Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

ECONOMIC AI

Estimation of Treatment Effects with Text and Images

Prof. Dr. Martin Spindler
Universität Hamburg & Economic AI

PSI Conference
8.6.-11.6.2025

Co-authors:

- Prof. Victor Chernozhukov (MIT)
- Dr. Philipp Bach (UHH)
- Dr. Sven Klaassen (UHH & Economic AI)
- Jan Teichert-Kluge (UHH)
- Suhas Vijaykumar (UCSD)

Introduction

Causal Machine Learning

Why? 〉〉 **Causal Machine Learning** 〈〈 AI

CAUSAL MODELING

- Learning causal relationships
- Going beyond correlations
- Pioneers: Pearl, Rubin, Imbens (Nobel Prize 2021)

AI/ MACHINE LEARNING

- Learning complex patterns in data
- Correlation based
- Good at forecasting / prediction

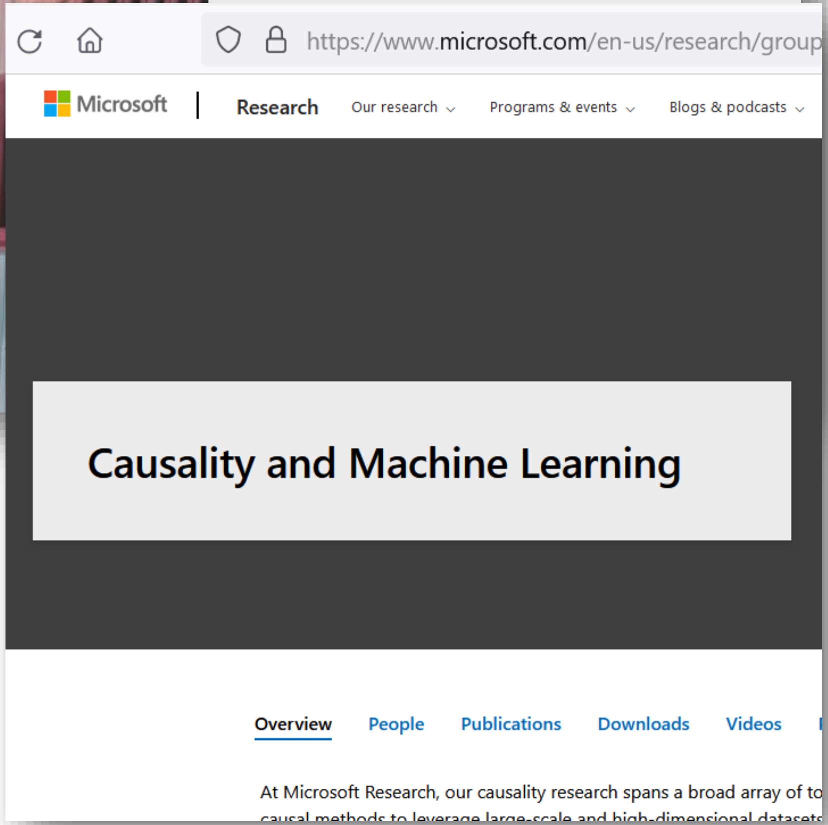
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Causal Machine Learning

WILL KNIGHT BUSINESS OCT 8, 2019 7:00 AM

An AI Pioneer Wants His Algorithms to Understand the 'Why'

Deep learning is good at finding patterns in reams of data, but can't explain how they're connected. Turing Award winner Yoshua Bengio wants to change that.

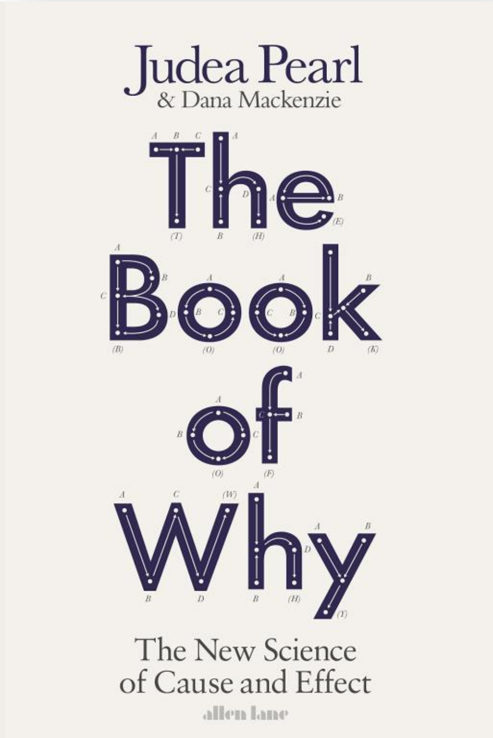

https://www.microsoft.com/en-us/research/group

Microsoft | Research Our research Programs & events Blogs & podcasts

Causality and Machine Learning

Overview People Publications Downloads Videos

At Microsoft Research, our causality research spans a broad array of to



DoubleML

doubleml.org

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Predictive vs. Causal ML

Predictive ML	Causal ML
How can we build a good prediction rule, $\boldsymbol{f}(\boldsymbol{X})$, that uses features $\boldsymbol{X}$ to predict? **Example:** mortality prediction *“How large it the risk that a patient will die in hospital?"*	What is the causal effect of a treatment $\boldsymbol{D}$ on an outcome $\boldsymbol{Y}$? **Example:** treatment evaluation *“Is a new drug better than an old one?“*

Methods

Randomized Control Trial

- “Gold standard” in medicine and science
- Widely used and well established
- But also non-standard situations
- Innovations: heterogenous TE, precision

