# CAUSAL MACHINE LEARNING & ITS USE FOR PUBLIC POLICY

Verein für Socialpolitik, Basel, September 2022

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

University of St.Gallen
Swiss Institute for Empirical Economic Research

# Recent (micro-) econometric history: Credibility revolution

*Successful improvement of the identification of causal effects & credibility of results*

## The Sveriges Riskbank Prize in Economic Sciences in Memory of Alfred Nobel

**Replication Crisis**

**p-value Hacking**

## 2000

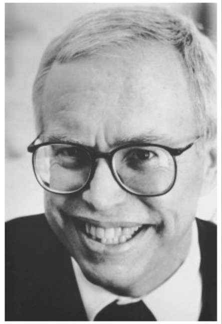


Photo from the Nobel Foundation archive.
James J. Heckman
Prize share: 1/2


Photo from the Nobel Foundation archive.
Daniel L. McFadden
Prize share: 1/2

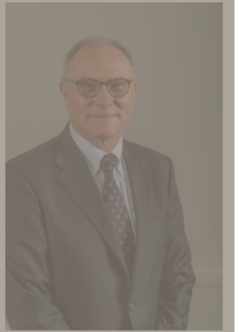
## 2019


© Nobel Media. Photo: A. Mahmoud
Abhijit Banerjee
Prize share: 1/3


© Nobel Media. Photo: A. Mahmoud
Esther Duflo
Prize share: 1/3


© Nobel Media. Photo: A. Mahmoud
Michael Kremer
Prize share: 1/3

## 2021


© Nobel Prize Outreach. Photo: Paul Kennedy
David Card
Prize share: 1/2


© Nobel Prize Outreach. Photo: Risdon Photography
Joshua D. Angrist
Prize share: 1/4


© Nobel Prize Outreach. Photo: Paul Kennedy
Guido W. Imbens
Prize share: 1/4

JH: "… methods for analyzing **selective samples**."

"for their **experimental approach** to alleviating global poverty."

JA & GI: "for their methodological contributions to the analysis of **causal relationships**".

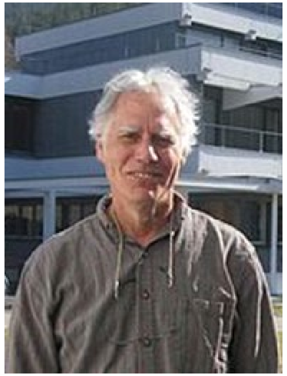
University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Econometrics benefited from other fields

**Potential Outcomes**

Donald Rubin
Statistican

**Dynamic Treatments**



James Robins
Epidemiologist

**Directed Acyclical Graphs (DAGs)**

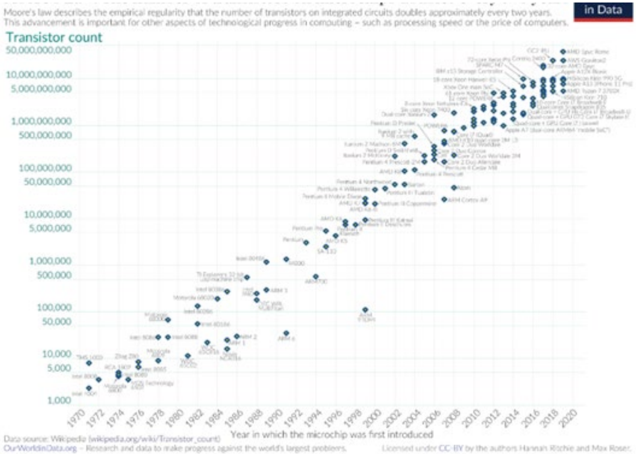
Judea Pearl
Computer Scientist

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Other recent developments

**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**

# Now & in the (near?) future | 1

## Causal Machine Learning

- 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 (this talk)
  - Epidemiology (personalized medicine, …)
  - Marketing (targeted marketing & political campaigns, …)
  - …

# 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.

# How useful is CML for different *research designs*? | 1

Different sets of identifying assumptions (research designs) identify causal parameters for different subpopulations

Usefulness depends on objective of estimation …

- Aggregate effects - heterogenous effects - direct decision support

# Does CML help for *aggregate* effects?

Useful when there are covariate**s** (*X*) and/or instrument**s** (*Z*)

- Flexible (& possibly efficient) ways to take *X*, *Z* into account
- *Selection-on-observables*, *IV* with *X*, *Z*, *DiD* with *X*

Not useful when covariates or instruments are not needed

- *Experiments*, standard *RDD*, *IV* & *DiD* with few *X*

# Does CML help for *heterogeneous* effects?

Useful when effects are identified for full population of interest

- Experiments, selection-on-observables (& IV with 100% compliance)

Limited usefulness when effects are identified for some subpopulation *only*

- *IV* & fuzzy *RDD*: *Compliers* & *local-to-cut-off compliers*
- Sharp *RDD*: Local-to-cut-off population
- *DiD*: Selected by previous assignment rule (*treated*)

Not useful when heterogenous effects cannot be identified

- Synthetic controls

# Does CML help for *algorithm-based or -assisted decision making?*

Decision rules must be based on pre-determined characteristics ($X$) only

- Compliance or treatment status unknown

Useful when effects are identified for full population of interest

- Experiments, selection-on-observables (& IV with 100% compliance)

Not useful when effects are identified for some subpopulation *only*

- Additional (homogeneity) assumptions needed

# Many topics not covered in this talk ...

CML in the private sector

Dynamic allocation & evaluation

- Bandits / reinforcement learning

More complex causal structures

- Dynamics (sequences)
- Mediation
- Networks
- ...

