Y^0$

Individualized (Conditional) Average Treatment Effects

$$IATE(x) = E(Y^1 - Y^0 \mid X = x) = E(Y \mid X = x, D = 1) - E(Y \mid X = x, D = 0)$$

Group (Conditional) Average Treatment Effects

$$GATE(h) = E(Y^1 - Y^0 \mid H = h) = E_{X|H=h} IATE(x)$$

Balanced GATE (BGATE)

$$BGATE(h;\tilde{x}) = E(Y^1 - Y^0 \mid H = h, \tilde{X} = \tilde{x}) = \ldots$$

Average Treatment Effects

$$ATE = E(Y^1 - Y^0) = E_X IATE(x)$$

Average Treatment Effects on the Treated

$$ATE = E(Y^1 - Y^0 \mid D = 1) = E_{X|D=1} IATE(x)$$

Effect estimation

Comprehensive (all-in-one) estimation methods

- One approach for the estimation of all parameters of interest (all-in-one)
- Option I: Use ML inside specific moment conditions (double/debiased machine learning, DML)
- Option II: Change a ML into a CML (causal forests, etc.)

Parameter specific estimation methods

- Use different methods for different parameters
  - Since methods are tuned to specific target of interest, they may perform better
- Many estimators, lot’s of different tuning parameters
- Substantial effort to understand specifics of estimators & monitor problems in all estimations
- Possible lack of internal consistency (GATEs may not add up to ATE, etc.)

Decision making (*optimal policy*)

Use disaggregated effects for decision making

- Find *X*-based rule for optimal allocation
  - Welfare function of decision maker
  - Constraints

Literature currently booming

Many open questions that are or will be addressed

- Methodological (statistical properties)
- Computational (in particular for multiple & cont. treatments)
- Ethical (‘fairness’)
- Practical (Black-box vs. explainable rules)


University of St.Gallen
Swiss Institute for
Empirical Economic Research
1 | Introduction
2 | Methodology for a special case
3 | An example: Active labour market policies in Flanders
4 | Conclusions & road ahead

University of St.Gallen
Swiss Institute for Empirical Economic Research

Empirical example

Labour Economics 80 (2023) 102306

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

ELSEVIER

Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium

Bart Cockx[$,#], Michael Lechner[$$,##], Joost Bollens[$$$,*]

[$] Department of Economics, Ghent University
[$$] Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen
[$$$] Vlaamse Dienst voor Arbeidsbemiddeling en Beroepsopleiding (VDAB)

Goal: Evaluation of participation in training programmes for unemployed

- Programmes are part of the active labour market policy of Flanders (Belgium)

3 types of training programmes considered

- Short & long vocational training, orientation training

Administrative data from Flemish employment service (about 60’000 observations)

University of St.Gallen
Swiss Institute for
Empirical Economic Research

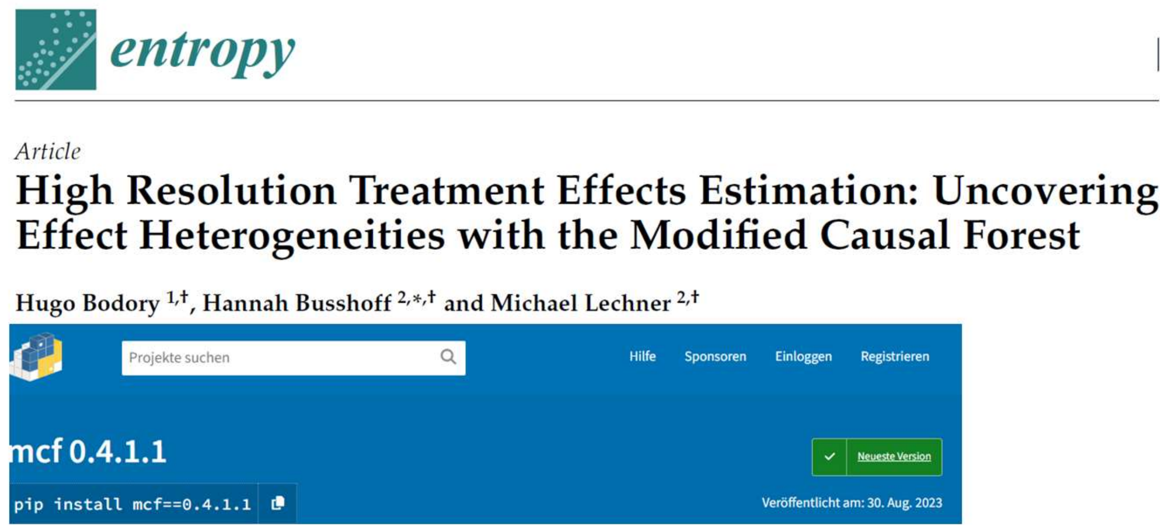
Data & Estimation

Empirical questions

- Did the programmes work on average?
- For whom did they (not) work?
- Could the allocation of unemployed to these programmes be improved?

