Sparsity-inducing Bayesian Causal Forest with Instrumental Variable

Jens Klenke[1,^], Lennard Maßmann[1,2,*]

[1]Chair of Econometrics, University of Duisburg-Essen
[2]Ruhr Graduate School in Economics

[*]E-mail: lennard.massmann@uni-due.de
[^]E-mail: jens.klenke@vwl.uni-due.de

July 25, 2024

Abstract

Your text:Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetuer odio sem sed wisi.

Contents

1 Introduction

Hahn et al. (2020)

1.1 Literature

- Using machine learning to infer heterogeneous effects in observational studies often focuses on CATE estimation under regular assignment mechanisms
- Focus in this work: Methods to discover heterogeneous effects in the presence of imperfect compliance
- Methods
  - tree-based
  - ensemble-of-trees
  - deep-learning-based methods
- BCF-IV: Discovers and estimates HTE in an interpretable way
  - BCF: BART-based semi-parametric Bayesian regression model, able to estimate HTE in regular assignment mechanisms, even with strong confounding
  - Use BCF to estimate $\hat{\tau}_C(x)$ and $\widehat{ITT}_Y(x)$ such that the conditional Complier Average Causal Effect $\hat{\tau}^{cace}(x) = \frac{\widehat{ITT}_Y(x)}{\hat{\tau}_C(x)}$
- - BART benefits in general
  - good performance in high-noise settings
  - shrinkage to/emphasize on low-order interactions
  - established software implementations ('BayesTree', 'bartMachine', 'dbarts')
- BART shortcomings
  - non-smooth predictions as BART prior produces stepwise-continuous functions
  - BART prior is overconfident in regions with weak common support

- Research proposal: Rewrite the BCF-IV model with SoftBART instead of BART prior to account for sparsity

1.2 Contribution

We contribute to the literature on machine learning techniques designed to uncover heterogeneous effects while addressing imperfect compliance by developing specialized adaptations of tree-based algorithms. Specifically, we generalize Bayesian Instrumental Variable Causal Forest (BCF-IV) from (Bargagli-Stoffi et al. 2022). BCF-IV is a semi-parametric Bayesian regression model that builds directly on the Bayesian Additive Regression Trees (BART) algorithm (Chipman et al., 2010). Instead of using BART fur pure predictions of outcomes, BCF-IV is designed to identify and estimate heterogeneous effects within the subpopulation of units that comply with the treatment assignment, known as compliers. Consequently, the estimated effects can be considered doubly local, representing subgroup effects within the compliers subpopulation. BCF-IV identifies heterogeneity through an interpretable tree structure, with each node representing a distinct subgroup. For each leaf node of the generated tree, BCF-IV supports multiple hypothesis test adjustments to control the Type I error rate (familywise error rate) or the false discovery rate.

In our work, we extend BCF-IV in several ways. First, we use the shrinkage prior adaptation of BART (SoftBART) as proposed in (Linero & Yang 2018, Linero 2018). More precisely, by dividing the conditional Intention-To-Treat (cITT) effects with the corresponding conditional Proportion of Compliers (cPC), one can show to arrive at the conditional Complier Average Causal Effect (cCACE). In the discovery step the original BCF-IV algorithm of (Bargagli-Stoffi et al. 2022) uses the Bayesian Causal Forest (BCF) (Hahn et al. 2020) to estimate the cITT effects. BCF is a nonlinear regression model that builds upon BART and is proposed for estimating heterogeneous treatment effects. It is specifically designed for scenarios characterized by small effect sizes, heterogeneous effects, and significant confounding by observables. First, BCF addresses the issue of highly biased treatment effect estimates in the presence of strong confounding by incorporating an estimate of the propensity function directly into the response model. Thereby, it induces a covariate-dependent prior on the regression function. Second, BCF allows for the separate regularization of treatment effect heterogeneity from the prognostic effect of control variables. Conventional response surface modeling approaches often fail to adequately model regularization over effect heterogeneity. Instead, BCF enables an informative shrinkage towards homogeneity such that one is able to control the degree of regularization over effect heterogeneity. In our work, we generalize the cITT effect estimation by replacing BCF with the Shrinkage Bayesian Causal Forest (SBCF) proposed by (Caron et al.

