Causal inference in industry and academia

Jana Mareckova – University of St. Gallen

WiDS 2024, Zürich

Short Introduction to Causal Inference

- **Causal inference** identifies true causal effects.
- **Causal Effect:** What would have happened to a population, group or individual under a different "treatment" (e.g. educational choice)?

Women in Data Science Worldwide | Zurich

Example in Academia / Public Policy

Unemployment Program Effectiveness

- **Scenario:** A government implements a new program to reduce unemployment duration.
- **Question:** What is the causal effect of the program on unemployment duration?
- **Outcome:** Assess whether the program effectively reduces the duration of unemployment.

Example in Industry

Marketing and Sales

- **Scenario:** A company launches a new marketing campaign.
- **Question:** What is the causal effect of the marketing campaign on sales?
- **Outcome:** Determine the true impact of the campaign on sales figures.

Estimation Methods

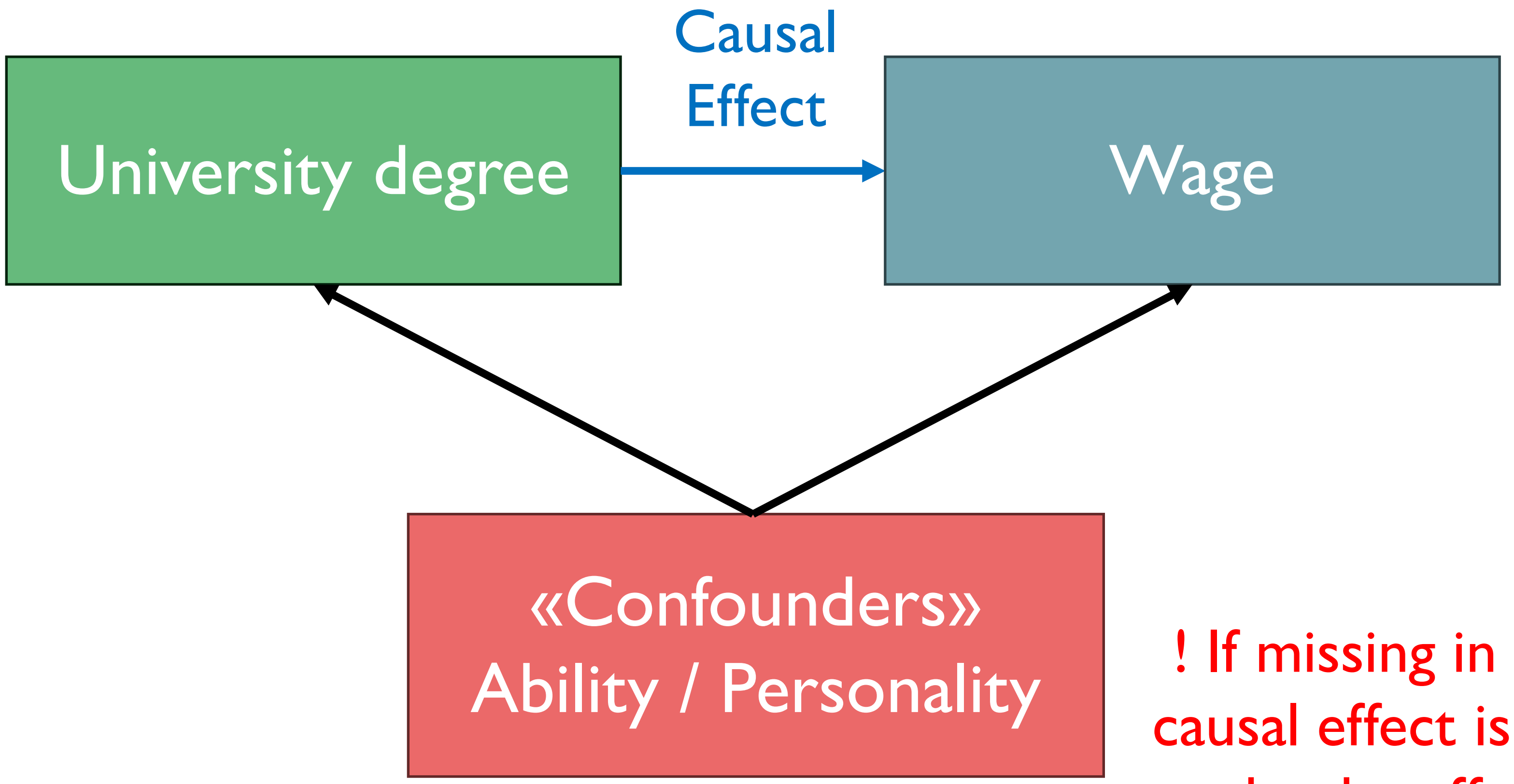
Lechner and Mareckova (2024) „**Comprehensive Causal Machine Learning**“, arXiv Working Paper