Estimation

- > 200’000 parameters
- Modified Causal Forest
  - Free Python code available on PyPI


entropy

Article

High Resolution Treatment Effects Estimation: Uncovering Effect Heterogeneities with the Modified Causal Forest

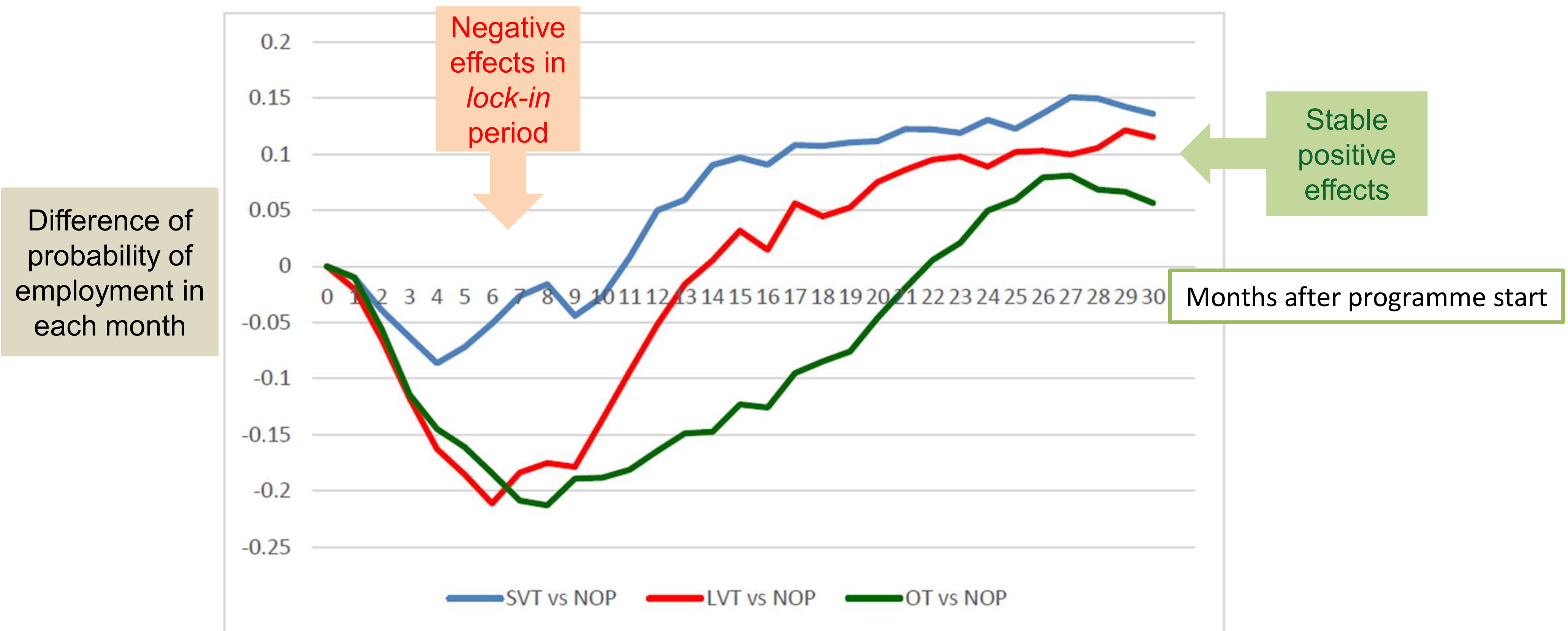
Hugo Bodory [1,†], Hannah Busshoff [2,*,†] and Michael Lechner [2,†]

University of St.Gallen
Swiss Institute for Empirical Economic Research

SVT: Short vocational training
LVT: Long vocational training
OT: Orientation training
NOP: Nonparticipation

