Shrinkage 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

October 7, 2024

Abstract

This paper focuses on improving the estimation of heterogeneous treatment effects in observational studies under sparsity and conditions of imperfect compliance using instrumental variables (IV). Traditional IV methods, such as two-stage least squares (2SLS), often impose linearity assumptions that may not hold in complex empirical settings. To address these limitations, the Bayesian Instrumental Variable Causal Forest (BCF-IV) framework has been developed to estimate the conditional Complier Average Causal Effect (CACE) non-parametrically while retaining interpretability. BCF-IV, based on the Bayesian Additive Regression Trees (BART) algorithm, identifies treatment effect heterogeneity within the subgroup of compliers using 2SLS leafwise. This research contributes in two significant ways by proposing the Shrinkage Bayesian Instrumental Variable Causal Forest (SBCF-IV) algorithm. First, the paper adopts the Shrinkage Bayesian Causal Forest (SBCF) algorithm, which integrates shrinkage priors to enable more precise treatment effect estimation in the presence of sparse data. Second, the paper refines the discovery of heterogeneous subgroups by incorporating posterior splitting probabilities into the decision-making process. These probabilities are used to scale the cost function during subgroup construction, leading to more accurate identification of meaningful subgroups. The approach of SBCF-IV enhances the ability to manage sparse data and improves the detection of variables that drive treatment effect heterogeneity. Overall, a simulation study suggests improved adaptability and interpretability in estimating conditional CACE, particularly in scenarios with sparsity, confounding and nonlinearity.

Contents

1 Introduction

Using machine learning to infer heterogeneous effects in observational studies often focuses on different forms of average treatment effect estimations under regular assignment mechanisms **?**. In this work, we focus on methods to discover and estimate heterogeneous treatment effects in the presence of imperfect compliance via an irregular assignment mechanism using instrumental variables (IV). Many IV frameworks are rooted in the two-stage least squares (2SLS) approach. In the first stage, the treatment is predicted using the instruments, followed by the second stage, where the predicted values of the treatment are utilized to estimate the unconfounded effect of the treatment on the outcome. According to standard IV assumptions, the variation in the predicted treatment variable should be independent of confounding variables, thereby ensuring that the estimated effect remains unbiased. Consequently, the second-stage results can be interpreted causally, provided that the instruments meet the necessary validity criteria. However, a key limitation of 2SLS methods is their reliance on linearity assumptions, which may introduce challenges in certain empirical contexts. Therefore, nonparametric IV approaches based on series approximations, control functions or stratification have been proposed which relax the assumptions of linearity and additivity like (**??**). In this work, we focus on stratification as a remedy to relax 2SLS while allowing for heterogeneity in the Complier Average Causal Effect (CACE) which is the treatment effect conditional on being a complier under an Instrumental Variable (IV) framework. Current methods like tree-based ensemble models or neural networks have been proposed to estimate heterogeneous treatment effects under imperfect compliance and simulation studies suggest precise treatment effect estimates (**??**). However, they lack comprehensibility due to complex, nonlinear parameterizations of the feature space. Additionally, algorithms relying on random forests need huge datasets for convergence and asymptotic results while neural networks are required to search through a lot of possible and sensitive tuning parameter configurations. Single tree-based algorithms have been proposed to mitigate the issues of asymptotics and parameter tuning while retaining interpretability (Bargagli Stoffi & Gnecco 2020, **?**, **?**). However, those single tree algorithms suffer from instability and worse predictive quality. Specifically, this paper generalizes the Bayesian Instrumental Variable Causal Forest (BCF-IV) algorithm proposed in Bargagli-Stoffi et al. (2022) to estimate conditional Complier Average Causal Effect (cCACE) accurately when there are many covariates leading to a sparse dataset. 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 the predictive BART algorithm for pure forecasts 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. In our work, we extend BCF-IV in several ways. As a first contribution, we use the shrinkage prior adaptation of SoftBART as proposed in (Linero & Yang 2018, Linero 2018) instead of the usual priors of BART. More precisely, by dividing the conditional Intention-To-Treat effects (cITT) with the corresponding conditional Proportion of Compliers, one can show to arrive at the 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 and find heterogeneous subgroups. 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. The second contribution of our work revolves around the usage of posterior splitting probabilities to improve the disovery of meaningful heterogeneous subgroups. 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`.

