Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence

Michael Knaus, Michael Lechner, Anthony Strittmatter
University of St. Gallen
michael.knaus@unisg.ch, mcknaus.github.io

Swiss Institute for Empirical Economic Research
University of St.Gallen
75 NRP Big Data National Research Programme

Motivation

- Recent decades have seen **huge advances** in the **identification** and **estimation** of **average causal effects.**
- **Less progress** in the estimation of **conditional/individualized/personalized average treatment** effects until recently many **new estimators** were proposed in **different literatures** (statistics, epidemiology, econometrics, computer science).
- Currently **limited knowledge** about the **relative performance** of different methods because we are **currently lacking** ...
  - ... unifying theoretical framework
  - ... simulation evidence of finite sample performance in realistic (economic) settings

⇒ **Limited guidance** for practitioners

⇒ **RQ1**: Can we **organize** the **estimators** proposed so far in a meaningful way?

⇒ **RQ2**: What is the **relative performance** of **causal machine learning** estimators for **heterogeneous causal effects** in **observational studies** with **binary treatments**?

Notation

Observed variables:
- Y: outcome of interest
- D: binary treatment indicator
- X: covariates

Estimand of interest:
- Conditional average treatment effect: $\tau(x) = E[Y^1 - Y^0 \mid X = x]$
- Y^d: potential outcomes

Nuisance parameters for estimators:
- $p(x) = P[D = 1 \mid X = x]$
- $\mu(x) = E[Y \mid X = x]$
- $\mu(d, x) = E[Y \mid D = d, X = x]$

Approaches

Generic approaches combined with **Random Forest** (RF) and **Lasso**:
- Conditional mean regression (CMR)
- Modified covariate method (MCM) [1]
- R-Learning [2]
- Modified outcome methods (MOM) with
  - inverse probability weighting (IPW)
  - doubly robust estimator (DR)

Estimator specific approach:
- Causal Forest (CF) [3]

Empirical Monte Carlo Study

Mimics **active labor market policy evaluation** with Swiss administrative data. Empirical Monte Carlo Study informs data generating processes (DGPs) as much as possible by real data. Total of 24 different DGPs with varying ...
- size of heterogeneity (no/medium/large)
- random noise in heterogeneity (yes & no)
- selection (yes & no)
- sample size (1000 & 4000)

Main findings

- **Four estimators** perform **consistently well** in all settings
- **All four** use **treatment *and* outcome information** for estimation
- **Lasso based** estimators **non-normal** due to excess kurtosis
- **Random Forest based** estimators remarkably **normal**

References

[1] L. Tian, A. A. Alizadeh, A. J. Gentles, and R. Tibshirani. A Simple Method for Estimating Interactions between a Treatment and a Large Number of Covariates. *J. Am. Stat. Assoc.*, 2014.

[2] X. Nie and S. Wager. Quasi-oracle estimation of heterogeneous treatment effects. *arXiv:1712.04912*, 2017.

[3] S. Athey, J. Tibshirani, S. Wager, et al. Generalized random forests. *Annals of Statistics*, 47(2):1148–1178, 2019.

Overview of estimators

- **Conditional mean regressions** is the simplest approach taking the differences of estimated conditional expectations for treated ($\hat{\mu}(1,x)$) and non-treated ($\hat{\mu}(0,x)$):

$$\hat{\tau}_{CMR}(x) = \hat{\mu}(1,x) - \hat{\mu}(0,x). \quad (1)$$

- The **remaining generic estimators** can be summarized as solving a **weighted minimization problem with modified outcomes,**