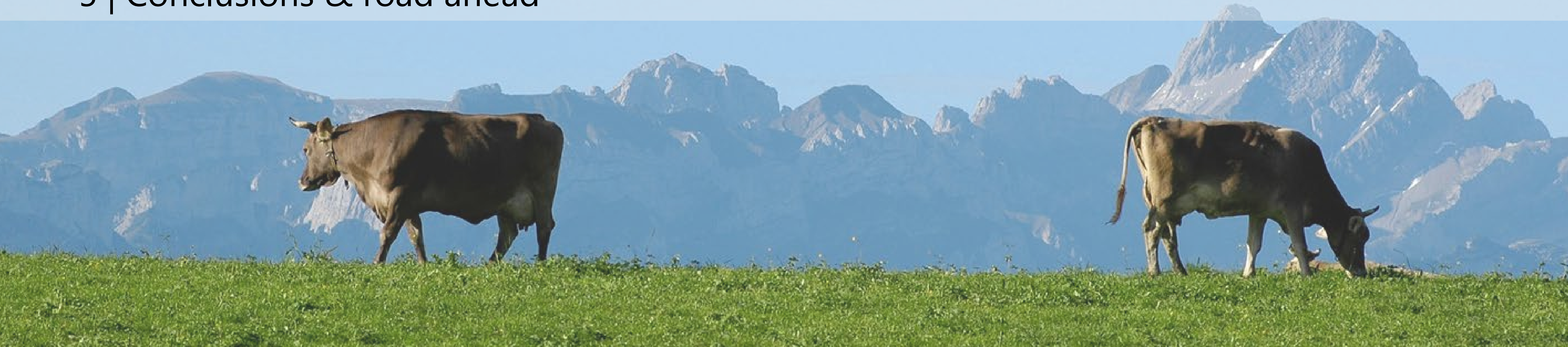
ML for other purposes

- Variable generation
  - Text
  - Pictures
  - Natural language processing
  - ...
- Prediction

...

University of St.Gallen
Swiss Institute for
Empirical Economic Research





# Machine Learning | 1

## What is machine learning?

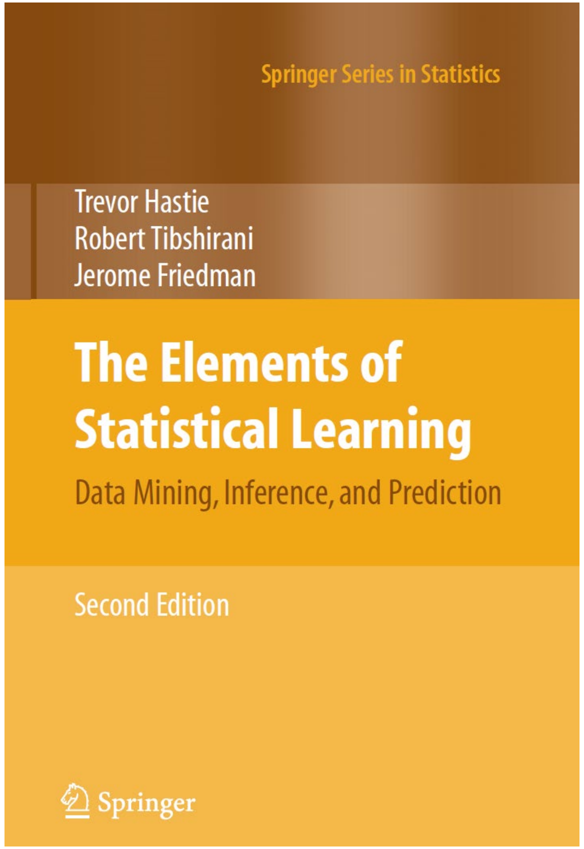
- Everything & nothing ➔ Here: *Statistical Learning*
- Flexible prediction methods

## Types of ML

- **Supervised** (*y, x)* & unsupervised learning (*x*)
- Classification (discrete *y*) & **regression** (cond. expectations)

## Examples of supervised, regression ML

- All classical econometric estimators
- Regularized, shrinkage estimators (Lasso, Ridge, ELN, …)
- Neural Networks
- Trees & Random Forests



This table oversimplifies (a bit)

# Classical Econometrics & Machine Learning

| Issue | Classical econometrics | Supervised statistical learning |
|---|---|---|
| **Target of interest (θ)** | Structural & causal parameters (*low dimensional*) | Prediction ($Ey\|x$) or classification of $y$ |
| **Sample analogue of θ** | -- | $y$ |
| **Judging quality of estimation** | *Indirect* (fit, …), in-sample | *Direct* ($\widehat{y}$ vs $y$), out-of-sample |
| **Inference & theoretical properties** | Very important | Less important (irrelevant?) |
| **Sample size (*N*)** | Large $N$ is nice to have | Large $N$ may be required |
| **# of variables (k)** | Much smaller than $N$ | Smaller or larger than $N$ |
| **Preferred model complexity** | Simple, likely to be parametric (linearity popular) | Complicated (overparametrized; nonlinear) |
| **Names of methods** | Boring | Cool |