The estimation of the Complier Average Causal Effect (CACE) which is the treatment effect conditional on being a complier under an Instrumental Variable (IV) framework.

- 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

2 Potential outcomes and irregular assignmnet

We follow the setup of Bargagli Stoffi & Gnecco (2020) and Bargagli-Stoffi et al. (2022) who use the usual notation of the Rubin's causal model. Given a set of N indiviudals, indexed by $i = 1, \ldots, N$, we denote with Y_i a generic outcome variable, with W_i a binary treatment indicator, with X an $N \times P$ matrix of P control variables, and with X_i the i-th P-dimensional row vector of covariates. Let us define the pair of potential outcomes $Y_i(W_i)$ by using the Stable Unit Treatment Value Assumption (SUTVA).

Assumption 2.1. Stable unit treatment value assumption (SUTVA).

$$\text{If } W_i = w, \text{ then } Y_i(w) = Y_i^{obs} \ , \forall w \in \{0,1\} \ , \forall i \in \{1, ..., N\}.$$

Therefore, it holds that $Y_i(W_i = 1) = Y_i(1)$ for an indiviudal i under assignment to the treatment group and $Y_i(W_i = 0) = Y_i(0)$ under assignment to control group. Despite one is unable to observe both potential outcomes at the same time for each individual one is able to observe the potential outcome that aligns with the assigned treatment status such that

$$Y_i^{\text{obs}} = Y_i(1)W_i + Y_i(0)(1 - W_i).$$

Under the following strong ignorability assumptions of Assumption 2.2 and Assumption 2.3 we operate under the regular assignment mechanism which, through Assumption 2.2, prevents the existence of unmeasured confounding and, by Assumption 2.3, allows for unbiased treatment effect estimation in support of the covariate space.

Assumption 2.2. Unconfoundedness.

$$W_i \perp\!\!\!\perp (Y_i(1), Y_i(0)) \, | X_i), \text{ or equivalently,}$$

$$Pr(W_i|(Y_i(1), Y_i(0), X_i) = Pr(W_i|X_i).$$

Assumption 2.3. Positivity.

$$\epsilon < p(X_i = x) < 1 - \epsilon \text{ with probability 1, } \forall \epsilon > 0, \forall x \text{ in support of } X_i.$$

Using strong ignorability, one can define, for instance, the CATE to analyze heterogeneous treatment effects by

Definition 2.1. Conditional Average Treatmnet Effect (CATE).

$$\tau(x) = \mathbb{E}[Y_i^{\text{obs}} \mid W_i = 1, X_i = x] - \mathbb{E}[Y_i^{\text{obs}} \mid W_i = 0, X_i = x].$$

In this work, we deviate from this regular assignment mechanism that is implied by Assumptions 2.2 and 2.3 and operate under the scenario of an irregular assignment mechanism. This irregular assignment mechanism allows for a violation of compliance between (quasi-)randomized treatment assignment, Z_i, and treatment receipt, W_i, such that just the assignment to treatment is assumed to be unconfounded through a valid choice of Z_i but the treatment receipt might be confounded. A remedy to the issue of confoundedness in treatment receipt is the Instrumental Variable (IV) approach. A proper instrumental variable, Z_i, affects treatment receipt, W_i, while not being allowed to affect Y_i directly such that treatment receipt depends on the treatment assignment by $W_i(Z_i)$. Based on the functional relation between W_i and Z_i and in case of a binary treatment variable, one can define four subgroups of individuals, G_i, such that

Definition 2.2. Subgroups G_i.

