Comprehensive Causal Machine Learning

Michael Lechner
Jana Mareckova

COMPIE 2024

June 24, 2024

Motivation

- Estimation of treatment effects at different levels of granularity (population, groups, individuals) within one estimation framework under unconfoudedness

Parameters of interest:

- Average Treatment Effect (ATE): population
- Conditional Average Treatment Effects (CATEs):
  - Group Average Treatment Effect (GATE): larger group of interest (e.g. age groups)
  - Individualized Average Treatment Effect (IATE): unit with specific values of many characteristics

Comprehensive Estimation

- Estimation of effects at different levels of granularity in a binary or multiple treatment setting
- Statistical inference
- Prediction of treatment effects on new sample that contains covariates only

Estimation Methods

- Causal Machine Learning Methods offer these options:
  - **Double Machine Learning (DML)**, Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018)
  - **Generalized Regression Forest (GRF)**, Athey, Tibshirani, and Wager (2019)
  - **Modified Causal Forest (MCF)**, Lechner (2018)

→ **Comprehensive Causal Machine Learning**

DML

- ATE, GATE and IATE estimated via a regression of a doubly robust score on a constant, group indicators or covariates, respectively

- DML is asymptotically semiparametrically efficient for ATE
- Consistent estimation of GATEs with uniform convergence (Semenova & Chernozhukov, 2021)
- Upper bound on the IATE error (Kennedy, 2023)

GRF

- IATEs estimated as a solution of local moment condition using forest weights
- GATE and ATE estimated as in DML

- IATE estimator is consistent and asymptotically normal with a known convergence rate

MCF

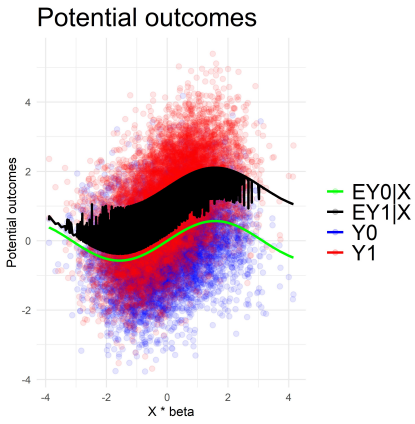
- Extension of CF (Wager and Athey (2018), *JASA*) using a splitting rule accounting for heterogeneity and selection into treatment

- IATE estimated in leaves as a difference between average outcomes of two treatment groups, consistent and asymptotically normal
- GATE and ATE are corresponding aggregates of IATEs, consistent and asymptotically normal
- MCF can reach smaller variance due its smoothing character for GATEs with large number of groups

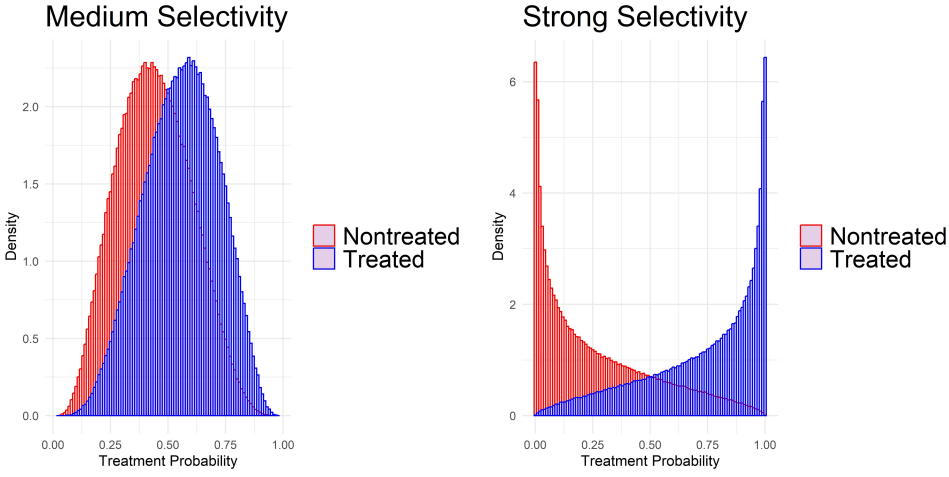
Simulation

- 75 different DGPs
- Treatment heterogeneity: Zero, Linear, Nonlinear, Quadratic, **Step-Function**
- Selection into Treatment: None, Medium, Strong
- Number of total and active confounders, Distribution of confounders, R^2 of potential outcome equations, Sample size N, Number of treatments, Treatment shares for 2 treatments, Number of GATEs
- RMSE, bias, variance, coverage probability

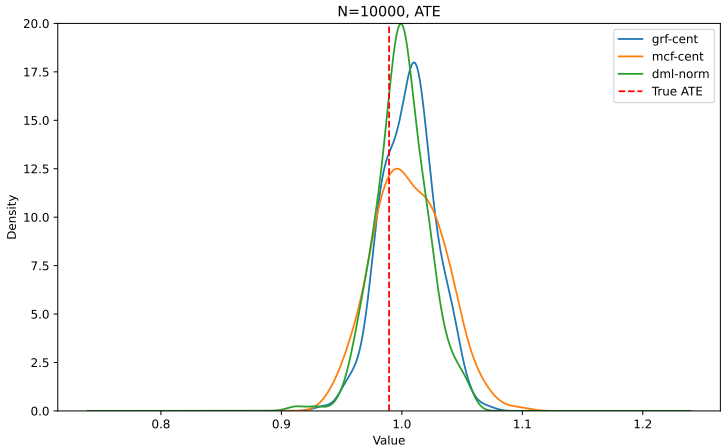
Treatment heterogeneity - Step-function:

