

Comprehensive Causal Machine Learning

Michael Lechner, Jana Mareckova
University of St. Gallen
michael.lechner@unisg.ch, jana.mareckova@unisg.ch

1. Motivation

Goal:

- Compare Causal Machine Learning (CML) methods that comprehensively estimate treatment effects from theoretical and simulation-based perspectives under unconfoundedness

Comprehensive estimation:

1. Unified framework for estimation of effects using ML methods at all possible levels of granularity
2. Statistical Inference
3. Prediction of effects at higher levels of granularity

Contribution:

1. Results on relative performance of DML, GRF, MCF in a large-scale simulation study
2. Theoretical guarantees for MCF
3. Practical recommendations

2. Parameters

Average Treatment Effect (ATE):

- Average treatment effect for a population of interest

Group Average Treatment Effect (GATE):

- Average treatment effect for a larger group of interest (e.g. age groups)

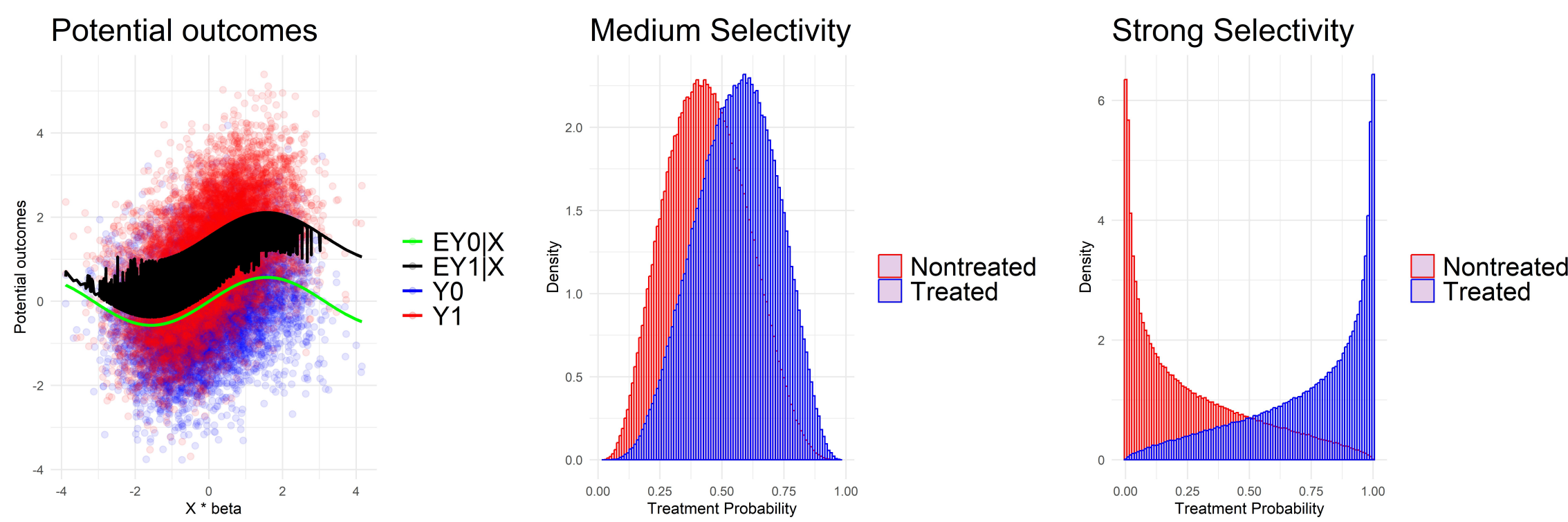
Individualized Average Treatment Effect (IATE):

- Average treatment effect for a unit with specific values of many characteristics

4. Simulation Study

Large-scale simulation study varying:

- **Treatment heterogeneity:** Zero, Linear, Nonlinear, Quadratic, **Step-Function**
- **Selection into Treatment:** None, Medium, Strong
- Number of total and active confounders
- **Type of confounders:** Uniform, Normal, **Mixed**
- R^2 of non-treated outcome equation: 0, **0.1**, 0.45
- **Sample size, N :** 2 500, 10 000
- **Number of Treatments:** 2, 4
- **Treatment shares for 2 treatments:** 25%, 50%
- **Number of GATEs:** 5, 10, 20, 40



