EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Causal Machine Learning

Welcome

Michael Knaus

March 10, 2023

Welcome!

Welcome to the Causal Machine Learning course

What is Causal Machine Learning?

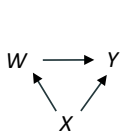
I honestly have no definition and different fields think differently about it

In this course you will learn an econometric perspective

We will cover recent advances in the effort to combine the powerful toolbox developed in Machine Learning (one mature literature) for conducting Causal Inference (another mature literature)

$\Rightarrow$ Creates a new strand of literature: Causal Machine Learning

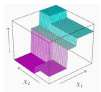
Combining two mature literatures



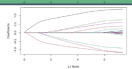
W → Y

Causal Inference
(focus on identification)

Z → W → Y



Supervised ML
(focus on prediction)



Causal ML
(focus on estimation)

Plan for today

1. What this course is about

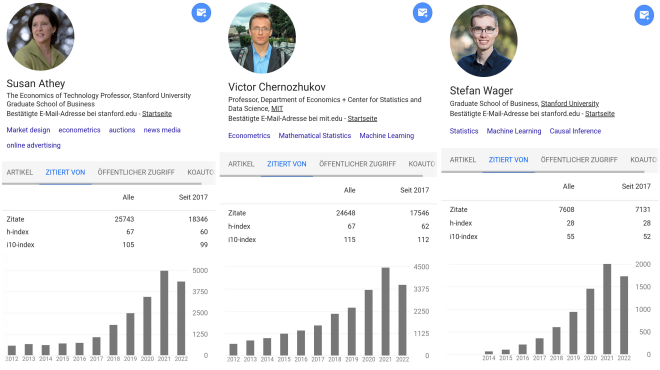
2. Method journey of the semester

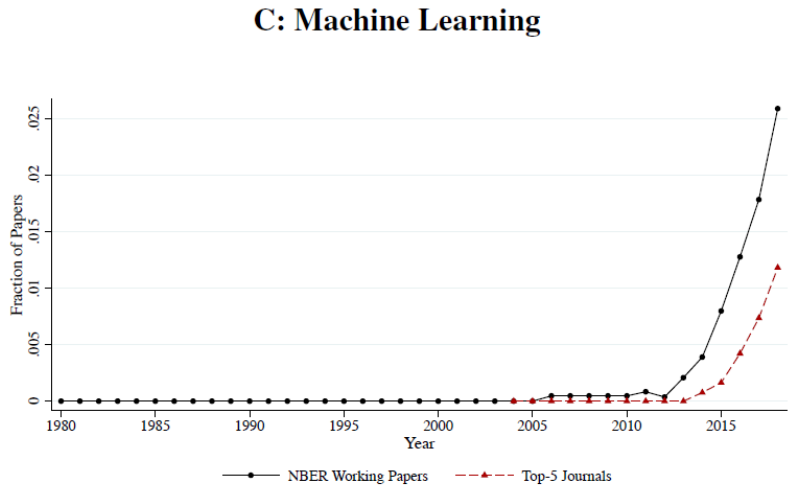
What this course is about

Machine Learning + Econometrics

The integration of ML into econometrics is arguably one of two methodological megatrends relevant for empirical economists (next to the better understanding of ”Dif-in-Difs”, see e.g. de Chaisemartin & D’Haultfœuille, 2022; Roth et al., 2022)

Three faces are representative for this development (accessed Oct 2022):





Source: Currie, Kleven, Zwiers (2020)