Average effects

Time evolution of the ATEs

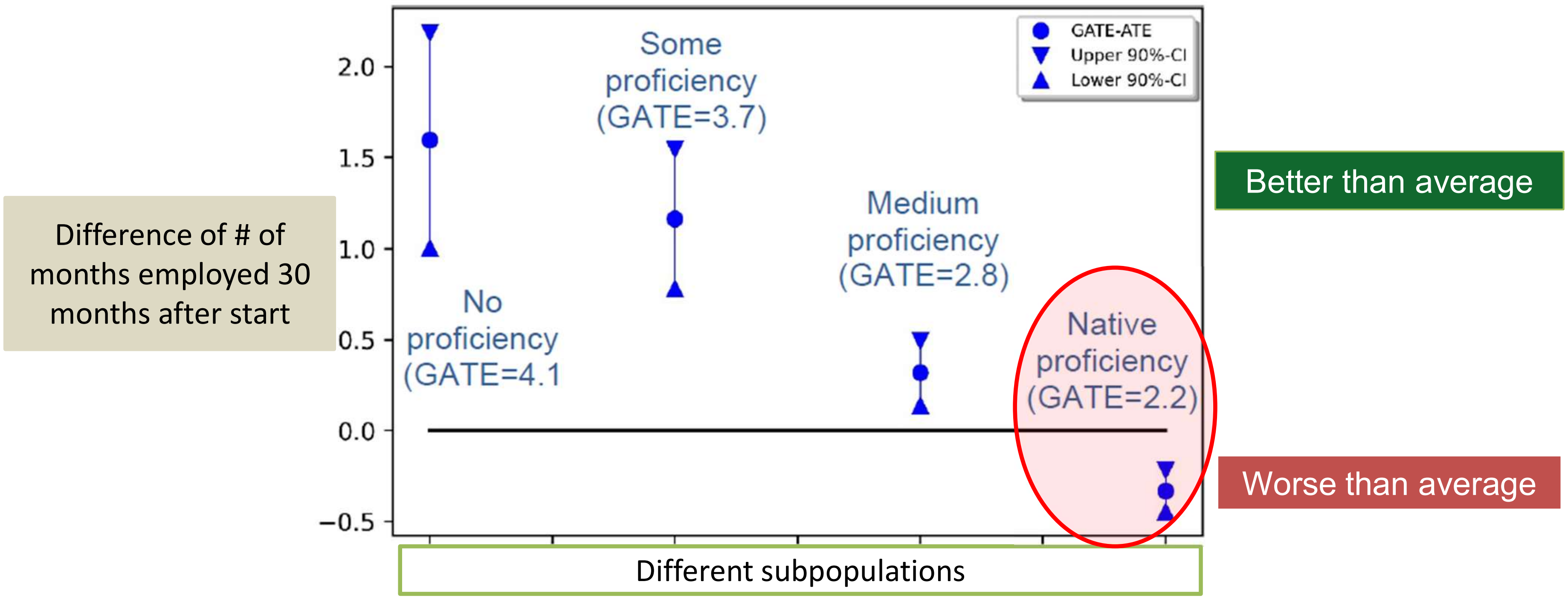


University of St.Gallen
Swiss Institute for Empirical Economic Research

SVT: Short vocational training
LVT: Long vocational training
OT: Orientation training
NOP: Nonparticipation

Group ATEs minus ATE | 1

SVT vs. NOP for **proficiency levels in Dutch**



Dutch proficiency displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. Dutch proficiency varies between no proficiency (0) and native proficiency (3).