Focus: Quality of point estimates and inference procedures

6. Summary

DML: Best for ATE, GATEs with small number groups and when selectivity is not too strong

GRF: Best for IATEs, is less affected by selectivity than DML for ATE and GATE

MCF is robust to strong selectivity, yields low MSE for GATEs with many groups

3. Estimators

DML: Double Machine Learning (Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) *The Econometrics Journal*)

- 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

GRF: Generalized Random Forest (Athey, Tibshirani, and Wager (2019), *The Annals of Statistics*)

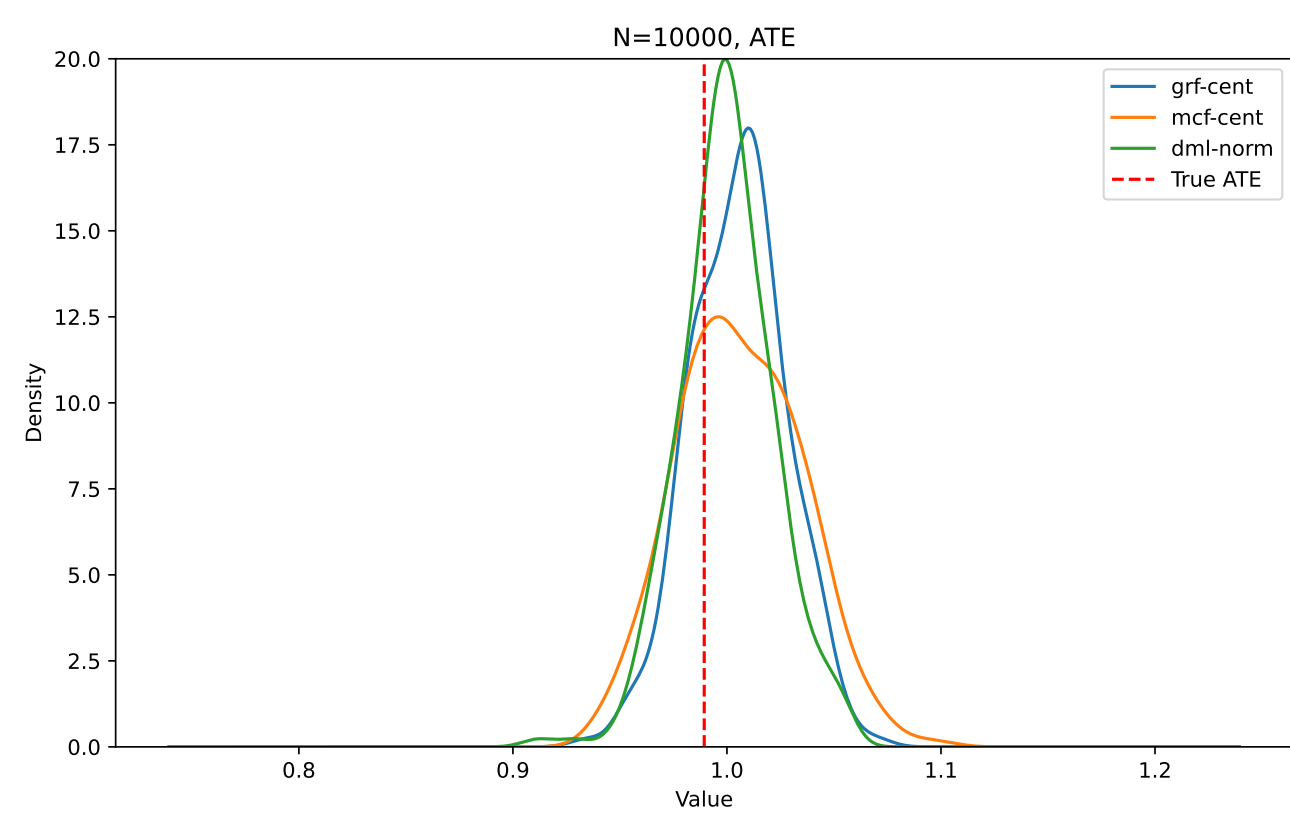
- IATEs estimated as a solution of local moment condition using forest weights; GATE and ATE estimated as in DML
- IATE estimator is consistent and asymptotically normal with a known convergence rate