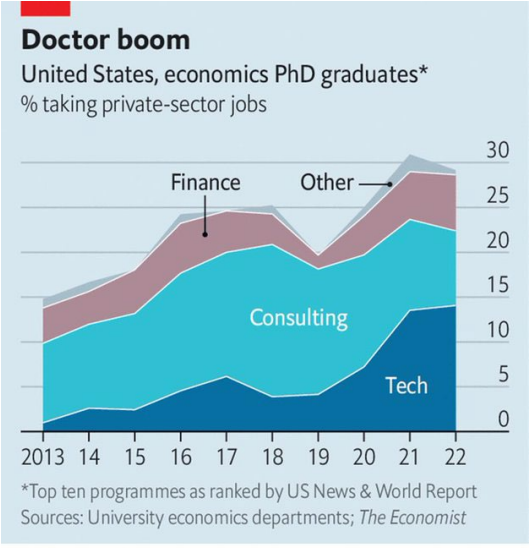
Machine Learning + Economics + Industry (1/3)

Big Tech firms, especially Amazon, but also Google, Meta, Uber, ... started to hire Economists on a large scale

Why?

"[...] economists offer skills that computer scientists and engineers often lack. They tend to have a good grasp of statistics, as well as a knack for understanding how incentives affect human behaviour. Most important, economists are adept at designing experiments to identify causal relationships between variables. ***Machine-learning engineers usually think in terms of prediction problems [...]. Economists can nail down the causal parameters [...].*** *"* The Economist Sep 7th 2022

Machine Learning + Economics + Industry (2/3)



The Economist

Machine Learning + Economics + Industry (3/3)

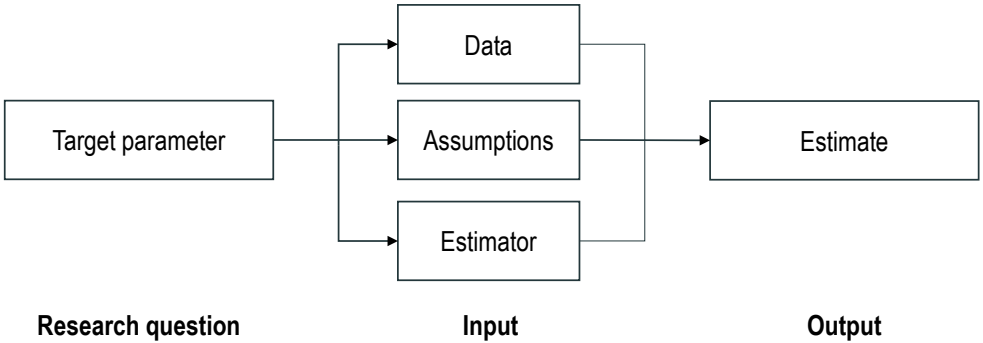
Good for you: The skills you acquire in this course and your studies in general are highly valuable

On a critical note:

Very stylized workflow of empirical economics

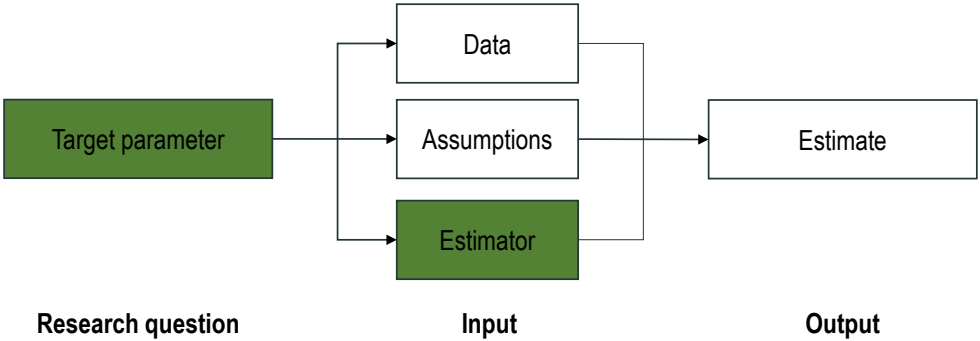
The big question is how to come from an (often causal) target parameter to a credible estimate of this parameter

Stylized workflow of empirical economics



Machine Learning/Data Science are currently used to improve all three inputs in different dimensions