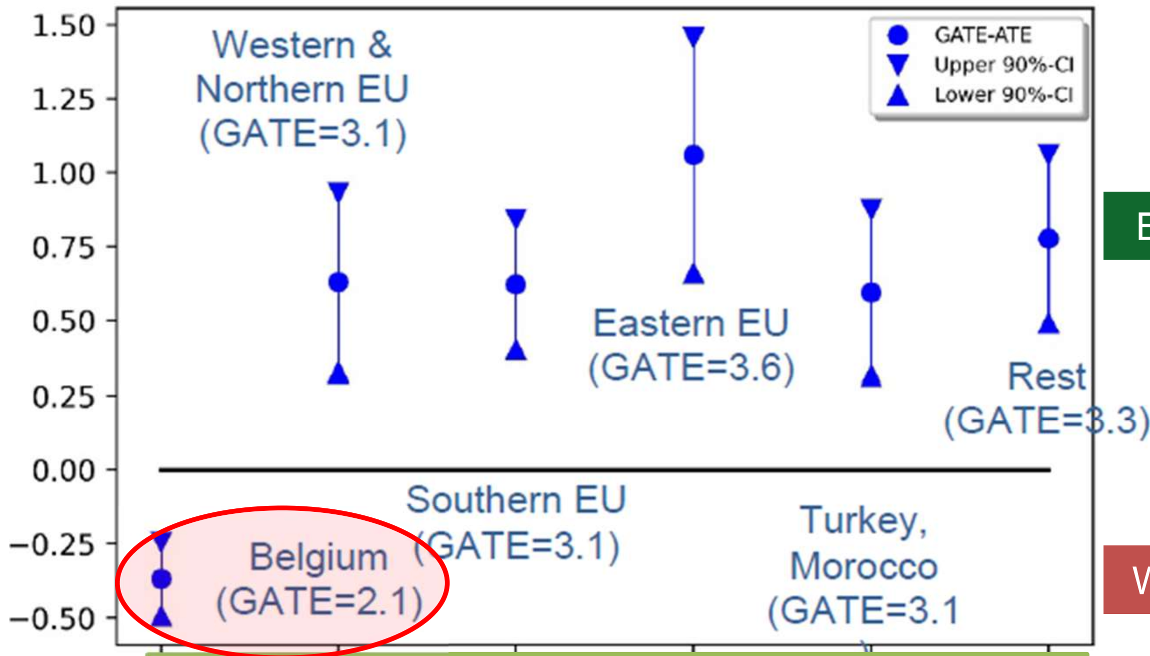
University of St.Gallen
Swiss Institute for Empirical Economic Research

SVT: Short vocational training
LVT: Long vocational training
OT: Orientation training
NOP: Nonparticipation

Group ATEs minus ATE | 2

SVT vs. NOP according to **country of birth**

Difference of # of months employed 30 months after start



Better than average

Worse than average

Different subpopulations

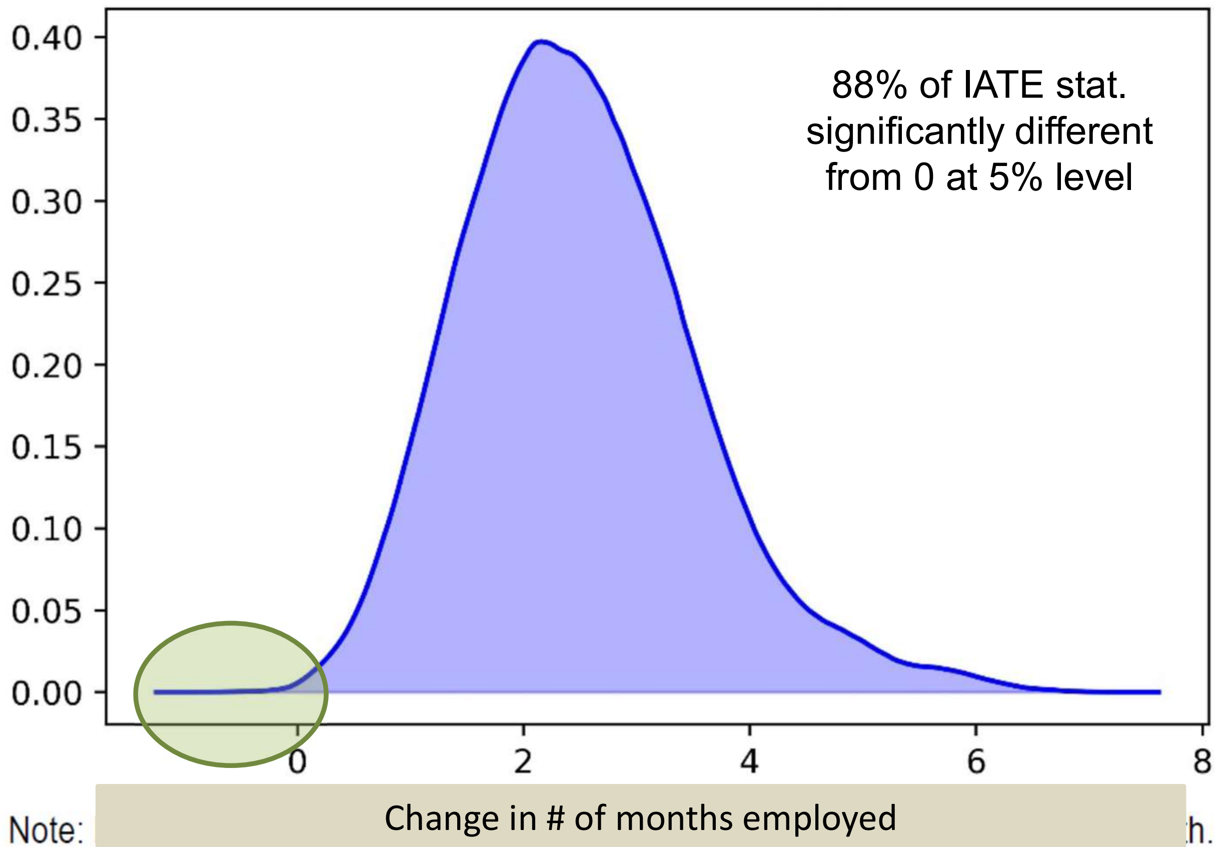
Country of birth displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. The vertical axis measures the deviation of the GATE from the ATE.