2022). SBCF extends BCF by using additional priors proposed in SoftBART that enable to adjust the influence of each covariate based on the number of corresponding splits in the tree ensemble. These priors enhance the model's adaptability to sparse data-generating processes and facilitate fully Bayesian feature shrinkage within the framework for estimating treatment effects. Consequently, it improves to uncover the moderating factors that drive heterogeneity when there is sparsity in the data. Moreover, this method allows the incorporation of prior knowledge regarding relevant confounding covariates and the relative magnitude of their impact on the outcome.

Second, BCF-IV uses a single binary tree based on the CART model of (Breiman1984) to analyse possible heterogeneity patterns within cCACE. In our work, we use posterior splitting probabilities retrieved from SBCF as a measure of variable importance within the `cost`-argument of `rpart`. These are scalings to be applied when considering splits, so the improvement on splitting on a variable is divided by its cost in deciding which split to choose in `rpart`.

Open questions for further contributions. Use Random Forests (Causal Rule Ensemble) instead of single CART for subgroup discovery? More rigorous Bayesian estimation by replacing `ivreg` with `brms` (implementation of credible intervals, estimation error, counterparts to something like weak instrument tests, bayesian model averaging)?

2 Potential Outcomes and IV

2.1 Potential Outcomes and ITT

- Y_i: outcome variable; W_i: treatment variable; Z_i: instrumental variable
- $\mathbb{X}$: $N \times P$ matrix of control variables
- G_i: sub-populations of units
  - $G_i = C : W_i(Z_i = 0) = 0, Wi(Zi = 1) = 1$
  - $G_i = D : W_i(Z_i = 0) = 1, W_i(Z_i = 1) = 0$
  - $G_i = AT : W_i(Z_i = 0) = 1, W_i(Z_i = 1) = 1$
  - $G_i = NT : W_i(Z_i = 0) = 0, W_i(Z_i = 1) = 0$

Given the Stable Unit Treatment Value Assumption (SUTVA), one can postulate the existence of potential outcomes $Y_i(W_i)$ such that $Y_i^{obs} = Y_i(1)W_i + Y_i(0)(1 - W_i)$.

One can directly get from the data the effect of the assignment to treatment:

$$\begin{aligned} ITT_Y &= \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] \\ &= \pi_C ITT_{Y,C} + \pi_D ITT_{Y,D} + \pi_{AT} ITT_{Y,AT} + \pi_{NT} ITT_{Y,NT}. \end{aligned}$$

2.2 IV assumptions

Assumptions to infer Complier Average Causal Effect (CACE),

$$\tau^{cace} = ITT_{Y,C} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[W_i|Z_i = 1] - \mathbb{E}[W_i|Z_i = 0]} = \frac{ITT_Y}{\pi_C},$$

and its conditional version (cCACE),

$$\tau^{cace}(x) = \frac{\mathbb{E}[Y_i|Z_i = 1, \mathbb{X}_i = x] - \mathbb{E}[Y_i|Z_i = 0, \mathbb{X}_i = x]}{\mathbb{E}[W_i|Z_i = 1, \mathbb{X}_i = x] - \mathbb{E}[W_i|Z_i = 0, \mathbb{X}_i = x]} = \frac{ITT_Y(x)}{\pi_C(x)}.$$

1. exclusion restriction: $Y_i(0) = Y_i(1),\ \text{for } G_i \in \{AT, NT\}$.
2. monotonicity: $W_i(1) \geq W_i(0) \rightarrow \pi_D = 0$.
3. existence of compliers: $P(W_i(0) < W_i(1)) > 0 \rightarrow \pi_C \neq 0$.
4. unconfoundedness of the instrument: $Z_i \perp (Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), W_i(0), W_i(1))$.