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# 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

# Notation

$D$ Treatment

$Y^0$ Potential outcome for $D=0$

$Y^1$ Potential outcome for $D=1$

$Y = D\, Y^1 + (1-D)\, Y^0$

$X$ All confounders & heterogeneity variables

$H$ Specific heterogeneity variables (low dimensional)

Data contains realisations of $D$, $Y$, $X$, $H$

# Identification

## Identifying assumptions

- $D$ independent of $Y^0$, $Y^1$ given $X$
- Common support, no interactions between units, exogeneity of $X$

## Implications

- Treated ($D=1$) & untreated ($D=0$) may have different distributions of $X$
- Distribution of unobservables (that influence $Y^d$) identical for $D=1$ & $D=0 \mid X$
- Credibility depends on the information available in the data

$D$: Treatment (0 or 1)
$Y^1$: Outcome when $D=1$
$Y^0$: Outcome when $D=0$
$X$: All confounder & heterogeneity variables
$Z$: Specific heterogeneity variables (low dim.)
Observable: $X,\ Z,\ Y=DY^1+(1-D)Y^0$

# Average effects at different aggregation levels

**Individualized (Conditional) Average Treatment Effects**

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

**Group (Conditional) Average Treatment Effects**

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

**Average Treatment Effects**

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

Effects for t treat non-treated
Quantile effect
...

$$IATE(x) = E(Y \mid X = x, D = 1) - E(Y \mid X = x, D = 0)$$
$$GATE(z) = E_{Z=z} IATE(x)$$
$$ATE = E_X IATE(x)$$

# The estimation problem | 1

## Naïve ML estimator

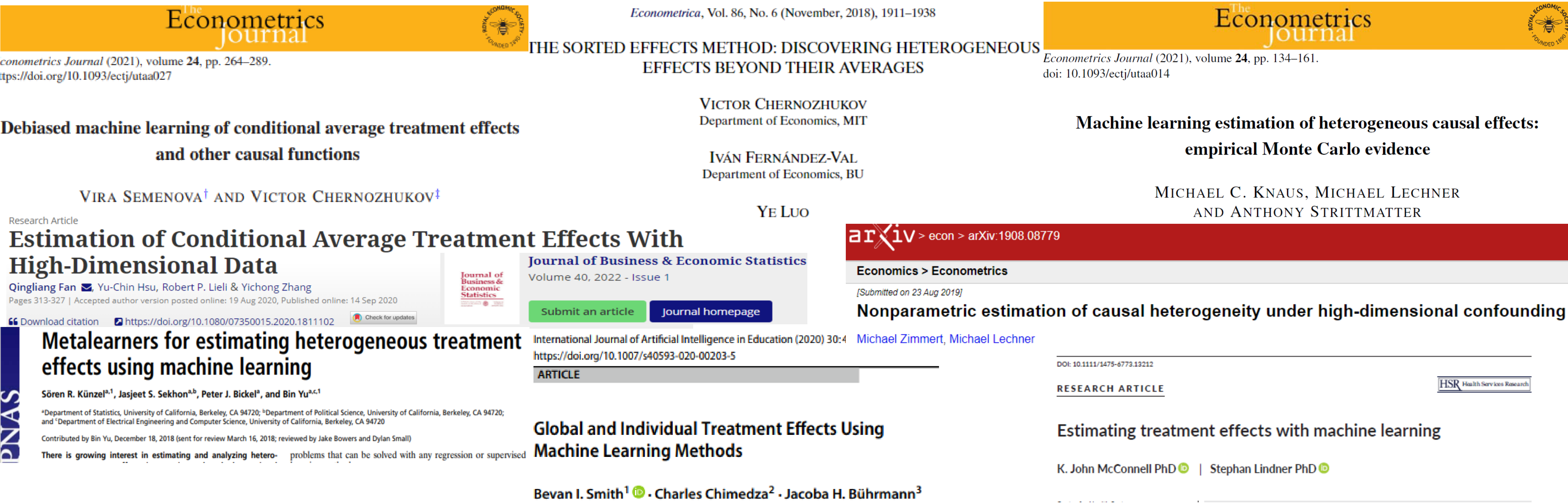
- ML estimation of *E(Y|X=x, D=d)* in treated (*d=1*) & non-treated subsample (*d=0*)
  - *IATE(x)*: Predictions of *Y* for treated - predictions of *y* for non-treated
  - *GATE(h)* : Average *IATE*($x_i$) *with* $h_i$ *same/similar to h*
  - *ATE*: Average *IATE*($x_i$)

## Naïve estimator may be a bad idea

- Predictions of ML estimators are usually biased
  - MSE optimal-prediction of *E(Y|X=x, D=d)* but inference may not work (too much bias)
- Estimating a difference well is different from estimating its components well
  - Only difference of estimation errors matters

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Huge literature on *best* estimator for each parameter



# Comprehensive vs. parameter specific approaches

Using different estimation concepts for different parameters is not attractive in practice

- Computationally intensive
  - Many estimators, lot's of different tuning parameters
- Substantial effort to understand specifics of estimators & monitor problems in all estimations
- There may be a lack of internal consistency (GATEs may not add up to ATE, etc.)
- …

Comprehensive estimation approaches

- Option I: Use ML inside specific moment conditions (double/debiased machine learning, DML)
- Option II: Change a ML into a CML (causal forests, etc.)

