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: $\tau(x) = E[Y^1 - Y^0 \mid X = x]$
- $Y^0$: potential outcomes

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

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 ($\hat{\mu}(0, x)$):
  $$\hat{\tau}_{CMR}(x) = \hat{\mu}(1, x) - \hat{\mu}(0, x).$$
- The remaining generic estimators can be summarized as solving a weighted minimization problem with modified outcomes,
  $$\min E \left\{ W \{ Y^* - \tau(X) \}^2 \right\}$$
- with weights $W$ and modified outcomes $Y^*$ in the following way ($T = 2D - 1 \in \{-1, 1\}$:

<table>
<thead>
<tr>
<th>Approach</th>
<th>$W$</th>
<th>$Y^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCM (w/ EA)</td>
<td>$D - p(X)$</td>
<td>$2TY$ or $2T(Y - \mu(X))$</td>
</tr>
<tr>
<td>R-Learning</td>
<td>$[D - p(X)]^2$</td>
<td>$\frac{Y - \mu(X)}{p(X)}$</td>
</tr>
<tr>
<td>MOM IPW</td>
<td>1</td>
<td>$\frac{D - p(X)}{p(X)(1-p(X))}$</td>
</tr>
<tr>
<td>MOM DR</td>
<td>1</td>
<td>$\hat{\mu}(1, X) - \hat{\mu}(0, X)$ + $\frac{D(Y - \hat{\mu}(1, X))}{p(X)} - \frac{(1-D)(Y - \hat{\mu}(0, X))}{(1-p(X))}$</td>
</tr>
</tbody>
</table>

- 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.

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)

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]

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

Baseline results

References