MCF: Modified Causal Forest (Lechner (2018), *arXiv*)

- Extension of CF (Wager and Athey (2018), *JASA*) using a splitting rule accounting for heterogeneity & selection
- IATE estimated in leaves as difference between average outcomes of two treatment groups. GATE and ATE are corresponding aggregates of IATEs.
- MCF can reach smaller variance due its smoothing character for GATEs with large number of groups

5. Results

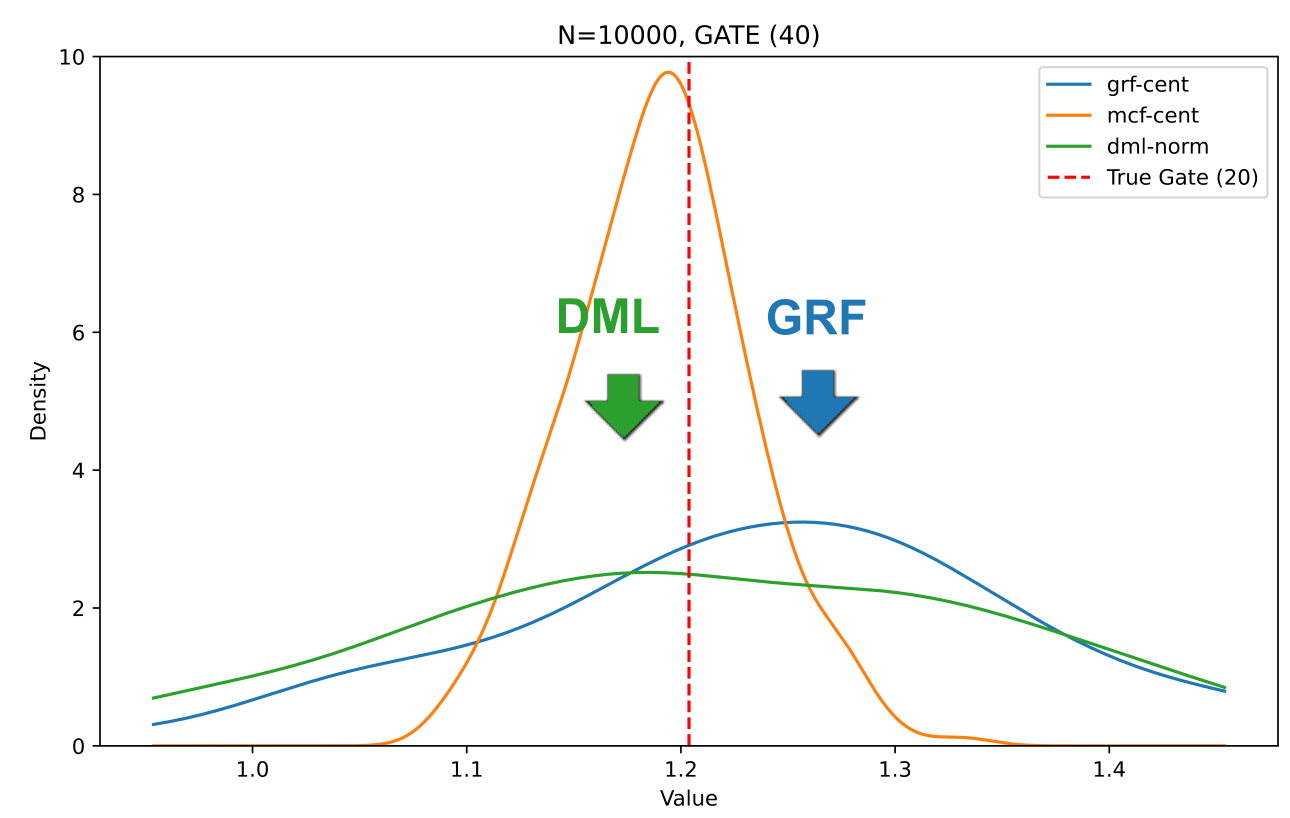
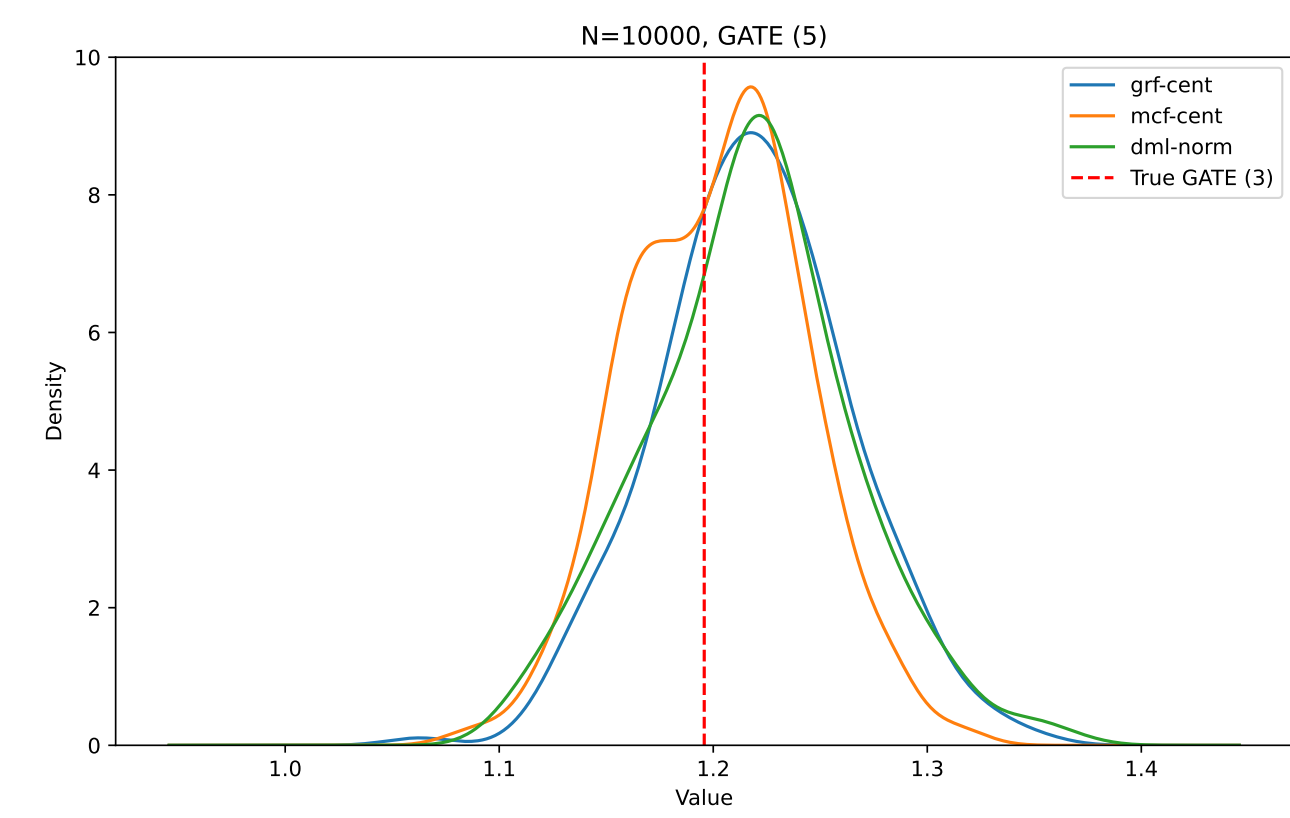
ATE: medium $\rightarrow$ strong selectivity



IATE

- GRF and MCF outperform DML - only exception: no selectivity into treatment 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

GATE: 5 $\rightarrow$ 40 groups



7. Practical Recommendations

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.