University of St.Gallen
Swiss Institute for Empirical Economic Research

SVT: Short vocational training
LVT: Long vocational training
OT: Orientation training
NOP: Nonparticipation

Individualized ATEs | 1

Distribution of estimated IATE of SVT vs. NOP



Individualized ATEs | 2

Characterisation of unemployed with high/low IATEs

- Form homogenous groups w.r.t. IATEs
- Compare means of covariates across groups
- Unsupervised ML: Here, k-means clustering

Findings

- Largest effects for *born outside Belgium*, *no good command of Dutch*, *older*, *low employability*
- Lowest effects for *born in Belgium*, *high employability*
- No gender differences

University of St.Gallen
Swiss Institute for
Empirical Economic Research

Allocation of individuals to programmes | 1

Target variable

- Expected increase in months of employment & reduction in months of unemployment
  - Both criteria are equally weighted

Results

- Observed (case workers): Allocation not correlated with estimated effects
- Black-Box (observed programme shares as capacity constraint)
  - + 1 month additional employment & 1 month reduced unemployment (for those reallocated)
- Shallow decision tree (observed programme shares as capacity constraint)
  - Gains only slightly smaller, but allocation rule is easy to understand

University of St.Gallen
Swiss Institute for Empirical Economic Research

SVT: Short vocational training
LVT: Long vocational training
OT: Orientation training
NOP: Nonparticipation

Allocation of individual unemployed to training programmes | 1

Decision tree of depth 3

Short training (SVT)

- Worked ≤ 20 months in last 2 years
- Unemployed ≤ 2 months last 10 years
- Born in Southern or Eastern EU, Turkey, Morocco

Long training (LVT)

- Worked > 20 months in last 2 years
- Worked > 105 months in last 10 years
- Living in specific areas

Orientation training (OT)

- Nobody

No programme participation (NOP)

- All others


University of St.Gallen
Swiss Institute for
Empirical Economic Research
1 | Introduction
2 | Methodology for a special case
3 | An example: Active labour market policies in Flanders
4 | Conclusions & road ahead

When does Causal Machine Learning help?

Identification

- No

Estimation & interpretation of effects of policy

- A lot ➔ a much richer set of (causal) information can be extracted from the data

Implementation of policy

- (B)GATEs for targeting larger groups
- Estimated allocation rules (*optimal policy*) for targeting at very specific level
  - Algorithm-based or -assisted allocation of policy

A remark on coding & software

Almost all researchers publish *R* and/or *Python* packages for their methods

- Other software plays only a minor role
- *Python* or *R*?
  - Industry: Python is close to be the ML & CML standard in industry
  - Academia: R is important but Python is catching-up (overtaking?)

Some coding skills in R and Python are useful when using these packages

Some dangers & possible pitfalls

Many effects are estimated: Researchers should resist selecting the *most interesting* effects

- Possible safeguard: Estimate only few GATEs guided by theory (& full reporting)

Wrong interpretation of effects

- Example: A GATE is a descriptive tool for causal effects (its not causal moderation!)