Observational Data / RWD

- Learning causal effects from RWD
- Additional assumptions required
- Toolbox for causal inference in such situations
- Innovation: ML / AI for complex data

Hybrid Methods: Instrumental Variables

- Invented in economics but has become popular more broadly
- Used in settings when Randomized Trials are not feasible and/or new policies / policy predictions are needed
- Can handle large dimensions with solid statistical properties

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Double Machine Learning

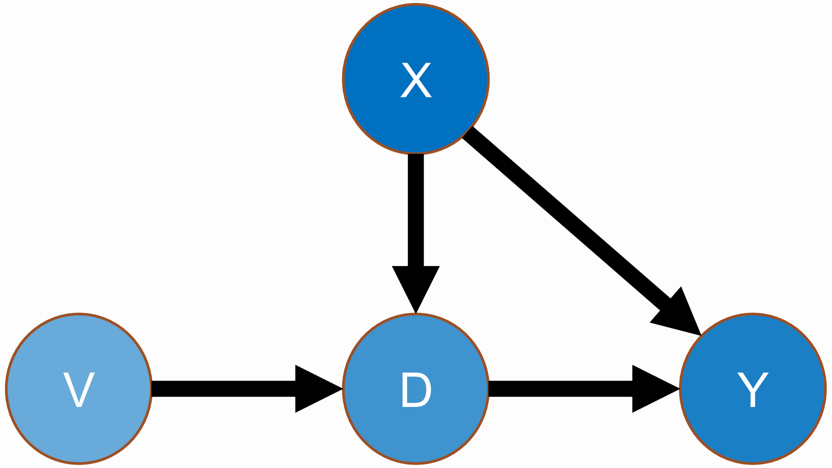
Motivation

We consider the linear regression model in a high-dimensional setting (potentially p >> n)

$$\boxed{Y = D\theta_0 + X_1\beta_1 + \ldots X_p\beta_p + \varepsilon, \quad \mathrm{E}[\varepsilon \mid X, D] = 0,}$$

- Y - outcome variable
- D - policy/treatment variable
- θ_0 - parameter of interest
- $\beta = (\beta_1, \ldots, \beta_p)'$ - nuisance parameter
- $X = (X_1, \ldots, X_p)'$ is a vector of other covariates, called "controls" or "confounders" in the sense that

$$D = \gamma' X + \nu, \quad E[\nu|X] = 0.$$



Partially Linear Regression

Partially linear regression (PLR) model

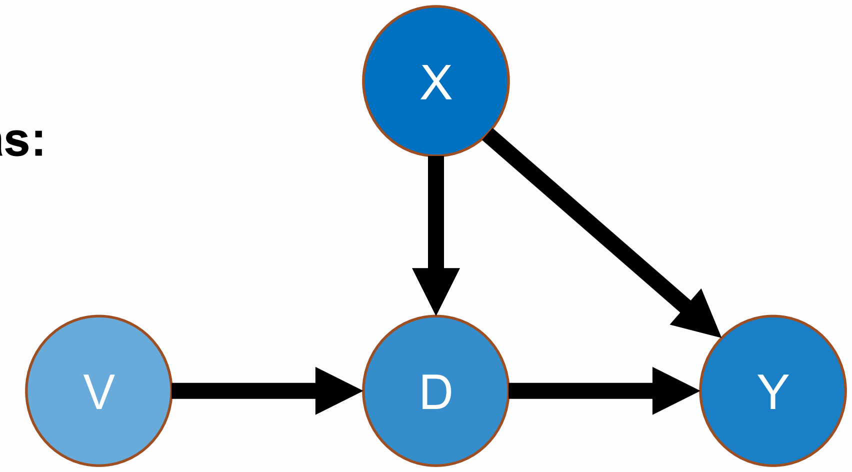
$$
\begin{aligned}
Y &= D\theta_0 + g_0(X) + \zeta, & \mathbb{E}[\zeta \mid D, X] &= 0,\\
D &= m_0(X) + V, & \mathbb{E}[V \mid X] &= 0,
\end{aligned}
$$

with

- Outcome variable $\boldsymbol{Y}$
- Policy or treatment variable of interest $\boldsymbol{D}$
- High-dimensional vector of confounding covariates $\boldsymbol{X} = (\boldsymbol{X_1}, \dots, \mathbf{X_p})$
- Stochastic errors $\boldsymbol{\zeta}$ and $\boldsymbol{V}$

Problem of simple "plug-in" approaches: Regularization bias:

If we use an ML model to estimate $\hat{\boldsymbol{g}}$ and simply plug in the predictions $\hat{\boldsymbol{g}}$, the final estimate on $\boldsymbol{\theta_0}$ will not be unbiased and neither be asymptotically normal