University of St.Gallen
Swiss Institute for Empirical Economic Research

# Comprehensive methodologies | 1

## Double / debiased machine learning (DML): Theory

- Main idea
  - Use specific moment condition that fulfils **N**eyman **O**rthogonality **C**ondition
    - **Here**: Dependence of moment condition on propensity score (*P(D=1\X=x)* & outcome equations (*EY|X=x, D=d)*) is such that small errors in those (*nuisance*) functions do not affect distribution of estimator
  - 1st step: Use ML to estimate nuisance functions
  - 2nd step: Solve moment conditions given estimated nuisance functions
- This principle has very wide applicability to many estima
- It is related to *Double Robustness* (in treatment effect est
  - DR in parametrics: OK if propensity score *or* outcome equatio
  - DR in CML: OK if propensity score and outcome equations ar

The Econometrics Journal

*Econometrics Journal* (2018), volume **21**, pp. C1–C68.
doi: 10.1111/ectj.12097

**Double/debiased machine learning for treatment and structural parameters**

VICTOR CHERNOZHUKOV†, DENIS CHETVERIKOV‡, MERT DEMIRER†, ESTHER DUFLO†, CHRISTIAN HANSEN§, WHITNEY NEWEY† AND JAMES ROBINS‖

**Nice technical survey**

Semiparametric Doubly Robust Targeted Double Machine Learning: A Review *†

Edward H. Kennedy
Department of Statistics & Data Science
Carnegie Mellon University

# Extensions of DML theory

*Econometrica*, Vol. 90, No. 3 (May, 2022), 967–1027

## AUTOMATIC DEBIASED MACHINE LEARNING OF CAUSAL AND STRUCTURAL EFFECTS

VICTOR CHERNOZHUKOV
Department of Economics, Massachusetts Institute of Technology

WHITNEY K. NEWEY
Department of Economics, Massachusetts Institute of Technology and NBER

RAHUL SINGH
Department of Economics, Massachusetts Institute of Technology

*Econometrica*, Vol. 90, No. 4 (July, 2022), 1501–1535

## LOCALLY ROBUST SEMIPARAMETRIC ESTIMATION

VICTOR CHERNOZHUKOV
Department of Economics, MIT

JUAN CARLOS ESCANCIANO
Department of Economics, Universidad Carlos III de Madrid

HIDEHIKO ICHIMURA
Department of Economics, University of Arizona and Department of Economics, University of Tokyo

WHITNEY K. NEWEY
Department of Economics, MIT and NBER

JAMES M. ROBINS
Epidemiology, School of Public Health, Harvard University

## Automatic Debiased Machine Learning for Dynamic Treatment Effects

Victor Chernozhukov, MIT
Whitney Newey, MIT
Rahul Singh, MIT
Vasilis Syrgkanis, Microsoft Research

Published as a conference paper at ICLR 2022

## RIESZNET AND FORESTRIESZ: AUTOMATIC DEBIASED MACHINE LEARNING WITH NEURAL NETS AND RANDOM FORESTS

**Victor Chernozhukov**, MIT
**Whitney K. Newey**, MIT
**Víctor Quintas-Martínez**, MIT
**Vasilis Syrgkanis**, Microsoft Research

## Automatic Debiased Machine Learning via Neural Nets for Generalized Linear Regression*

Victor Chernozhukov, *MIT*
Whitney K. Newey, *MIT*
Victor Quintas-Martinez, *MIT*
Vasilis Syrgkanis, *Microsoft Research*

## A Simple and General Debiased Machine Learning Theorem with Finite Sample Guarantees

**Victor Chernozhukov**, MIT Economics, vchern@mit.edu
**Whitney K. Newey**, MIT Economics, wnewey@mit.edu
**Rahul Singh**, MIT Economics, rahul.singh@mit.edu

University of St.Gallen
Swiss Institute for Empirical Economic Research

# DML | Example: ATE for binary treatment

- Moment condition

$$\psi(Y,X,D,ATE,g,m) = g(1,X) - g(0,X) + \frac{[Y - g(1,X)]D}{m(X)} - \frac{[Y - g(0,X)](1-D)}{1-m(X)} - ATE$$

$$g_0(d,x) = E(Y \mid X = x, D = d); \qquad m_0(x) = P(D = 1 \mid X = x)$$

$$E\psi(Y,X,D,ATE_0,\eta_0) = 0 \qquad \eta = [g_0(0,x), g_0(1,x), m_0(x), \forall x]$$

$$NOC: \partial_\eta E\psi(Y,X,D,ATE_0,\eta)|_{\eta=\eta_0} = 0$$

- Estimator

$$\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N} \hat{g}_{-i}(1,x_i) - \hat{g}_{-i}(0,x_i) + \frac{[y_i - \hat{g}_{-i}(1,x_i)]d_i}{\hat{m}_{-i}(x_i)} - \frac{[y_i - \hat{g}_{-i}(0,x_i)](1-d_i)}{1-\hat{m}_{-i}(x_i)}$$