$$G_i = \begin{cases} C, & \text{if } W_i(Z_i = 0) = 0, W_i(Z_i = 1) = 1 \\ D, & \text{if } W_i(Z_i = 0) = 1, W_i(Z_i = 1) = 0 \\ AT, & \text{if } W_i(Z_i = 0) = 1, W_i(Z_i = 1) = 1 \\ NT, & \text{if } W_i(Z_i = 0) = 0, W_i(Z_i = 1) = 0 \end{cases}.$$

where C, D, AT and NT are abbreviations for Compliers, Defiers, Always-Takers, Never-Takers. The proportion of individuals that belong to each subgroup is defined as π_{G_i}, i.e. the proportion of compliers reads π_C. Considering the distinction between Z_i and W_i, the Intention-To-Treat (ITT) effect is defined as

Definition 2.3. Intention-To-Treat (ITT) effect.

$$ITT_Y = \mathbb{E}\left[Y_i | Z_i = 1\right] - \mathbb{E}\left[Y_i | Z_i = 0\right].$$

The ITT effect refers to the instrument's average effect. Based on the subgroups in Definition 2.2 and their proportions π_{G_i}, an IV setting can be formalized by the usual four IV assumptions following **?**.

Assumption 2.4. Classical IV assumptions with a binary treatment.

Exclusion restriction:	$Y_i(0) = Y_i(1), \text{ for } G_i \in \{AT, NT\}.$
Monotonicity:	$W_i(1) \geq W_i(0) \rightarrow \pi_D = 0.$
Existence of compliers:	$P(W_i(0) < W_i(1)) > 0 \rightarrow \pi_C \neq 0.$
Unconfoundedness of IV:	$Z_i \perp (Y_i(0,0), Y_i(0,1), Y_i(1,0), Y_i(1,1), W_i(0), W_i(1)).$

If Assumptions 2.4 hold, the Complier Average Causal Effect (CACE) can be identified.

Definition 2.4. Complier Average Causal Effect (CACE).

$$\tau_{\text{CACE}} = \frac{\text{ITT}_Y}{\pi_C} = \frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[W_i \mid Z_i = 1] - \mathbb{E}[W_i \mid Z_i = 0]},$$

The CACE can be estimated from observational data where the numerator represents the average effect of the instrument, also referred to as the Intention-To-Treat (ITT) effect. The denominator represents the overall proportion of indiviudals that comply with the treatment assignment, also referred to as the proportion of compliers **?**. CACE is also sometimes referred to as Local Average Treatment Effects (LATE, see **?**) and represents the estimate of the causal effect of the assignment to treatment on the principal outcome, Y_i, for the subpopulation of compliers **?**. Consider the system of two simultaneous equations

$$\begin{aligned} Y_i^{obs} &= \alpha + \tau_{CACE} W_i + \epsilon, \\ W_i &= \pi_0 + \pi_C Z_i + \eta_i, \end{aligned}$$

with $\mathbb{E}(\epsilon_i) = \mathbb{E}(\eta_i) = 0$ and we assume by the first equation a linear projection of W_i onto Z_i with $\mathbb{E}(Z_i \eta_i) = 0$. Then, **?** and **?** show that τ_{CACE} can be estimated by a Two Stage Least Square (TSLS) estimator which is consistent and asymptotic normal as displayed in **?**.

In this work, we follow Bargagli-Stoffi et al. (2022) and consider the conditional

version of the CACE,

Definition 2.5. Conditional CACE (cCACE).

$$\tau_{\text{CACE}}(x) = \frac{\text{ITT}_Y(x)}{\pi_C(x)} = \frac{\mathbb{E}[Y_i \mid Z_i = 1, X_i = x] - \mathbb{E}[Y_i \mid Z_i = 0, X_i = x]}{\mathbb{E}[W_i \mid Z_i = 1, X_i = x] - \mathbb{E}[W_i \mid Z_i = 0, X_i = x]}.$$

The cCACE is a straightforward extension of the CACE in Definition 2.4 presented in Bargagli-Stoffi et al. (2022). A natural subgroup-related estimator $\tau_{\text{CACE}}(x)$ can be defined by acknowledging $\boldsymbol{X}_i \in \mathbb{X}_j$ with $\mathbb{X}_j$ being a pre-specified subgroup.

