Meta-Learners for Estimation of Causal Effects: Finite Sample Cross-Fit Performance

YSEM, 2022

Gabriel Okasa

TIS-EPFL
Chair for Technology and Innovation Strategy
Swiss Federal Institute of Technology in Lausanne, Switzerland

EPFL

Motivation

Research Question:

- What is the finite sample performance of machine learning based meta-learners using cross-fitting for estimation of heterogeneous causal effects?

Motivation

Meta-Learners:

- flexibility in estimation of heterogeneous causal effects
- generality in the choice of the learning method (Künzel et al. 2019)
- lack of unifying simulation evidence for assessment of meta-learners

Cross-Fitting:

- overfitting bias due to estimation of nuisance functions (Chernozhukov et al. 2018)
- sample-splitting and cross-fitting to reduce bias and regain efficiency
- lack of simulation evidence for assessment of estimation procedures

Methods

Data Inputs:

- treatment indicator $W_i \in \{0, 1\}$
- outcome variable Y_i
- covariates X_i

Nuisance Functions:

- propensity score function $e(x) = \mathbb{P}[W_i = 1 \mid X_i = x]$
- response function $\mu(x) = \mathbb{E}[Y_i \mid X_i = x]$

Meta-Learning:

- treatment effect function $\tau(x) = \zeta\big(W_i, X_i, Y_i, e(x), \mu(x)\big)$

EPFL

Methods

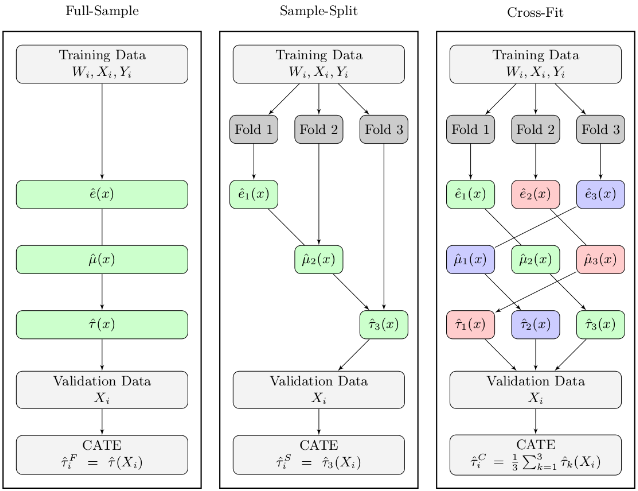


Figure 1: Illustration of the full-sample, sample-splitting and cross-fitting procedure.

Methods

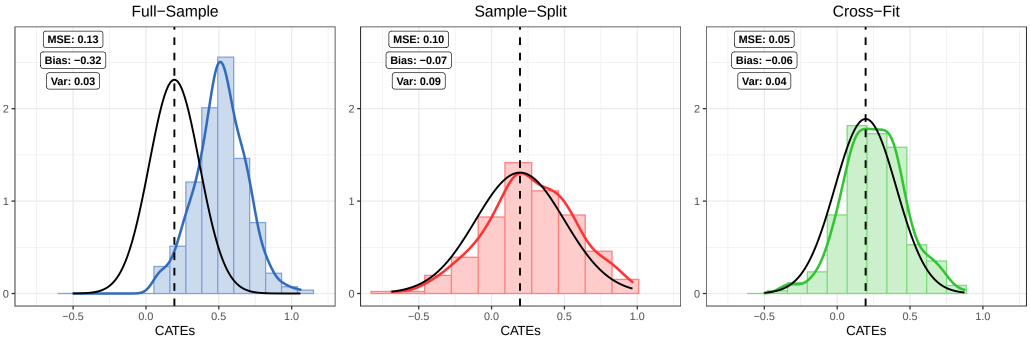


Figure 2: CATE distributions under full-sample, sample-splitting and cross-fitting estimation.

Methods

Framework:

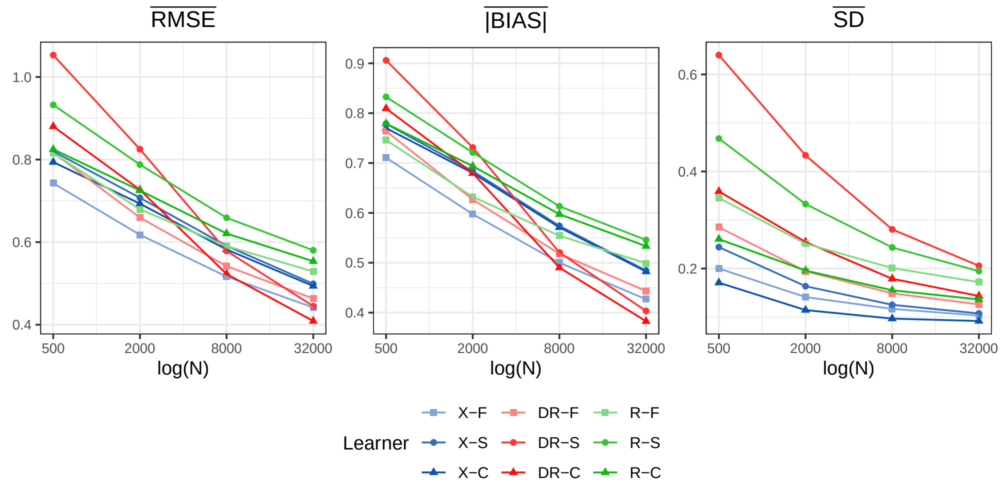
- identification based on the selection-on-observables strategy
- implementations based on the full-sample, sample-splitting and cross-fitting
- meta-learners based on the random forest algorithm

Simulation Study:

- synthetic and empirical simulations
- DGPs with unequal treatment shares, non-linearities and large-dimensions
- varying sample sizes up to 32’000 observations

Results

Figure 3: Results for Main Simulation: unbalanced treatment and nonlinear CATE



Results

Estimation Procedures:

- sample-splitting effectively reduces the bias in large samples
- cross-fitting additionally regains the full sample size efficiency
- full-sample estimation preferable in small samples when using machine learning

Meta-Learners:

- varying impacts of the estimation procedures on the performance of meta-learners
- X-learner suitable for imbalanced treatment shares in any version and sample size
- DR-learner suitable for balanced treatment shares using cross-fitting in large samples

Results

Takeaway:

- The performance of meta-learners varies greatly but the choice of the meta-learner and the estimation procedure can be guided by observable data characteristics.

Thank You for Your Attention!

gabriel.okasa@epfl.ch
okasag.github.io

References

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. "Double/debiased machine learning for treatment and structural parameters". In: *Econometrics Journal* 21.1 (2018), pp. 1–68.

Künzel, Sören R., Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. "Metalearners for estimating heterogeneous treatment effects using machine learning". In: *Proceedings of the National Academy of Sciences* 116.10 (2019), pp. 4156–4165.