ML-estimated (X-fitted) functions

$$\sqrt{N}\frac{\widehat{ATE} - ATE_0}{E[\psi(Y,X,D,ATE_0,g_0,m_0)^2]} \xrightarrow{d} N(0,1) \qquad \text{(efficient, Hahn, 1998)}$$

# Nice operationalisation of DML for programme evaluation

The Econometrics Journal

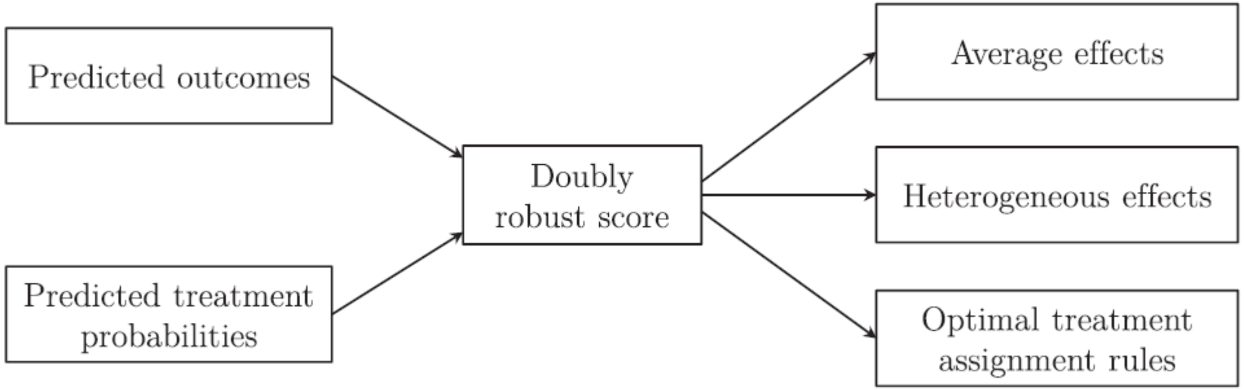
ROYAL ECONOMIC SOCIETY · FOUNDED 1890

*Econometrics Journal* (2022), volume **00**, pp. 1–26.
https://doi.org/10.1093/ectj/utac015

**Double machine learning-based programme evaluation under unconfoundedness**

MICHAEL C. KNAUS

$$\boxed{\begin{aligned}&\hat{g}_{-i}(1,x_i)-\hat{g}_{-i}(0,x_i)+\\&+\frac{[y_i-\hat{g}_{-i}(1,x_i)]d_i}{\hat{m}_{-i}(x_i)}-\frac{[y_i-\hat{g}_{-i}(0,x_i)](1-d_i)}{1-\hat{m}_{-i}(x_i)}\end{aligned}}$$



University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Comprehensive methodologies | Change the ML estimator | Example

## Causal Tree

PNAS

**Recursive partitioning for heterogeneous causal effects**

Susan Athey[a,1] and Guido Imbens[a]

[a]Stanford Graduate School of Business, Stanford University, Stanford, CA 94305

## Causal Forest

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2018, VOL. 113, NO. 523, 1228–1242, Theory and Methods
https://doi.org/10.1080/01621459.2017.1319839

Taylor & Franci
Taylor & Francis Group

Check for updates

**Estimation and Inference of Heterogeneous Treatment Effects using Random Forests**

Stefan Wager and Susan Athey

## Modified Causal Forest (mcf)

- Some small changes to make CF more comprehensive (& improve on it)
  - IATEs are weighted means of *y*
  - Obtain GATEs, ATEs by aggregating the weights …

arXiv > econ > arXiv:2209.03744

Economics > Econometrics

*[Submitted on 8 Sep 2022]*

**Modified Causal Forest**

Michael Lechner, Jana Mareckova

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Comprehensive methodologies | 3

Somewhere in-between these ‘worlds’ of DML and CF

- Local maximum likelihood estimation
- RF provides the local weighting scheme

*The Annals of Statistics*
2019, Vol. 47, No. 2, 1148–1178
https://doi.org/10.1214/18-AOS1709


**GENERALIZED RANDOM FORESTS**

BY SUSAN ATHEY*, JULIE TIBSHIRANI† AND STEFAN WAGER*

…

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# 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

## It raises many questions

- Methodological (statistical properties)
- Computational (in particular for multiple & cont. treatments)
- Ethical
- Practical

*Econometrica*, Vol. 72, No. 4 (July, 2004), 1221–1246

STATISTICAL TREATMENT RULES FOR HETEROGENEOUS POPULATIONS

BY CHARLES F. MANSKI[1]

*Econometrica*, Vol. 86, No. 2 (March, 2018), 591–616

WHO SHOULD BE TREATED? EMPIRICAL WELFARE MAXIMIZATION METHODS FOR TREATMENT CHOICE

TORU KITAGAWA
Cemmap and Department of Economics, University College London

ALEKSEY TETENOV

Machine Learning (2022) 111:2741–2768

**Optimal policy trees**

Maxime Amram[1] · Jack Dunn[1] · Ying Daisy Zhuo[1]