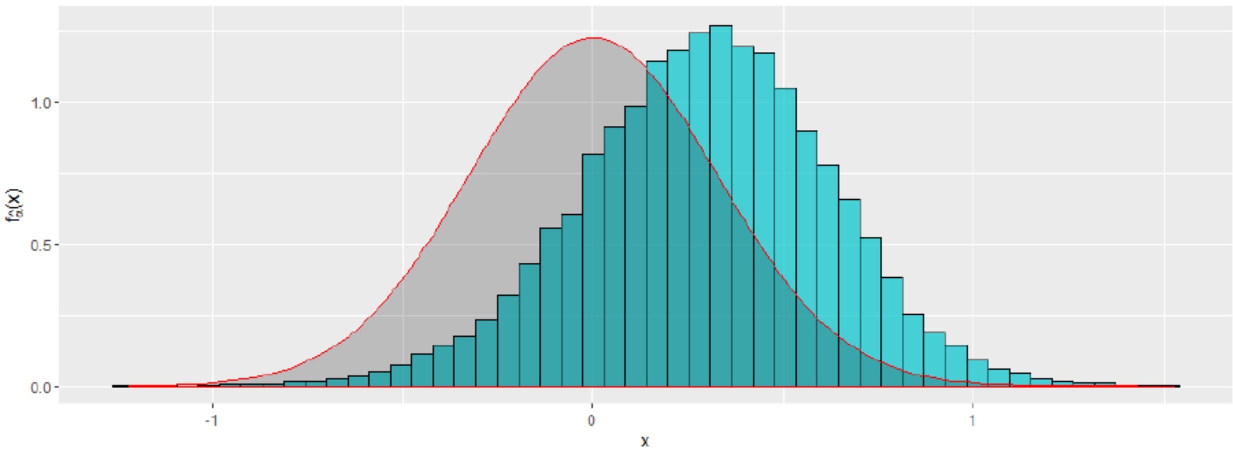


„Naive“ or Prediction-Based ML Approach is Bad

Naive/Textbook Inference:

1. Select controls terms by running Lasso (or variants) of Y_i on X_i
2. Estimate θ_0 by least squares of Y_i on D_i and selected controls, apply standard inference

The distribution of $\hat{\theta}_0 - \theta_0$ looks like this:

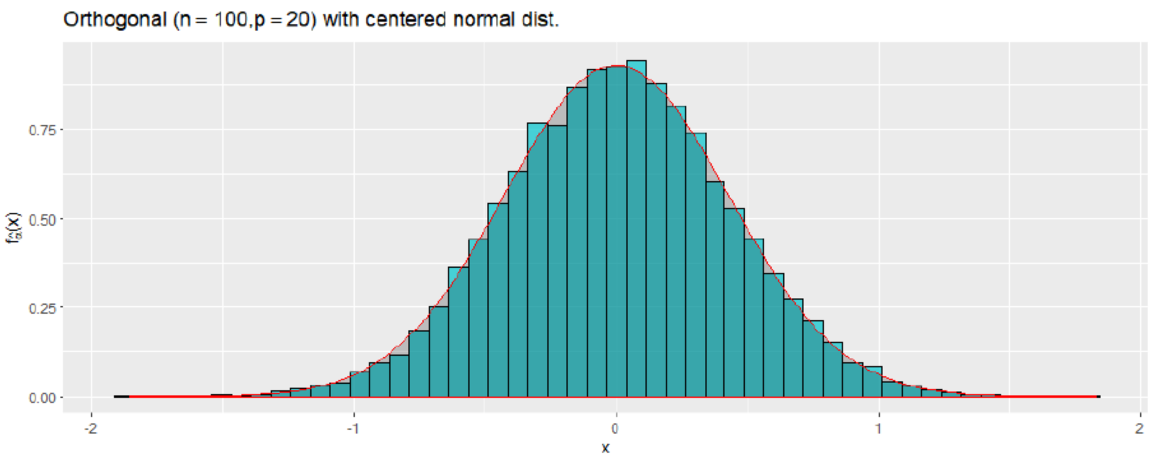


The Double Machine Learning Approach

1. Predict Y and D using X by $\mathrm{E}[Y|X]$ and $\mathrm{E}[D|X]$, obtained using Lasso, Random Forest or other "best performing" ML tools.
2. Residualize $W = Y - \mathrm{E}[Y|X]$ and $V = D - \mathrm{E}[D|X]$
3. Regress W on V to get θ_0

Frisch-Waugh-Lovell (1930s) style with ML methods

The distribution of $\hat{\theta}_0 - \theta_0$ looks like this:



The Key Ingredients of DML

1. Neyman Orthogonality

The inference is based on a score function $\boldsymbol{\psi}(\boldsymbol{W};\boldsymbol{\theta},\boldsymbol{\eta})$ that satisfies

$$\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{W};\boldsymbol{\theta}_0,\boldsymbol{\eta}_0)] = \mathbf{0}$$

Where $\boldsymbol{W} := (\boldsymbol{Y},\boldsymbol{D},\boldsymbol{X},\boldsymbol{Z})$ and with $\boldsymbol{\theta}_0$ being the unique solution that obeys the **Neyman orthogonality condition**

$$\partial_{\boldsymbol{\eta}}\, \mathbb{E}[\boldsymbol{\psi}(\boldsymbol{W};\boldsymbol{\theta}_0,\boldsymbol{\eta})]\Big|_{\boldsymbol{\eta}=\boldsymbol{\eta}_0} = \mathbf{0}$$

- For many models the Neyman orthogonal score functions are linear in $\boldsymbol{\theta}$

$$\boxed{\psi(W;\theta,\eta) = \psi_a(W;\eta)\theta + \psi_b(W;\eta)}$$

- The estimator $\tilde{\boldsymbol{\theta}}_0$ then takes the form

$$\boxed{\tilde{\theta}_0 = -\left(\mathbb{E}_N[\psi_a(W;\eta)]\right)^{-1}\mathbb{E}_N[\psi_b(W;\eta)]}$$

PLR example: Orthogonality by including the first-stage regression, i.e., the regression relationship of the treatment variable $\boldsymbol{D}$ and the regressors $\boldsymbol{X}$