This course



We focus on extending the menu of target parameters and estimators in your empirical toolbox

What to expect? (1/2)

You will learn recent Causal ML methods for estimation of average and personalized effects of

- an intervention/policy/treatment W
- on an outcome Y
- while adjusting for confounding variables X (optional)
- using exogeneous variation of an instrument Z (optional)

and how to recommend assignment of W in a data-driven way

To this end, we first need to recap/learn the Causal Inference and ML basics

Important: The methods we cover do NOT provide more credible identification BUT more clever/data-driven and comprehensive estimation

What to expect? (2/2)

The literature is moving very fast and there are no established textbooks that could serve as basis for such a course, yet

There are open questions regarding basically everything that I show you

$\Rightarrow$ No best practices that I could teach, yet

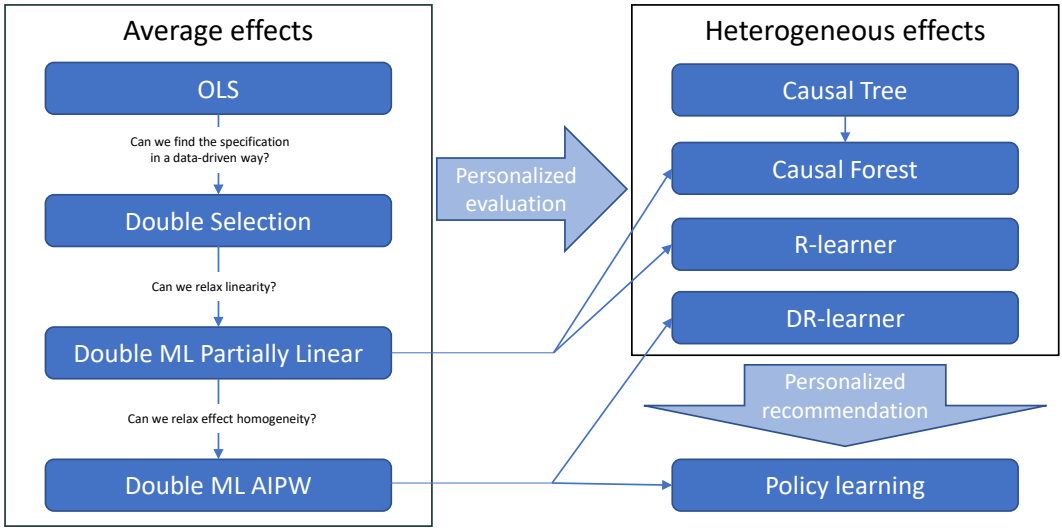
My goal is to introduce key ideas underlying popular Causal ML methods and to connect dots between them and concepts from econometrics 101

This enables you to apply the basic methods we discuss and allows you to faster understand more complex methods building on the same ideas

However, do not expect that you will only learn settled recipes that just require pushing buttons as you are used to with more mature methods

Method journey of the semester

Journey


Average effects
OLS
Can we find the specification in a data-driven way?
Double Selection
Can we relax linearity?
Double ML Partially Linear
Can we relax effect homogeneity?
Double ML AIPW
Personalized evaluation
Heterogeneous effects
Causal Tree
Causal Forest
R-learner
DR-learner
Personalized recommendation
Policy learning

Example use case of the full pipeline

Evaluation of training programmes for job-seekers

Some shameless self-promotion

The Econometrics Journal

Econometrics Journal (2022), volume **25**, pp. 602–627.
https://doi.org/10.1093/ectj/utac015

Double machine learning-based programme evaluation under unconfoundedness

MICHAEL C. KNAUS

Swiss Institute for Empirical Economic Research, University of St. Gallen, Varnbüelstrasse 14, 9000 St. Gallen, Switzerland.
Email: michael.knaus@unisg.ch

PDF available from Journal or arXiv

Ceterum censeo a fancy method alone is not a credible identification strategy