*Econometrica*, Vol. 89, No. 1 (January, 2021), 133–161

POLICY LEARNING WITH OBSERVATIONAL DATA

SUSAN ATHEY
Stanford Graduate School of Business, Stanford University

STEFAN WAGER

**Operations Research**

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Offline Multi-Action Policy Learning: Generalization and Optimization

Zhengyuan Zhou, Susan Athey, Stefan Wager

University of St.Gallen
Swiss Institute for
Empirical Economic Research

1 | Introduction

2 | Machine learning & classical econometrics

3 | Methodology for a special case

4 | An example: Active labour market policies in Flanders

5 | Conclusions & road ahead

University of St.Gallen
Swiss Institute for
Empirical Economic Research

# Empirical example

arXiv > econ > arXiv:1912.12864

**Economics > Econometrics**

*[Submitted on 30 Dec 2019 (v1), last revised 6 May 2020 (this version, v2)]*

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

Bart Cockx, Michael Lechner, Joost Bollens

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

**mcf 0.2.6**

Latest version

```
pip install mcf
```

Released: Aug 12, 2022

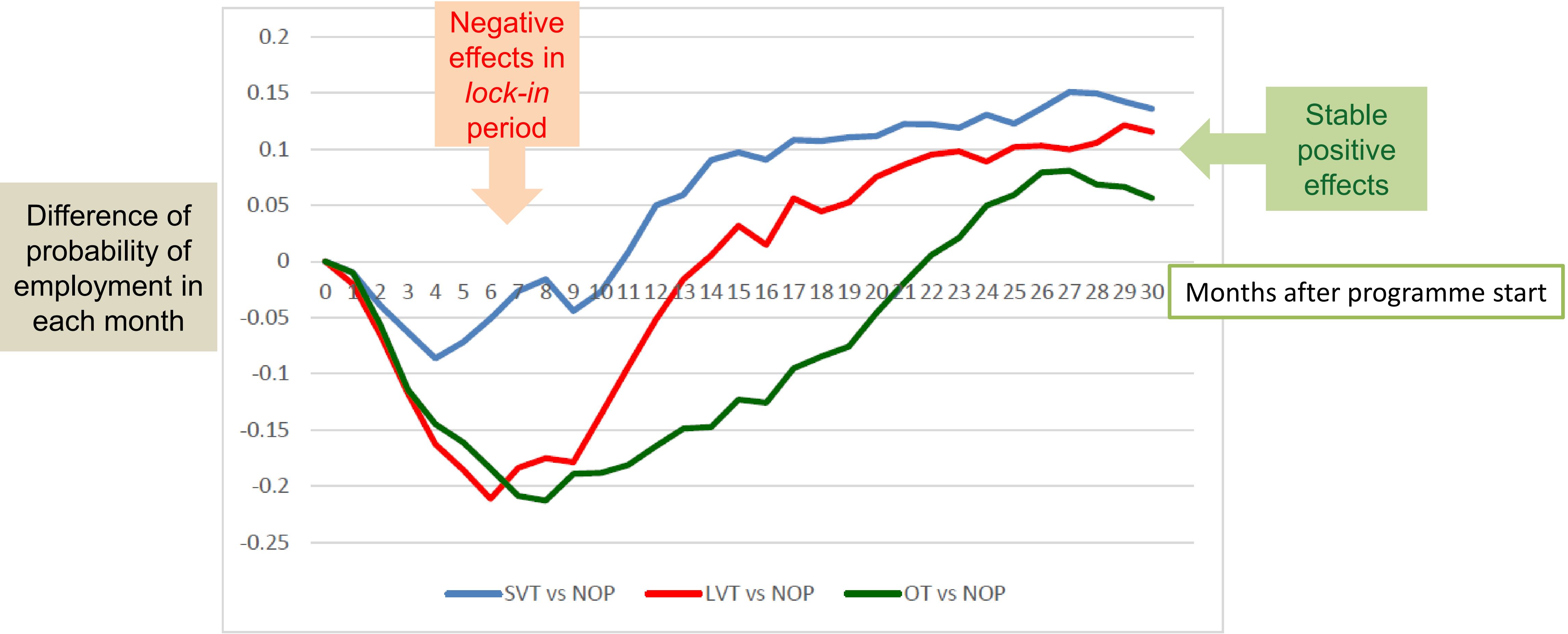
mcf is a powerful package to estimate heterogeneous treatment effects for multiple treatment models in a selection-on-observables setting and learn optimal policy rules