- Comparing methods estimating **causal effects at population, group and individual level** (under selection-on-observables)

  - **Double Machine Learning** (Chernozhukov et al., 2018)
  - **Generalized Random Forest** (Athey et al., 2019)
  - **Modified Causal Forest** (Lechner, 2018)

Selection-on-observables – Main assumption:

Unconfoundedness:



! If missing in the model →
causal effect is contaminated
by the effect of these

Estimation Methods - DML

DML:

- Population, group and individual effects estimated in two steps:
  1. Estimate doubly robust scores
  2. Regress them on constant, group indicators or individual covariates
- DML is asymptotically semiparametrically efficient for population effects
- DML has valid statistical inference on group effects
- DML error on individual effects is bounded

Estimation Methods - GRF

GRF:

- Individual effects estimated as a solution of local moment condition using forest weights
- Forest accounts for effect heterogeneity in splitting rule and selection into treatment by data transformation
- Group and population effects estimated as in DML
- Individual effects are pointwise consistent and asymptotically normal with a known convergence rate

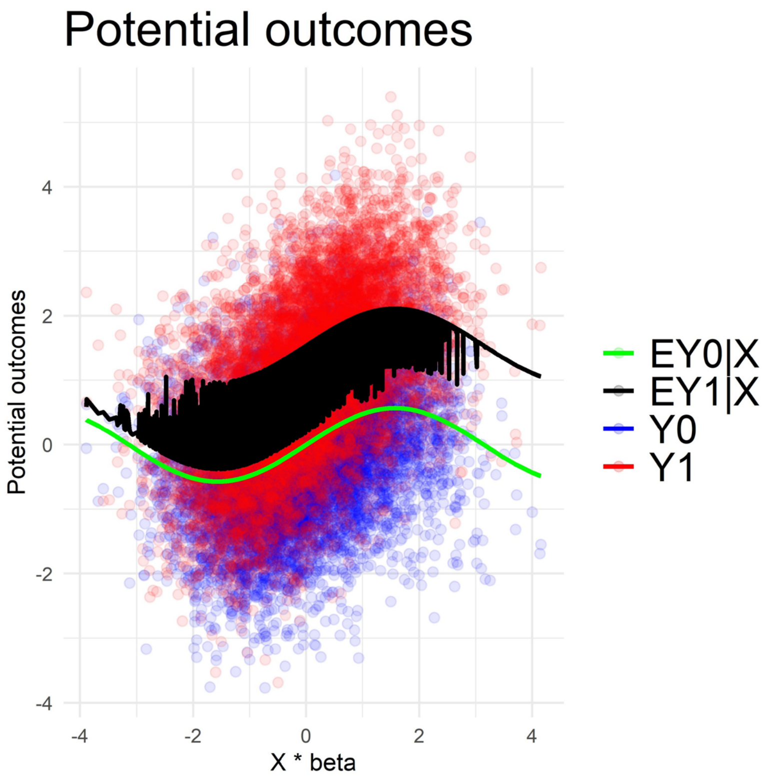
Estimation Methods - MCF

MCF:

- Splitting rule accounts for effect heterogeneity and selection into treatment
- Individual effects estimated in final leaves as difference between outcomes in two treatment groups.
- Group and population effects are corresponding aggregates of individual effects.
- Individual, group and population effects are consistent and asymptotically normal
- MCF can reach smaller variance due its smoothing character for group effects with large number of groups

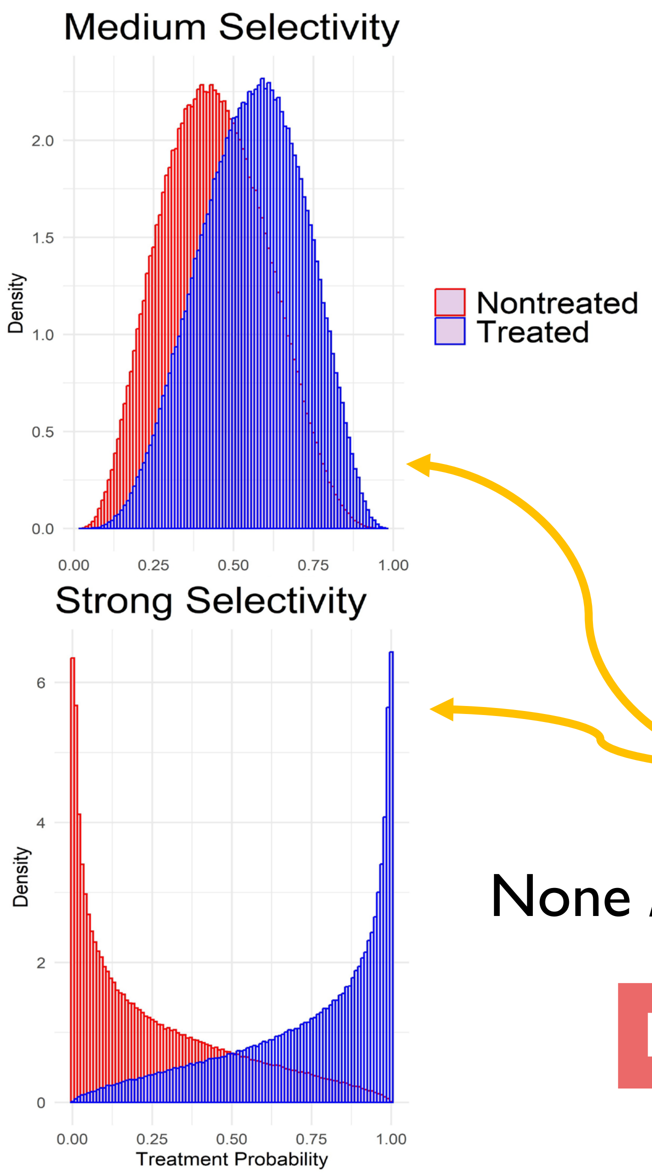
Simulation Designs

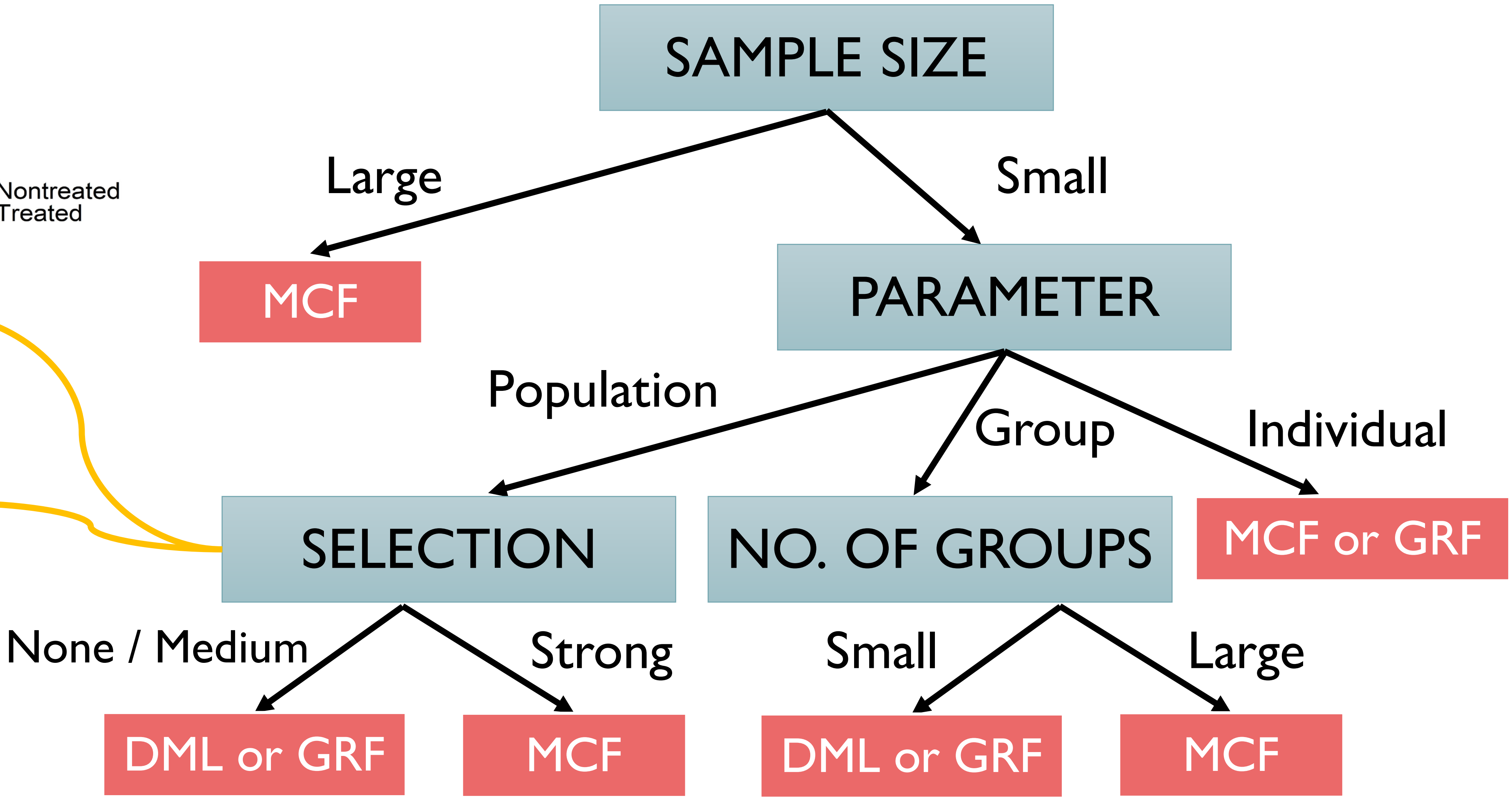
- Treatment heterogeneity: Zero, Linear, Nonlinear, Quadratic, **Step-Function**
- Selection into Treatment: None, Medium, Strong
- Number of total and active variables
- Type of variables
- R^2 of non-treated outcome equation
- R^2 of treatment effect
- Sample size: 2 500, 10 000
- Number of Treatments: 2, 4
- Treatment shares for 2 treatments: 25%, 50%
- Number of groups: 5, 10, 20, 40