2.3 Estimation

The conditional CACE can be estimated in a generic sub-sample (i.e., for each $\mathbf{X}_i \in \mathbb{X}_j$, where $\mathbb{X}_j$ is a generic node of the discovered tree, like a non-terminal node or a leaf) as:

3 Sparsity-inducing Bayesian Causal Forest with Instrumental Variable

3.1 Sparse Bayesian Causal Forest

3.2 Sparse BCF-IV

4 Simulation study

Performance criteria according to bargagli-stoffi:

1. Average number of truly discovered heterogeneous subgroups corresponding to the nodes of the generated CART (proportion of correctly discovered subgroups);

2. Monte Carlo estimated bias for the heterogeneous subgroups:

$$\text{Bias}_m(I_{\text{inf}}) = \frac{1}{N_{\text{inf}}} \sum_{i=1}^{N_{\text{inf}}} \sum_{l=1}^{L} \left(\tau_{\text{cace},i}(\ell) - \hat{\tau}_{\text{cace},i}(\ell, \Pi_m, I_{\text{inf}}) \right), \tag{4.0.1}$$

$$\text{Bias}(I_{\text{inf}}) = \frac{1}{M} \sum_{m=1}^{M} \text{Bias}_m(I_{\text{inf}}), \tag{4.0.2}$$

where Π_m is the partition selected in simulation m, L is the number of subgroups with heterogeneous effects (i.e., two for the case of strong heterogeneity and four for the case of slight heterogeneity), and N_{inf} is the number of observations in the inference sample.

3. Monte Carlo estimated MSE for the heterogeneous subgroups:

$$\text{MSE}_m(I_{\text{inf}}) = \frac{1}{N_{\text{inf}}} \sum_{i=1}^{N_{\text{inf}}} \sum_{l=1}^{L} \left(\tau_{\text{cace},i}(\ell) - \hat{\tau}_{\text{cace},i}(\ell, \Pi_m, I_{\text{inf}}) \right)^2, \tag{4.0.3}$$

$$\text{MSE}(I_{\text{inf}}) = \frac{1}{M} \sum_{m=1}^{M} \text{MSE}_m(I_{\text{inf}}); \tag{4.0.4}$$

4. Monte Carlo coverage, computed as the average proportion of units for which the estimated 95% confidence interval of the causal effect in the assigned leaf includes the true value, for the heterogeneous subgroups:

$$C_m(I_{\text{inf}}) = \frac{1}{N_{\text{inf}}} \sum_{i=1}^{N_{\text{inf}}} \sum_{l=1}^{L} \left(\tau_{\text{cace},i}(\ell) \in \hat{\text{CI}}_{95} \left(\hat{\tau}_{\text{cace},i}(\ell, \Pi_m, I_{\text{inf}}) \right) \right), \tag{4.0.5}$$

$$C(I_{\text{inf}}) = \frac{1}{M} \sum_{m=1}^{M} C_m(I_{\text{inf}}). \tag{4.0.6}$$

5 Empirical application

6 Discussion

Bibliography

Bargagli-Stoffi, F. J., Witte, K. D. & Gnecco, G. (2022), 'Heterogeneous causal effects with imperfect compliance: A Bayesian machine learning approach', *The Annals of Applied Statistics* **16**(3), 1986–2009. Publisher: Institute of Mathematical Statistics.
URL: *https://projecteuclid.org/journals/annals-of-applied-statistics/volume-16/issue-3/Heterogeneous-causal-effects-with-imperfect-compliance–A-Bayesian-machine/10.1214/21-AOAS1579.full*

Caron, A., Baio, G. & Manolopoulou, I. (2022), 'Shrinkage Bayesian Causal Forests for Heterogeneous Treatment Effects Estimation', *Journal of Computational and Graphical Statistics* **31**(4), 1202–1214. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10618600.2022.2067549.
URL: *https://doi.org/10.1080/10618600.2022.2067549*

Hahn, P. R., Murray, J. S. & Carvalho, C. M. (2020), 'Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects (with Discussion)', *Bayesian Analysis* **15**(3), 965–1056. Publisher: International Society for Bayesian Analysis.
URL: *https://projecteuclid.org/journals/bayesian-analysis/volume-15/issue-3/Bayesian-Regression-Tree-Models-for-Causal-Inference–Regularization-Confounding/10.1214/19-BA1195.full*

Linero, A. R. (2018), 'Bayesian Regression Trees for High-Dimensional Prediction and Variable Selection', *Journal of the American Statistical Association* **113**(522), 626–636. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2016.1264957.
URL: *https://doi.org/10.1080/01621459.2016.1264957*

Linero, A. R. & Yang, Y. (2018), 'Bayesian Regression Tree Ensembles that Adapt to Smoothness and Sparsity', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **80**(5), 1087–1110.
URL: *https://academic.oup.com/jrsssb/article/80/5/1087/7048381*