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

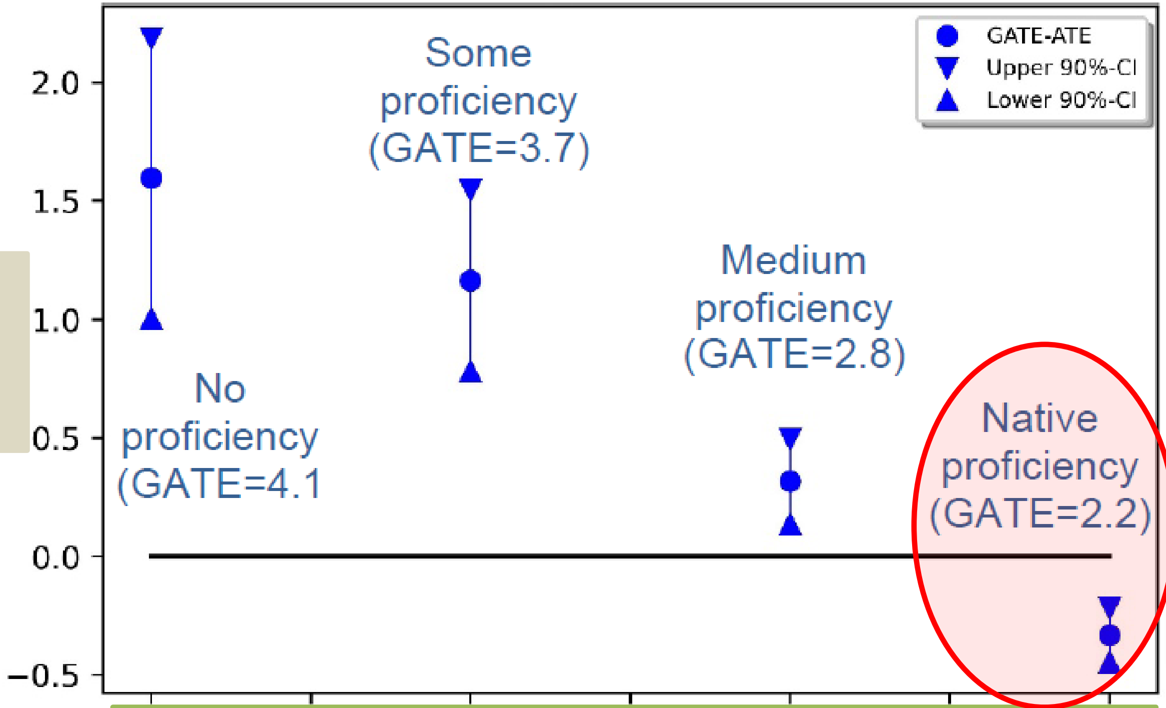


**SVT**: Short vocational training
**LVT**: Long vocational training
**OT**: Orientation training
**NOP**: Nonparticipation

# Group ATEs minus ATE | 1

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

Difference of # of months employed 30 months after start




Better than average

Worse than average

Different subpopulations

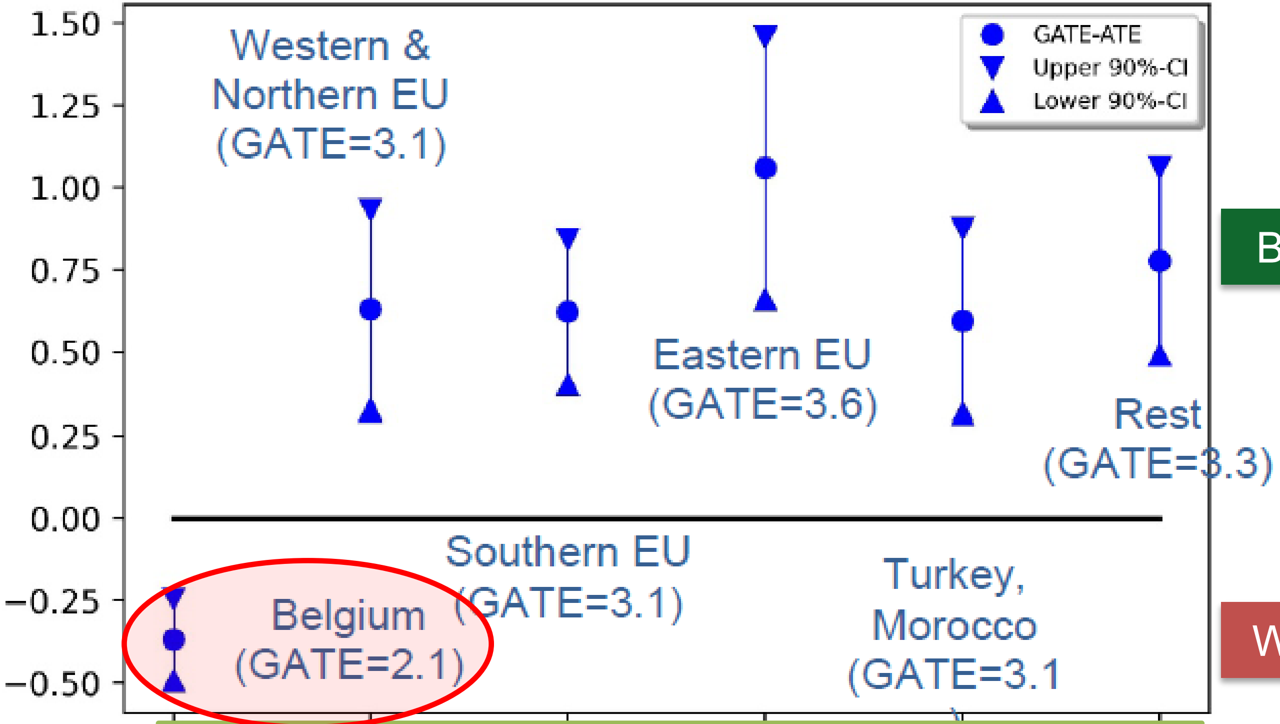
Dutch proficiency displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. Dutch proficiency varies between no proficiency (0) and native proficiency (3).