Orthogonal score function

$$\boxed{\psi(\cdot) = (Y - E[Y|X] - \theta(D - E[D|X]))(D - E[D|X])}$$

Neyman Orthogonality

The two strategies rely on very different moment conditions for identifying and estimating $\boldsymbol{\theta}_0$

$$\mathbb{E}[\psi(W, \theta_0, \eta_0)] = 0$$

Naive approach

$$\psi(W, \theta_0, \eta) = (Y - D\theta_0 - g_0(X))D$$

Regression adjustment score

$$\eta = g(X),$$
$$\eta_0 = g_0(X),$$

FWL partialling out

$$\psi(W, \theta_0, \eta_0) = ((Y - E[Y|X]) - (D - E[D|X])\theta_0)\,(D - E[D|X])$$

Neyman-orthogonal score (Frisch-Waugh-Lovell)

$$\eta = (g(X), m(X)),$$
$$\eta_0 = (g_0(X), m_0(X)) = (\mathbb{E}[Y \mid X], \mathbb{E}[D \mid X])$$

Both estimators solve the empirical analog of the moment conditions: $$\boxed{\frac{1}{n}\sum_{i=1}^{n} \psi(W_i, \theta, \hat{\eta}_0) = 0,}$$

where instead of unknown nuisance functions we plug-in their ML-based (hold-out) estimators

The Key Ingredients of DML

2. High-Quality Machine Learning Estimators

The nuisance parameters are estimated with high-quality (fast-enough converging) machine learning methods.

- Different structural assumptions on $\boldsymbol{\eta_0}$ lead to the use of different machine-learning tools for estimating $\boldsymbol{\eta_0}$ (Chernozhukov et al., 2018, Chapter 3)

3. Sample Splitting

To avoid the biases arising from overfitting, a form of sample splitting is used at the stage of producing the estimator of the main parameter $\boldsymbol{\theta_0}$.

- Cross-fitting performs well empirically (efficiency gain by switching roles)

Heterogenous Treatment effects

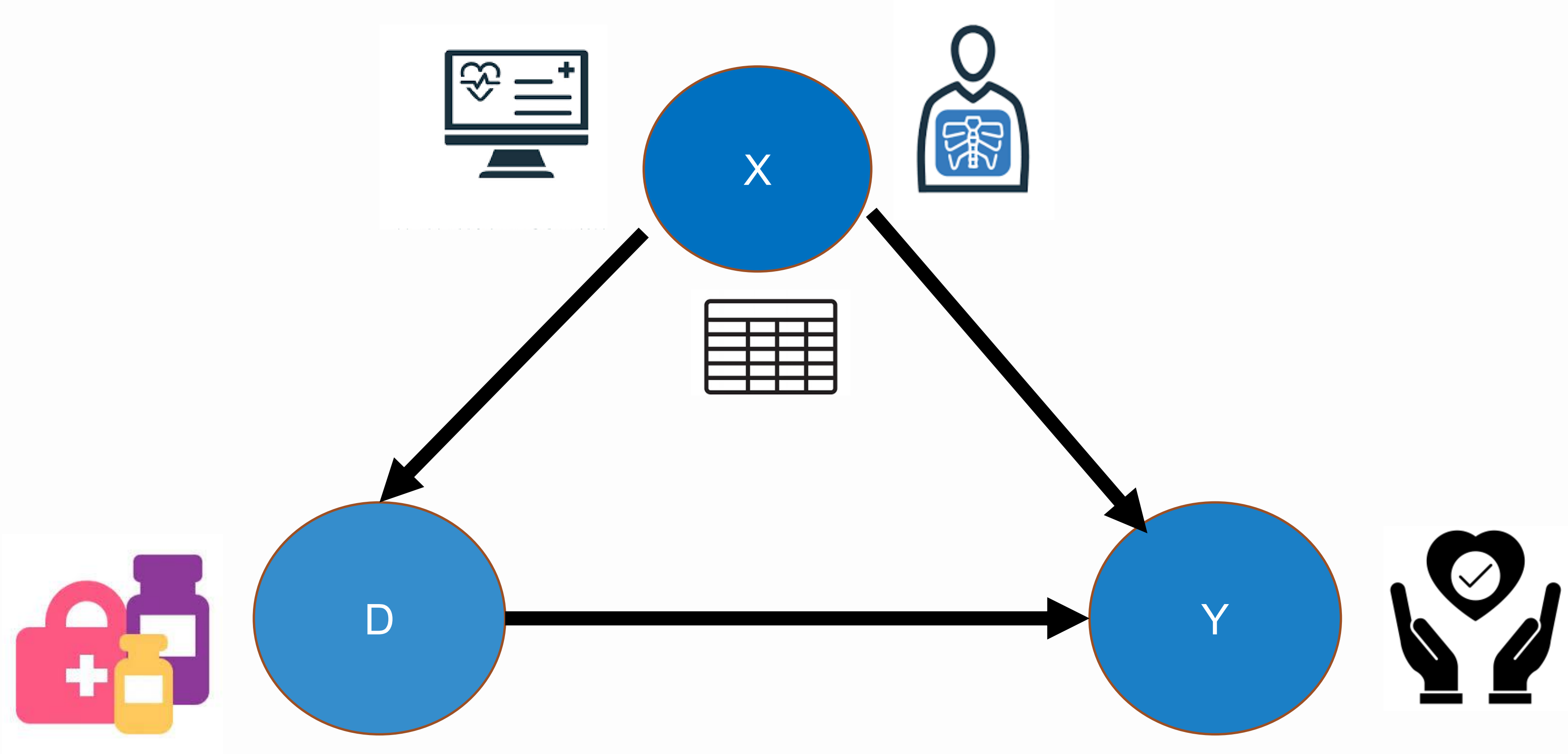
- Treatment effect depending on convariates
- $\theta(X)$
- Key for personalized medicine and targeted policies
- In the last years lot of progress on estimation of HTE with ML /AI
- Double Machine Learning also allows for valid inference
  of complex objects