Women in Data Science Worldwide | Zurich

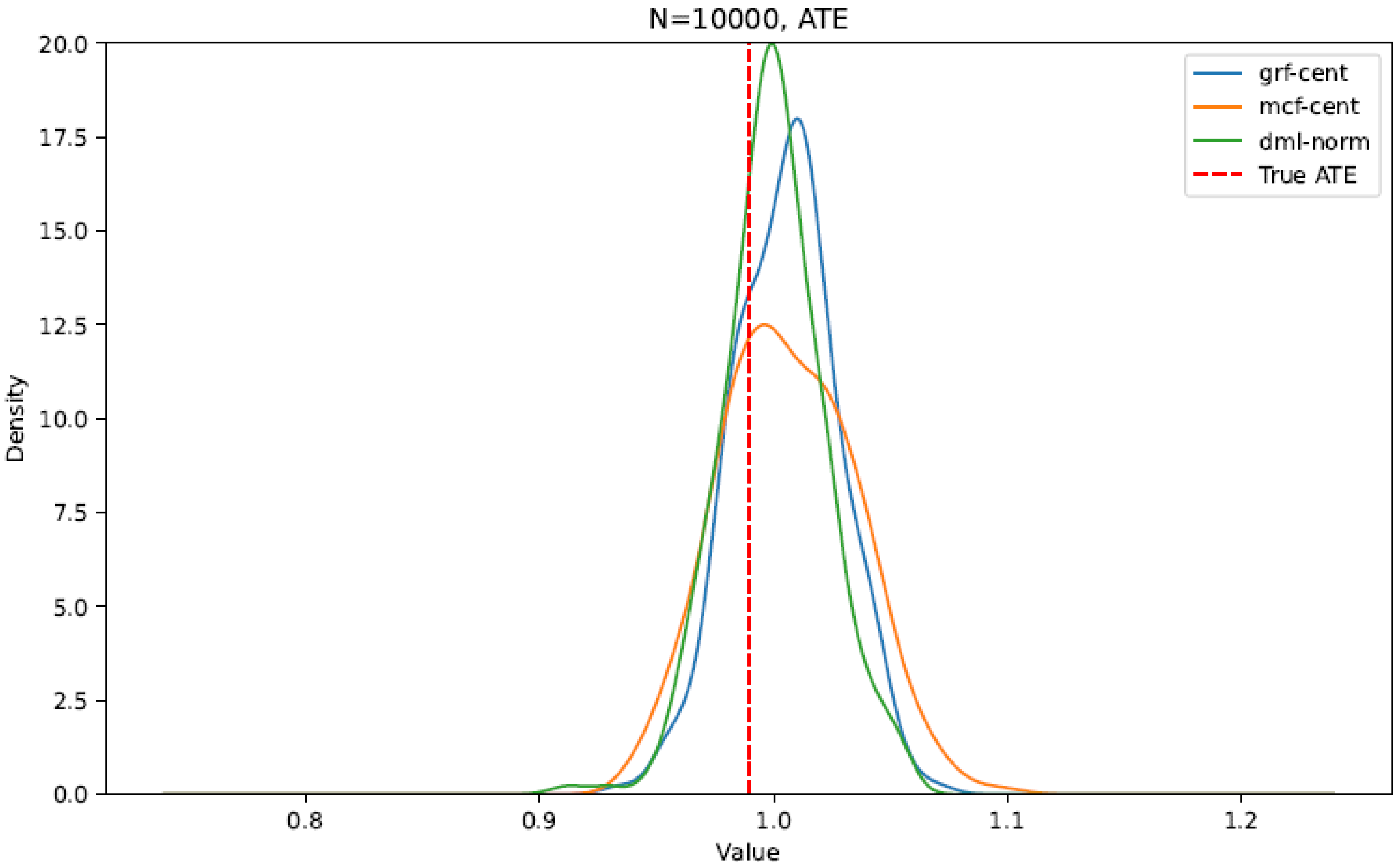
Practical Recommendations