$$\min_{\tau} E\left\{W\left[Y^* - \tau(X)\right]^2\right\} \quad (2)$$

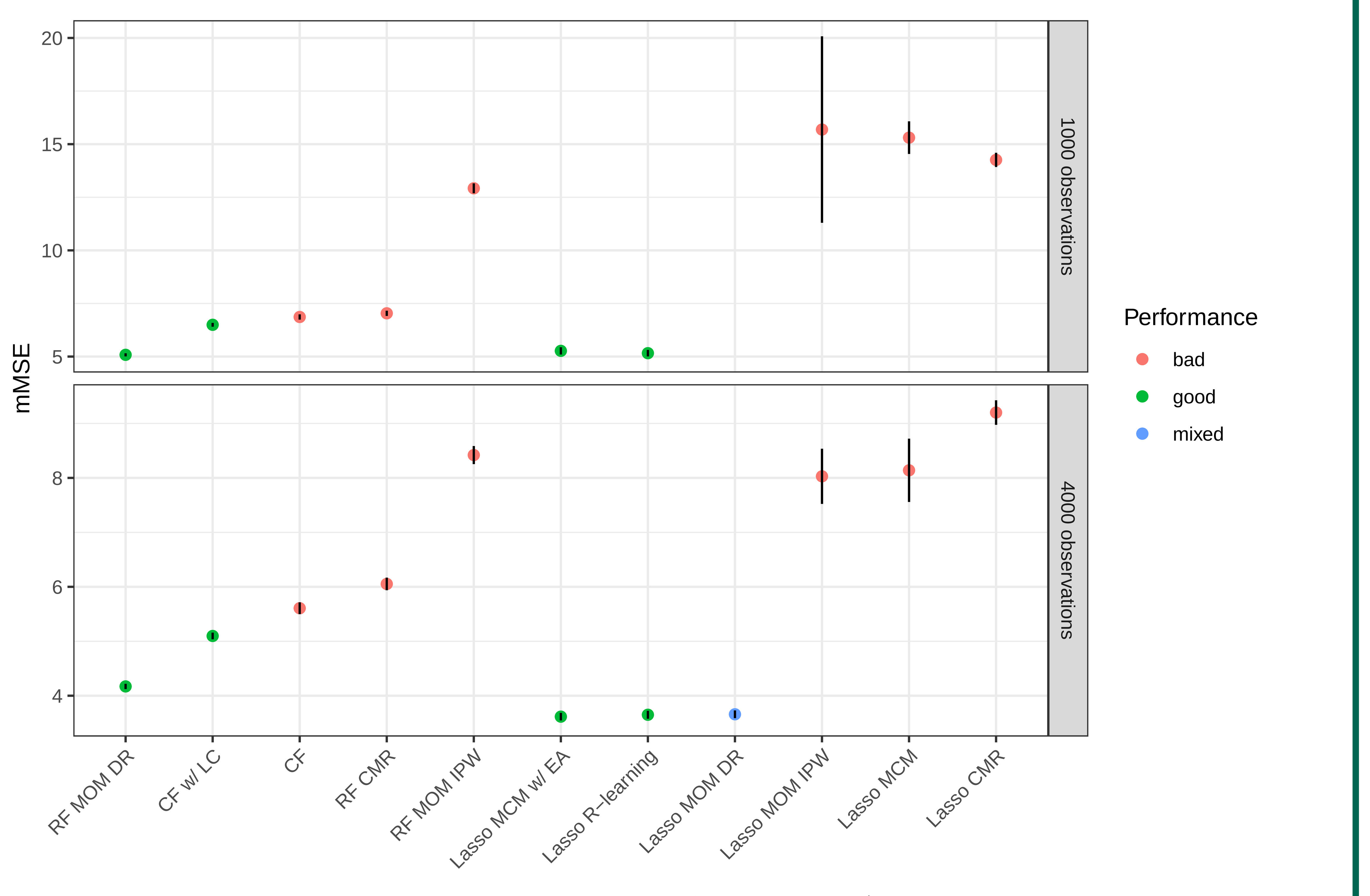
- with weights W and modified outcomes Y^* in the following way ($T = 2D - 1 \in \{-1, 1\}$):

Approach	W	Y^*
MCM (w/ EA)	$\left\lvert\frac{D-p(X)}{p(X)(1-p(X))}\right\rvert$	$2TY$ or $2T(Y - \mu(X))$
R-Learning	$[D - p(X)]^2$	$\frac{Y-\mu(X)}{D-p(X)}$
MOM IPW	1	$Y\frac{D-p(X)}{p(X)(1-p(X))}$
MOM DR	1	$\mu(1,X) - \mu(0,X) + \frac{D(Y-\mu(1,X))}{p(X)} - \frac{(1-D)(Y-\mu(0,X))}{(1-p(X))}$

- **Causal Forest** modifies the spitting criterion of standard Random Forests. It boils down to taking the **difference of two weighted means** where the **weights are tailored** to estimate heterogeneous effects. **Local centering** (LC) to **account** for **confounding**.

⇒ Combining the generic approaches with Random Forest and Lasso as well as some variants of the above estimators leads to a **total of eleven compared estimators**.

Baseline results



Notes: mMSE is the mean MSE of 10,000 out-of-sample CATEs in 2000/500 replications. Black lines indicate two standard (simulation) errors. The **mMSE of Lasso MOM DR for 1000 observations is 48**.