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DoubleMLDeep

in collaboration with ECONOMIC AI

Causal Infernece with text and images



Causal Infernece with text and images

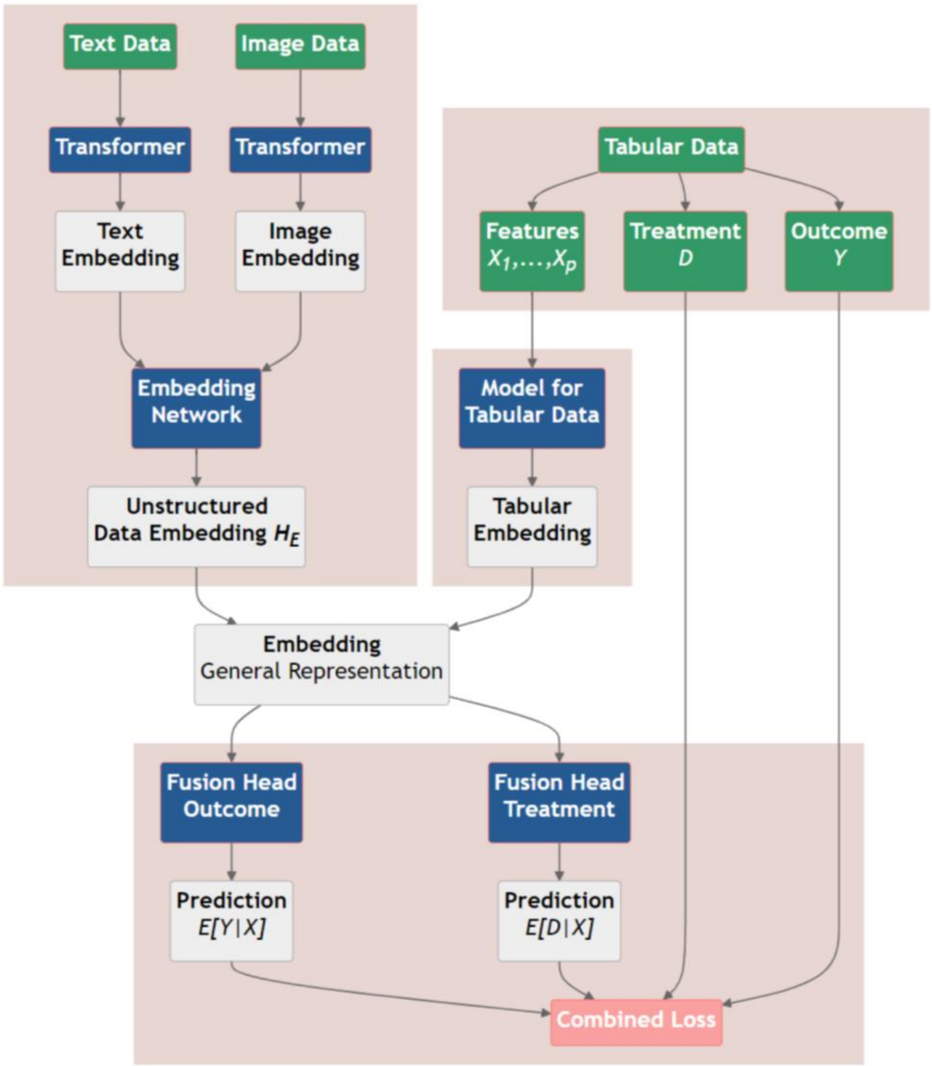
$$Y = D\theta_0 + g_0(X) + \zeta$$

Outcome

Treatment

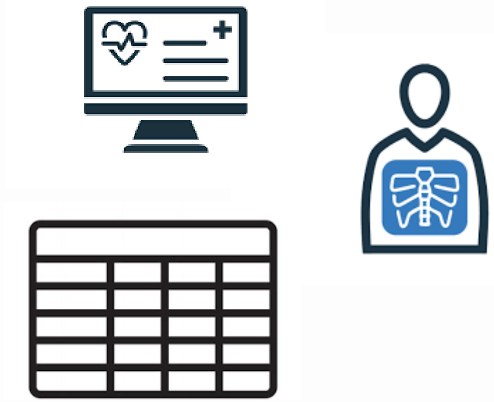
(Multimodal) Confounder
/ Control Variables

Causal Infernece with text and images



Causal Infernece with text and images

$$Y = D\theta_0 + g_0(X) + \zeta$$



Looking for collaboration partner and beta user!

(Multimodal) Confounder / Control Variables

ECONOMIC AI

ECONOMIC AI

arXiv > cs > arXiv:2402.01785

Search...

Help |

Computer Science > Machine Learning

[Submitted on 1 Feb 2024]

DoubleMLDeep: Estimation of Causal Effects with Multimodal Data

Sven Klaassen, Jan Teichert-Kluge, Philipp Bach, Victor Chernozhukov, Martin Spindler, Suhas Vijaykumar

This paper explores the use of unstructured, multimodal data, namely text and images, in causal inference and treatment effect estimation. We propose a neural network architecture that is adapted to the double machine learning (DML) framework, specifically the partially linear model. An additional contribution of our paper is a new method to generate a semi-synthetic dataset which can be used to evaluate the performance of causal effect estimation in the presence of text and images as confounders. The proposed methods and architectures are evaluated on the semi-synthetic dataset and compared to standard approaches, highlighting the potential benefit of using text and images directly in causal studies. Our findings have implications for researchers and practitioners in economics, marketing, finance, medicine and data science in general who are interested in estimating causal quantities using non-traditional data.

Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI); Econometrics (econ.EM); Methodology (stat.ME); Machine Learning (stat.ML)

MSC classes: 62, 91

ACM classes: I.2.0

Cite as: arXiv:2402.01785 **[cs.LG]**

(or arXiv:2402.01785v1 **[cs.LG]** for this version)

https://doi.org/10.48550/arXiv.2402.01785

Submission history

From: Martin Spindler [view email]

[v1] Thu, 1 Feb 2024 21:34:34 UTC (261 KB)

Universität Hamburg

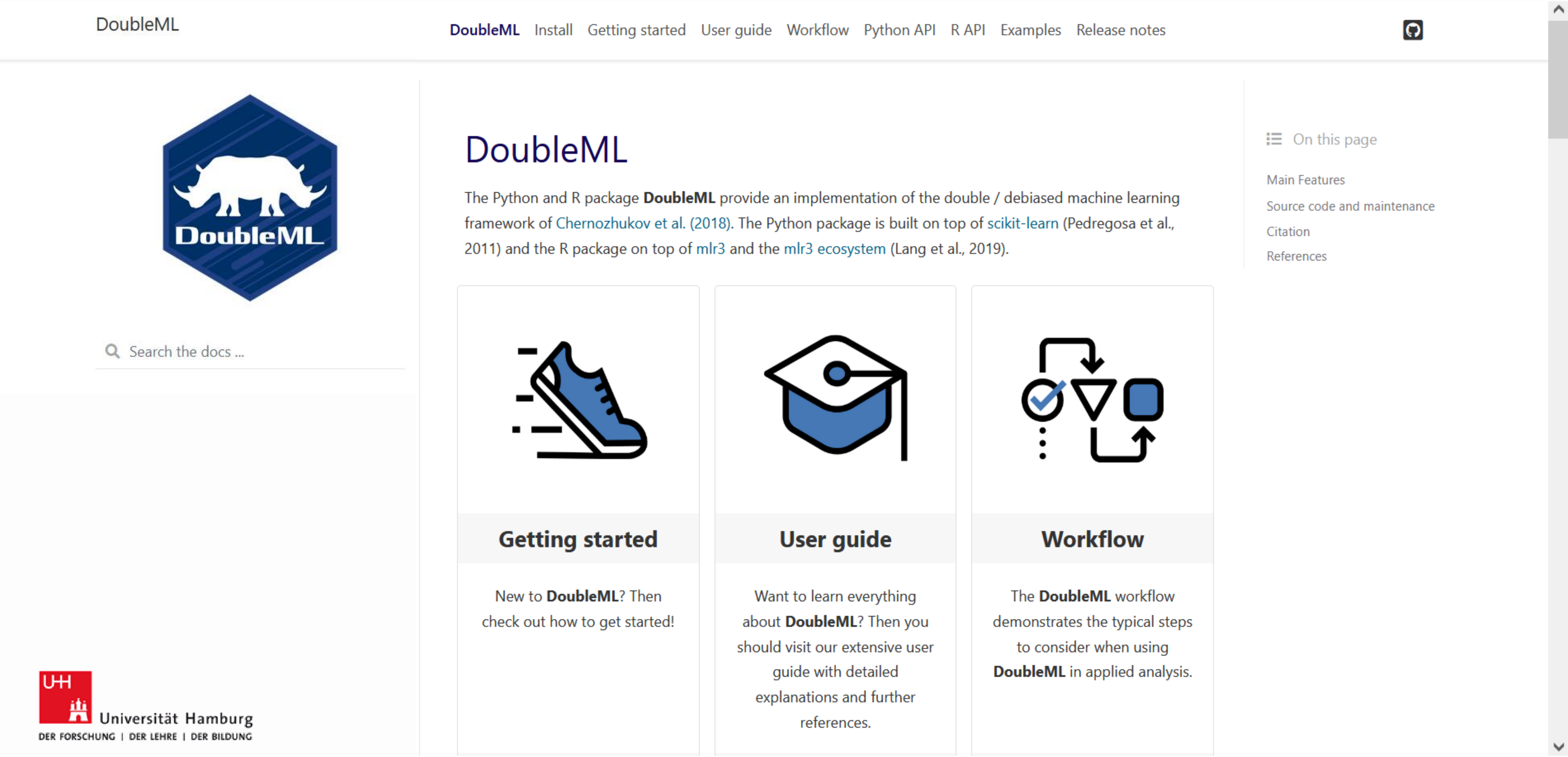
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