Definition 2.6. Estimator of cCACE.

$$\begin{aligned}\widehat{\tau}_{\text{CACE}}(x) &= \frac{\widehat{\text{ITT}}_Y(x)}{\widehat{\pi}_C(x)} \\ &= \frac{\frac{1}{N_{1,j}} \sum_{l:X_l \in \mathbb{X}_j} Y_l^{obs} Z_l - \frac{1}{N_{0,j}} \sum_{l:X_l \in \mathbb{X}_j} Y_l^{obs}(1 - Z_l)}{\frac{1}{N_{1,j}} \sum_{l:X_l \in \mathbb{X}_j} W_l Z_l - \frac{1}{N_{0,j}} \sum_{l:X_l \in \mathbb{X}_j} W_l(1 - Z_l)}\end{aligned}$$

Intuitively, Definition 2.6 implies to use TSLS subgroup-wise for every $\mathbb{X}_j$ under Assumptions 2.4. The system of two simulatenous equations from above can be conditionalized by

$$\begin{aligned}Y_{i,\mathbb{X}_j}^{obs} &= \alpha_{\mathbb{X}_j} + \tau_{\mathbb{X}_j}^{CACE} W_{i,\mathbb{X}_j} + \epsilon_{i,\mathbb{X}_j}, \\ W_{i,\mathbb{X}_j} &= \pi_{0,\mathbb{X}_j} + \pi_{C,\mathbb{X}_j} Z_{i,\mathbb{X}_j} + \eta_{i,\mathbb{X}_j},\end{aligned}$$

such that the reduced form reads

$$\begin{aligned}Y_{i,\mathbb{X}_j}^{\text{obs}} = &\left(\alpha_{\mathbb{X}_j} + \tau_{\text{CACE},\mathbb{X}_j} \pi_{0,\mathbb{X}_j}\right) + \\ &\left(\tau_{\text{CACE},\mathbb{X}_j} \pi_{C,\mathbb{X}_j}\right) Z_{i,\mathbb{X}_j} + \\ &\left(\varepsilon_{i,\mathbb{X}_j} + \tau_{\mathbb{X}_j}^{\text{CACE}} \eta_{i,\mathbb{X}_j}\right).\end{aligned}$$

The intercept $\left(\alpha_{\mathbb{X}_j} + \tau_{\text{CACE},\mathbb{X}_j} \pi_{0,\mathbb{X}_j}\right)$ and slope parameter $\left(\tau_{\text{CACE},\mathbb{X}_j} \pi_{C,\mathbb{X}_j}\right)$ can be estimated by ordinary least squares, given a sufficient number of $i.i.d$ observations in each subgroup $\mathbb{X}_j$. More information regarding theoretical properties of ... can be found in Appendix A of Bargagli-Stoffi et al. (2022). In Definition 2.6, the estimator $\widehat{\tau}_{\text{CACE}}(x)$ is defined by relying on the existence of accurately pre-specified

subgroups. The main contribution of BCF-IV in Bargagli-Stoffi et al. (2022) revolves around providing a full algorithm that (1) honestly splits the data into two disjunct subsamples $\mathcal{I}_{disc}, \mathcal{I}_{inf}$, (2) discovers heterogeneity in cCACE in an interpretable way using $\mathcal{I}_{disc}$ and (3) infers precise estimates of cCACE on $\mathcal{I}_{inf}$. The next chapter explains BCF-IV in detail and describes the extension to a sparsity-inducing version.