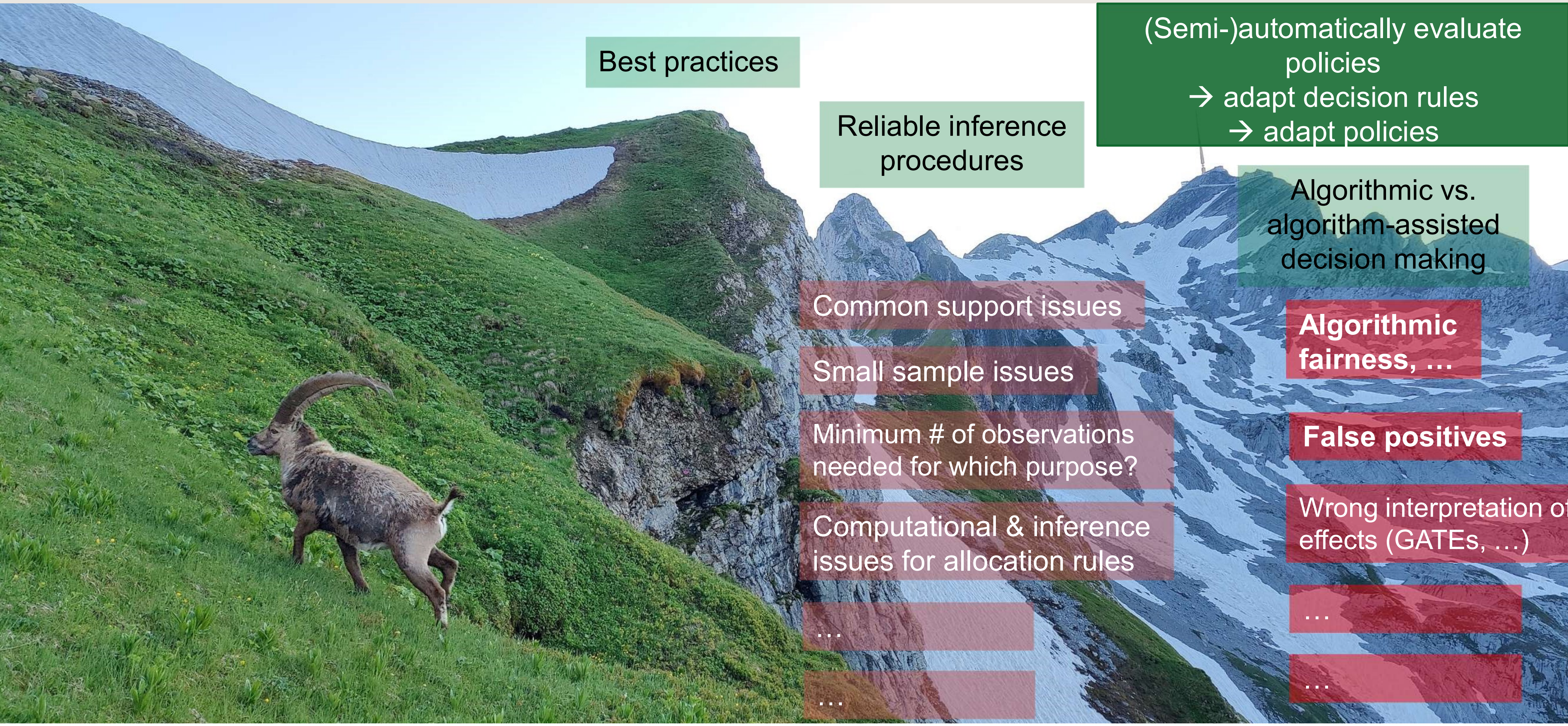
Common support issues

Insufficient sample size

- Reliable estimation of aggregate effects needs fewer observation than CATEs
- Complex functional forms (explicit or implicit) need more observations than simple ones
  - Estimators dividing by probabilities may be particularly vulnerable
- Robust inference needs more data than point estimates

University of St.Gallen
Swiss Institute for Empirical Economic Research

A map of the CML hike ahead



- Best practices
- Reliable inference procedures
- (Semi-)automatically evaluate policies
  → adapt decision rules
  → adapt policies
- Algorithmic vs. algorithm-assisted decision making

- Common support issues
- Small sample issues
- Minimum # of observations needed for which purpose?
- Computational & inference issues for allocation rules
- …
- …

- **Algorithmic fairness, …**
- **False positives**
- Wrong interpretation of effects (GATEs, …)
- …
- …

University of St.Gallen
Swiss Institute for
Empirical Economic Research

There are still many issues to investigate, but so far these methods seem to work nicely & appear to very helpful for understanding & improving public policies

Michael Lechner

Swiss Institute for Empirical Economic Research (SEW)
University of St. Gallen | Switzerland