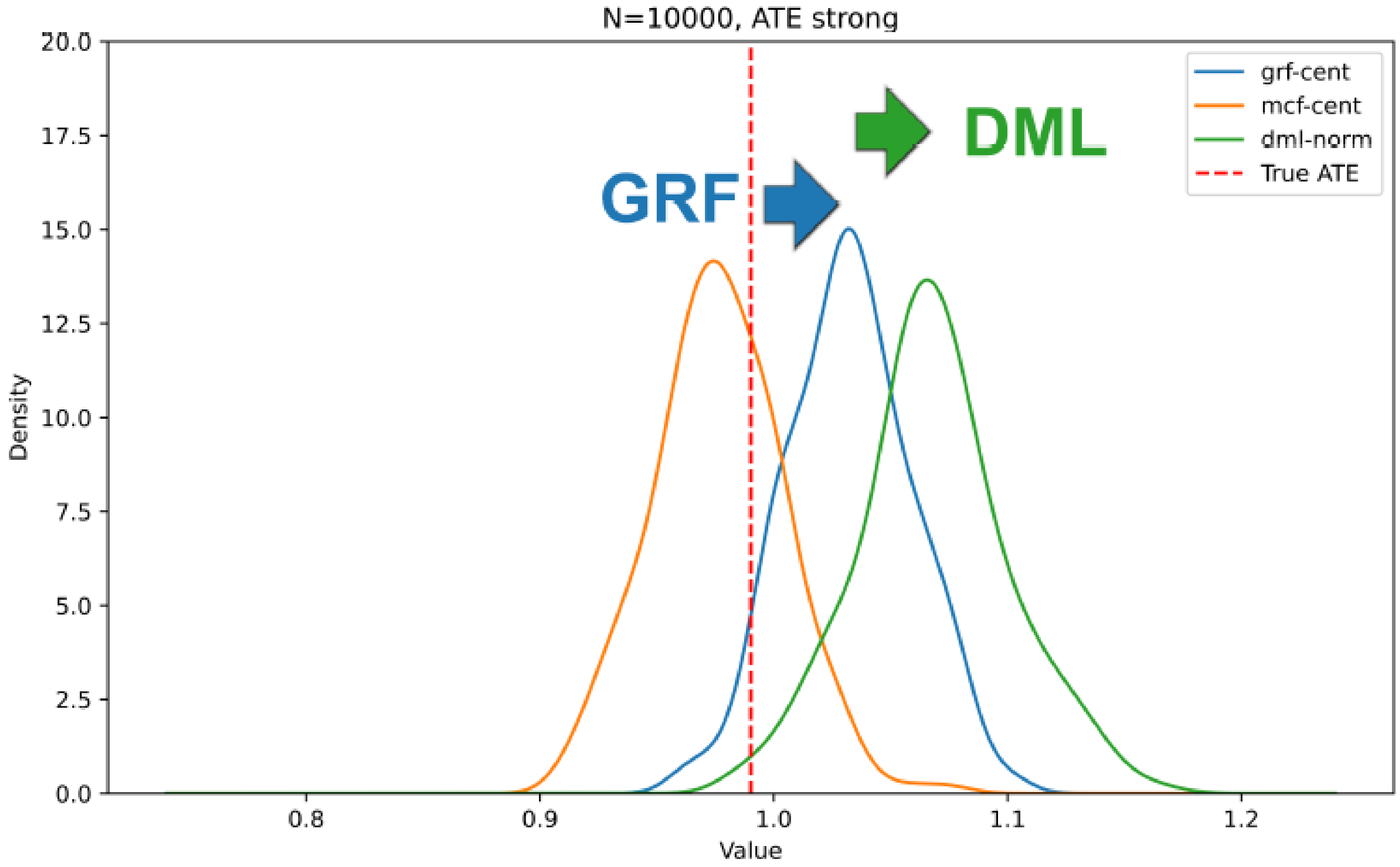



Q&A

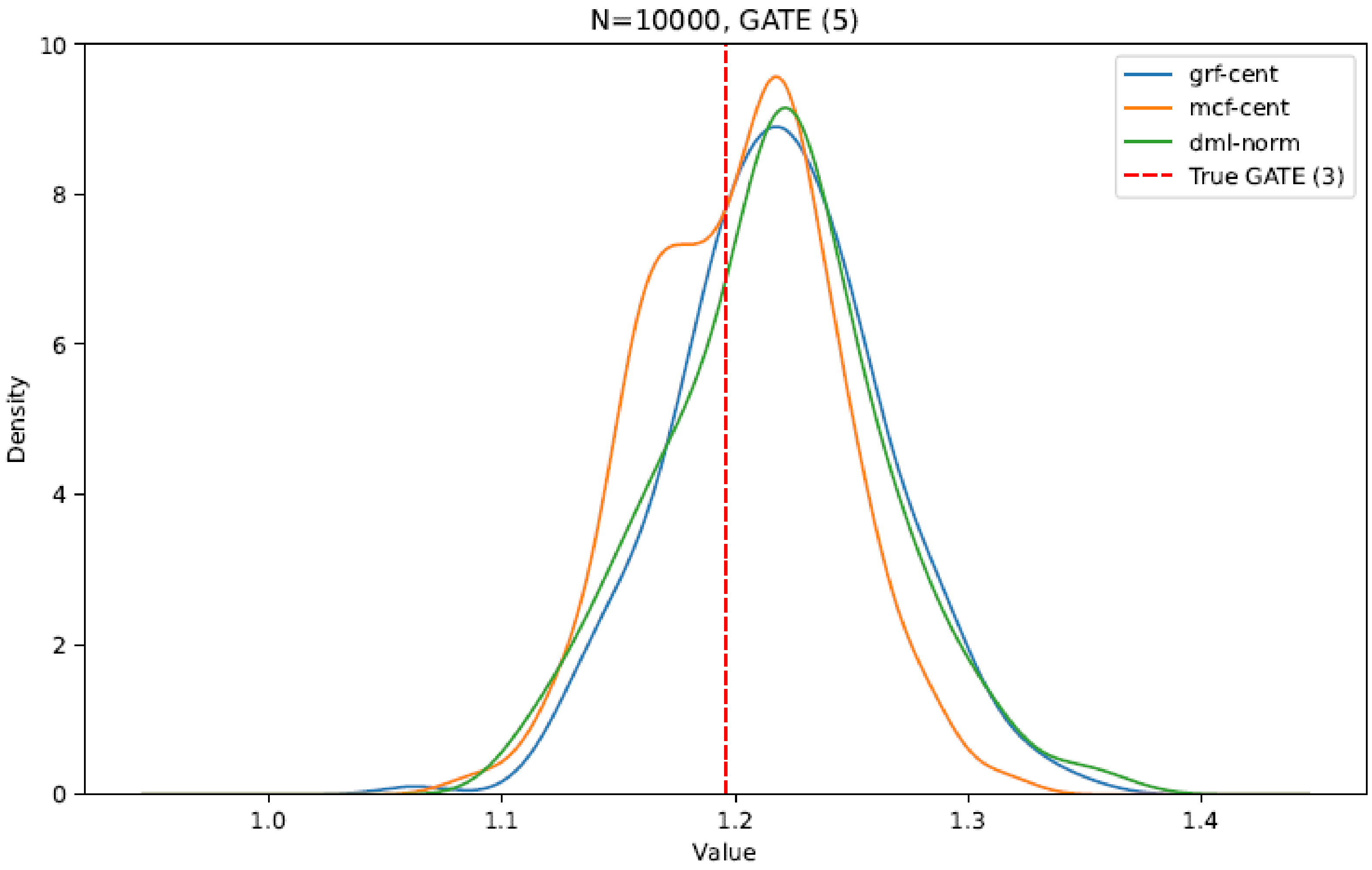
Population – medium selectivity



Population – strong selectivity



Groups – 5 groups



Groups – 40 groups