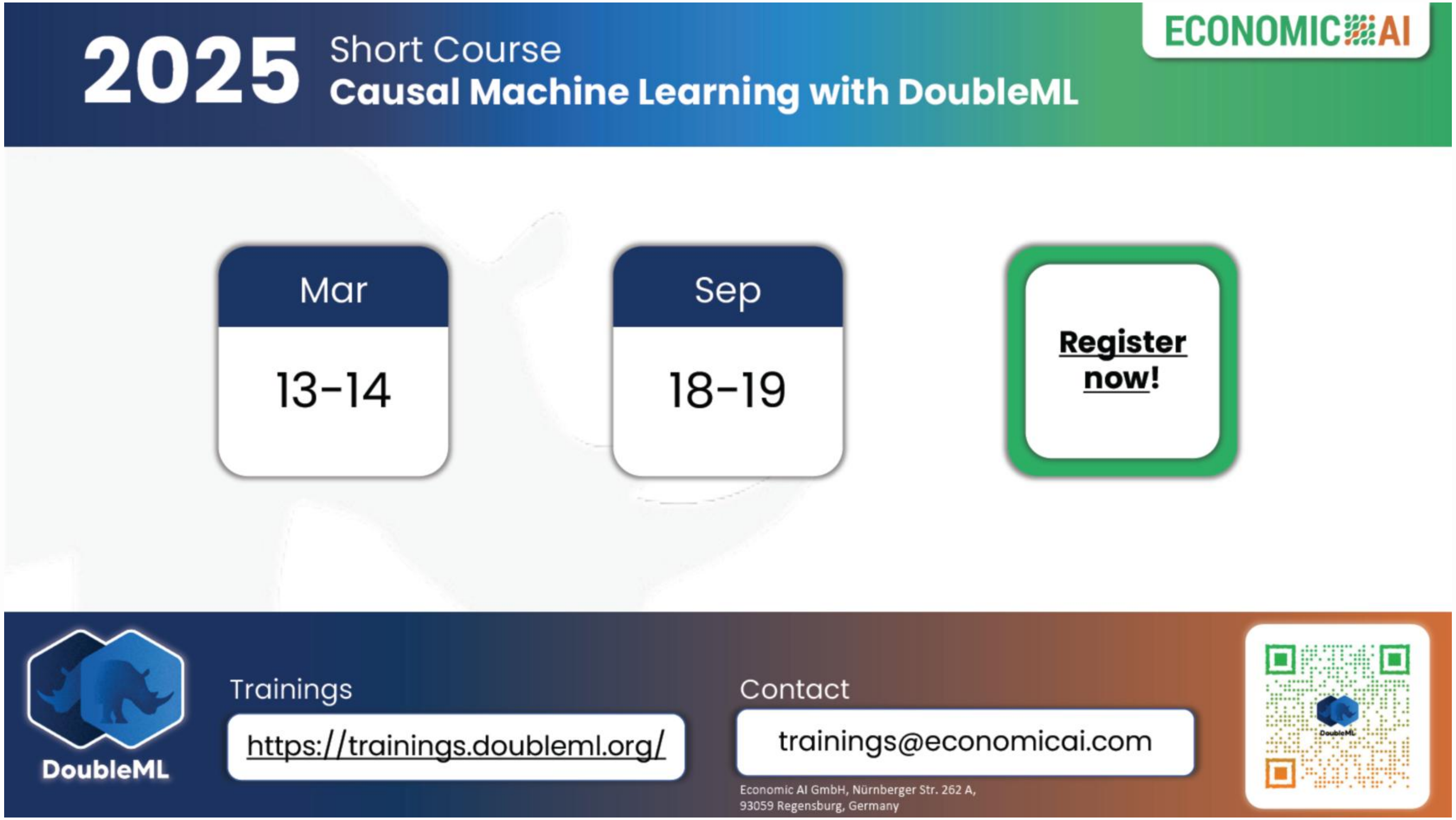
ECONOMIC AI

Resources

Online Resources



Online Resources



ECONOMIC AI
2025
Short Course
Causal Machine Learning with DoubleML
Mar
13-14
Sep
18-19
Register now!
DoubleML
Trainings
https://trainings.doubleml.org/
Contact
trainings@economicai.com
Economic AI GmbH, Nürnberger Str. 262 A,
93059 Regensburg, Germany

Webinar

Webinar: Causal AI for Industry

On July 2nd, 2025, 16:00 – 17:10 CET

The new field of Causal AI combines Causal Inference and Machine Learning to tackle complex, causal problems that conventional, correlation-based Predictive/Generative AI can't solve. **What value does Causal AI offer to industry**? Join our webinar for a hands-on introduction to Causal AI and discover its applications in targeted marketing, dynamic pricing, product optimization, advanced A/B testing, and many more.
Don't miss the opportunity – Register now to stay ahead.

Target Audience

Data Scientists, (Data Science) Managers, Statisticians, Analysts, and anyone interested in gaining a first understanding of Causal AI.

Timetable – July 2nd, 2025

ECONOMIC AI

	Contents	*Speaker*
16:00 – 16:05 CET	*Welcome Note*	Prof. Dr. Martin Spindler
16:05 – 16:20 CET	*What's new in DoubleML?*	Dr. Philipp Bach
16:20 – 16:40 CET	*Causal AI in Corporate Finance: Credit Ratings*	Dr. Helmut Wasserbacher
16:40 – 16:50 CET	*Upskilling for Causal AI*	Dr. Philipp Bach
16:50 – 17:10 CET	*Causal AI Projects & Closing + Q&A*	Jan Rabenseifner Prof. Dr. Martin Spindler