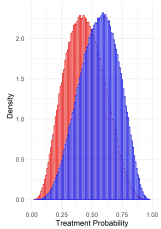


Selection into treatment:

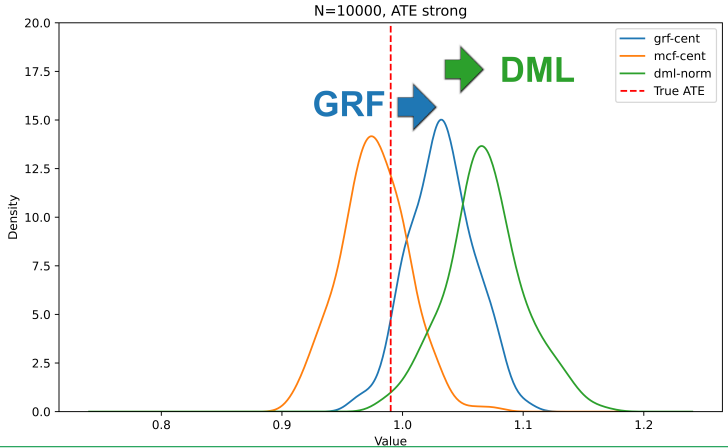


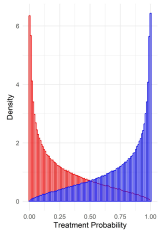
Results - ATE, medium selectivity



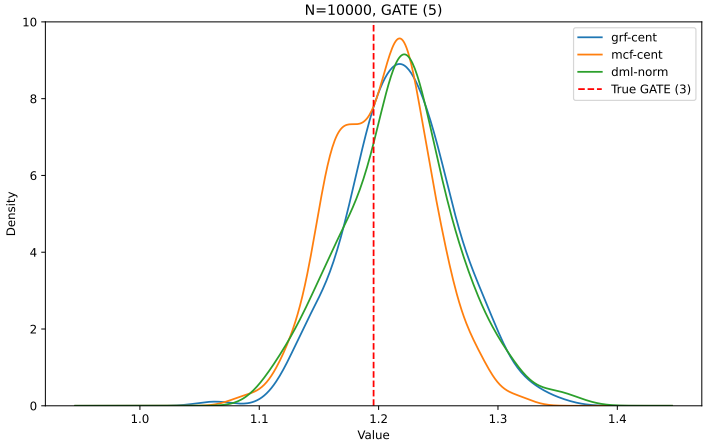


Results - ATE, strong selectivity

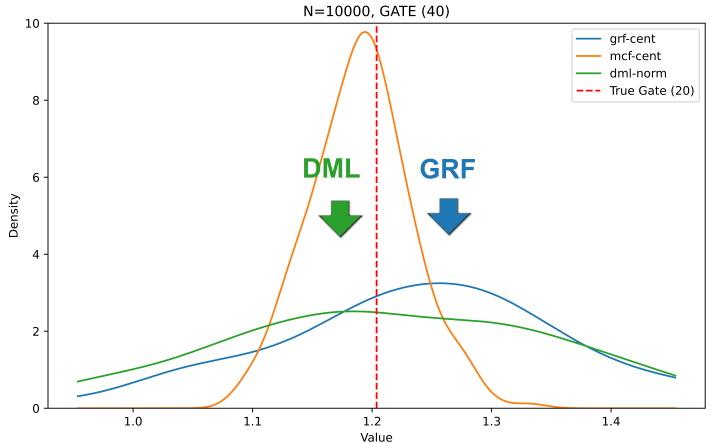




Results - 5 GATEs



Results - 40 GATEs



Results - IATE

- GRF and MCF outperform DML in terms of RMSE - only exception: experimental setting with not too much nonlinearity in treatment effects
- coverage is below nominal levels for all methods stemming mainly from too much bias in point estimates

Practical Recommendations

- No single framework dominating at all levels of granularity
- **Efficiency not crucial** (e.g. N large): Use MCF (most robust).
- **Efficiency is crucial** (e.g. N small): Use parameter-specific approach.
  - For ATE, GATE (few groups): Use DML or GRF. (MCF for robustness check)
  - For GATE (many groups): Use MCF.
  - For IATE: Use GRF or MCF.

Questions? Comments? Suggestions?

Paper on arXiv:



References I

- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). **Double/debiased machine learning for treatment and structural parameters.** The Econometrics Journal 21(1), 1-68.
- Athey, S., Tibshirani, J., & Wager, S. (2019). **Generalized random forests.** The Annals of Statistics, 47(2), 1148.
- Lechner, M. (2018). **Modified causal forests for estimating heterogeneous causal effects.** arXiv preprint arXiv:1812.09487.
- Semenova, V., & Chernozhukov, V. (2021). **Debiased machine learning of conditional average treatment effects and other causal functions.** The Econometrics Journal, 24(2), 264-289.

References II

- Kennedy, E. H. (2023). **Towards optimal doubly robust estimation of heterogeneous causal effects.** Electronic Journal of Statistics, 17 (2), 3008-3049.
- Wager, S., & Athey, S. (2018). **Estimation and inference of heterogeneous treatment effects using random forests.** Journal of the American Statistical Association, 113(523), 1228-1242.