**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

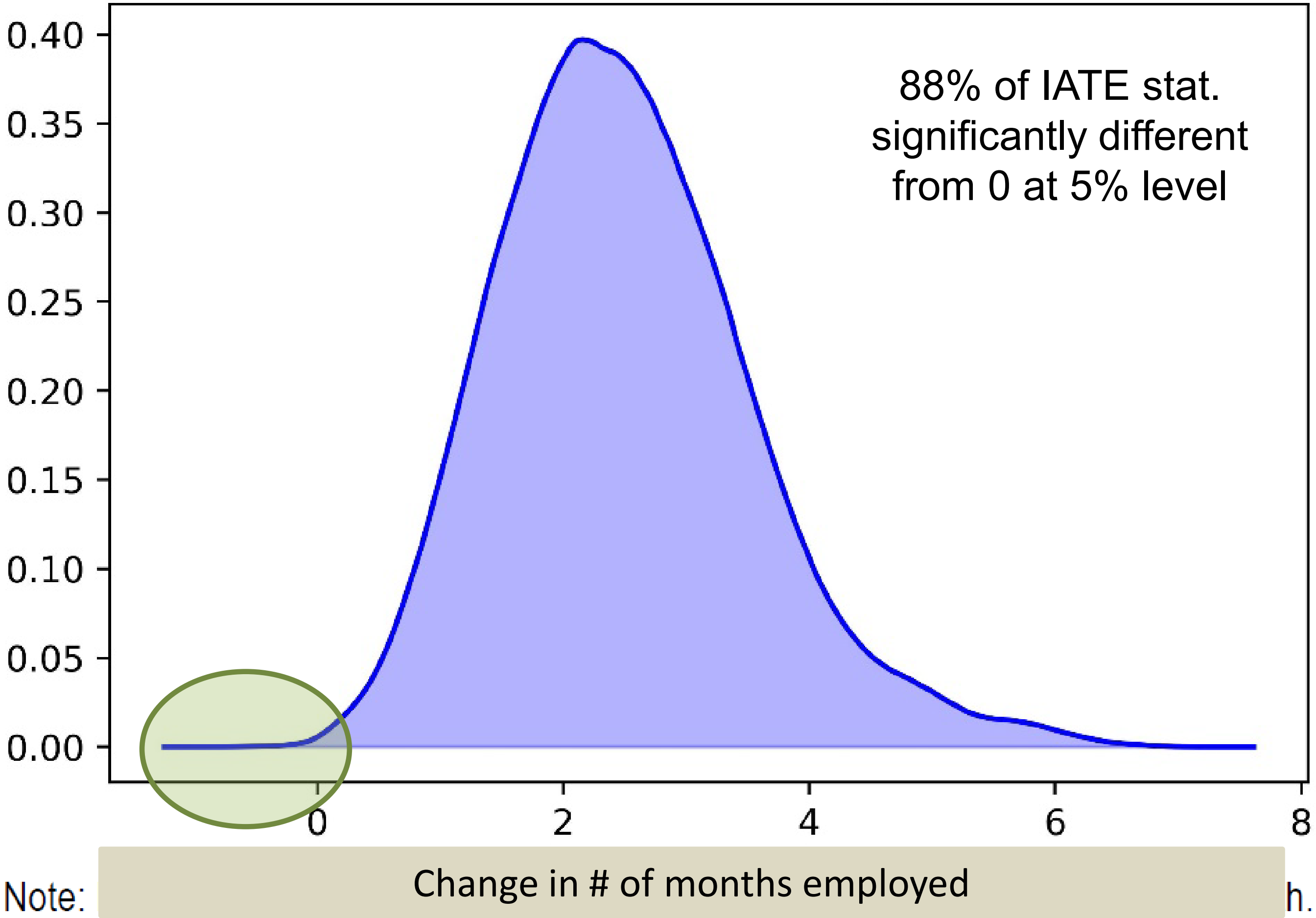
Country of birth displayed on horizontal axis. Vertical axis denotes difference of respective GATE with ATE. (GATE-ATE) and its 90% confidence interval shown. The vertical axis measures the deviation of the GATE from the ATE.

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




Note: h.

# 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

*Alternative method*

*Econometrica*, Vol. 86, No. 6 (November, 2018), 1911–1938

THE SORTED EFFECTS METHOD: DISCOVERING HETEROGENEOUS EFFECTS BEYOND THEIR AVERAGES

VICTOR CHERNOZHUKOV
Department of Economics, MIT

IVÁN FERNÁNDEZ-VAL
Department of Economics, BU

YE LUO

## 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

# 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 | Machine learning & classical econometrics

3 | Methodology for a special case

4 | An example: Active labour market policies in Flanders

5 | 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

- 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*?
  - Python close to be the ML & CML standard in industry
  - R import in research community, Python is gaining importance

Good programming skills are needed if

- Existing methods are adapted
- New methods developed

# 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

- 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

# The CML hike ahead