Register now via Eventbrite

Thank You for Your Attention

ECONOMIC AI

Prof. Dr. Martin Spindler

spindler@economicai.com

martin.spindler@uni-hamburg.de

linkedin.com/in/martin-spindler

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

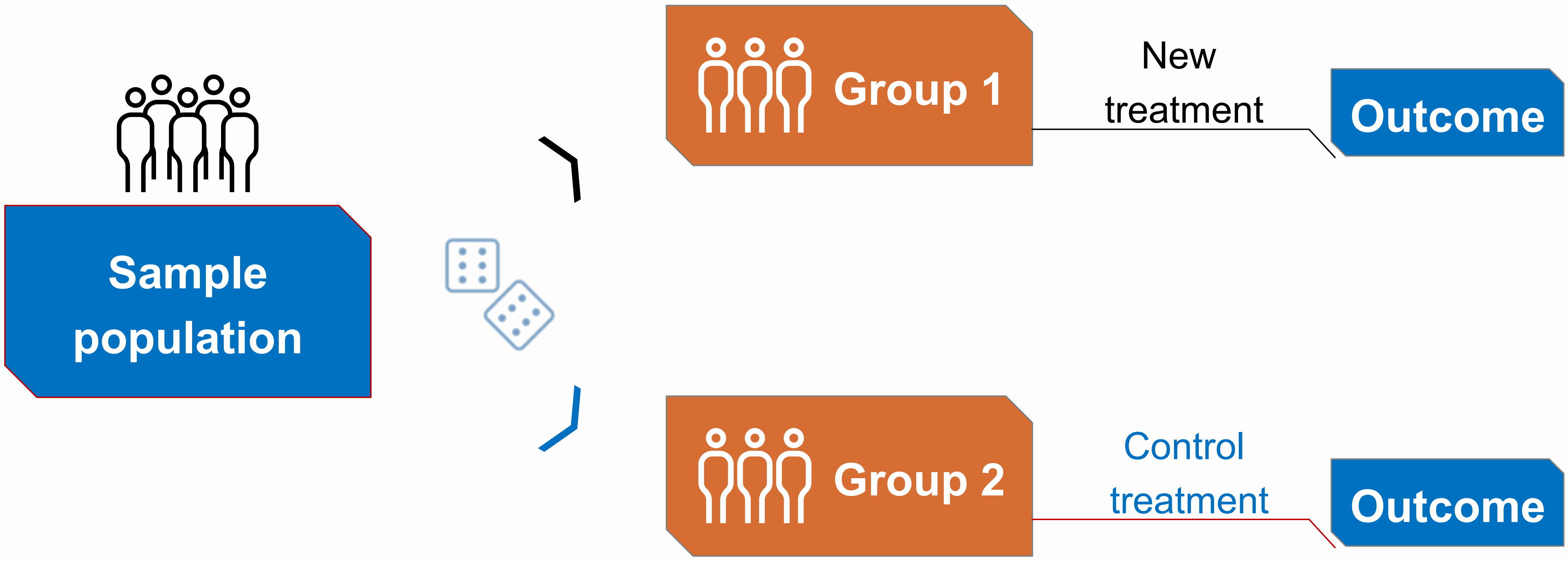
Appendix

Randomized Control Trials

ECONOMIC AI

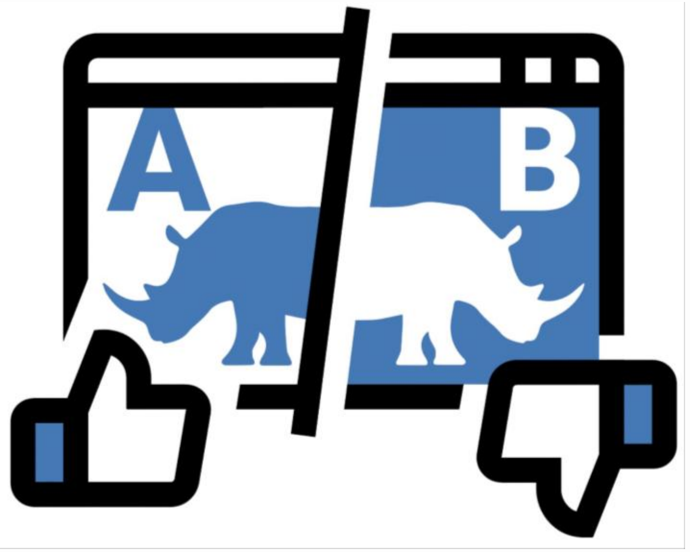
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Randomize Control Trial: Evaluating Causal Effects



Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Application: Randomized Experiments



- **General:** What is the effect of a certain variable $\boldsymbol{D}$ on a relevant outcome variable $\boldsymbol{Y}$?
- **Randomized experiments** are a direct way to estimate such effects (assuming they are conducted properly)

Challenges in practice:

1. No (pure) A/B-testing / experiments possible → observational data
2. A/B test suffers from low power
3. Heterogenous treatment effects

Solution with DoubleML

1. **Observational study**: Include control variables $\boldsymbol{X}$ which may also impact the variables $\boldsymbol{Y}$ or $\boldsymbol{D}$
2. Include covariates $\boldsymbol{X}$ that help to predict the outcome $\boldsymbol{Y}$ using ML methods
3. Detection of complex treatment effect patterns

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

RCTs powered by AI

More precise estimation with ML & AI

Heterogenous treatment effects & policy optimization

- „Personalized medicine“

Adaptive Experiments & Reinforcement Learning

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Observational Data / Real World Data

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Real World Evidence

- Additional assumptions are required for causal inference
- Important innovation: Heterogenous Treatment Effects
- Combining CI tools with ML for complex settings
- Multimodal data becoming more and more available (e.g. EHR, medical images)

Real World Evidence

Toolbox for causal inderence:

- Unconfoundedness assumption
  (propensity score matching, regression adjustment, doubly robust estimation, …)
- More realistic with more data (e.g. EHR)
- Instrumental Variables
- Difference in Difference Estimation
- Synthetic Controls
- Regression Discontinuity
- Panel Data

The Key Ingredients of DML

Illustration of the cross-fitting algorithm