3 Bayesian Causal Forests and Instrumental Variables

This paper proposes an algorithm for the estimation of cCACE for sparse data scenarios. More precisely, we propose an extension of the BCF-IV algorithm in Bargagli-Stoffi et al. (2022) to handle scenarios with many irrelevant covariates in the dataset. Section 3.1 describes the original BCF-IV algorithm while section 3.2 explains the proposed extension based on the Shrinkage Bayesian Causal Forest. As pointed out in the literature review of this paper, current ensemble methods that operate under an irregular assignment mechanism with imperfect compliance face some difficulties. Algorithms like Deep IV **?** and the Generalized Random Forest **?** provide precise cCACE estimates but are rather uninformative about relevant covariates, or subsets of possibly many covariates, that drive heterogeneity in cCACE. Tree-based methods like the Causal Tree with IV Bargagli Stoffi & Gnecco (2020) propose to estimate treatment effects under imperfect compliance and the existence of a suitable IV while retainig interpretability. However, although the single tree structure enables interpretability it also lacks of stability and replicability. Bargagli-Stoffi et al. (2022) argue to overcome those shortcomings using the BCF-IV algorithm as outlined in the following Section 3.1.

3.1 BCF-IV

The main steps of the BCF-IV algorithm are outlined in Algorithm 1. Details of these three steps of honest sample splitting, discovery of treatment effect heterogeneity and inference of treatment effects are discussed in Sections 3.1.1, 3.1.2 and 3.1.3.

3.1.1 Honest sample splitting

The first step concerns honest sample splitting. This step enables a data-driven discovery of heterogeneous subgroups such that there is no need to specify those subgroups beforehand. Defining subgroups before estimating treatment effects based on relevant data of the studied population is a challenging task. It requires deep knowledge about the intricacies of the treatment effect at hand and may be prone to

overlook relevant subgroups. Honest sample splitting as proposed in **?** is a remedy for those issues by making distinctions between model selection and treatment effect inference.

Algorithm 1: Bayesian Causal Forest with Instrumental Variable (BCF-IV)

Input : N units i (X_i, Z_i, W_i, Y_i), with feature vector X_i, treatment assignment (instrumental variable) Z_i, treatment receipt W_i, observed response Y_i

Output : A tree structure discovering the heterogeneity in the causal effects and estimates of the Complier Average Causal Effects (CACE) within its leaves.

1. The Honest Splitting Step:

- Randomly split the total sample into a discovery subsample (I_{dis}) and an inference subsample (I_{inf}).

2. The Discovery Step (performed on I_{dis}):
Estimation of the Conditional CACE:

- (a) Estimate the conditional Intention-To-Treat: $\widehat{\text{ITT}}(x)$.
- (b) Estimate the conditional proportion of compliers: $\widehat{\pi}_C(x)$.
- (c) Estimate the conditional CACE, $\widehat{\tau}_{\text{CACE}}(x)$, using the estimated values from (a) and (b).

Heterogeneous subpopulations discovery:

- (d) Discover the heterogeneous effects by fitting a decision tree using the data $(\widehat{\tau}_{\text{CACE}}(x), X_i)$.

3. The Inference Step (performed on I_{inf}):

- (a) Estimate the $\widehat{\tau}_{\text{CACE}}(x)$ for all discovered subpopulations (i.e., nodes and leaves) in the tree discovered in Step 3(d).
- (b) Perform multiple hypothesis tests and adjust p-values to control for the familywise error rate or, less stringently, the false discovery rate.
- (c) Run weak-instrument tests within every node and discard nodes where weak-instrument issues are detected.

3.1.2 Discovery of heterogeneous subgroups

To estimate the conditional CACE given in Definition 2.6, we need some functional expression for the conditional expected value for the outcome, Y_i, as well as for the treatment indicator, W_i. The Bayesian Causal Forest (BCF) algorithm is proposed in Hahn et al. (2020) to use the Bayesian Additive Regression Trees (BART) algorithm **?** to estimate the CATE of Definition 2.1 in a regular assignment mechanism. BART is related to the CART algorithm of **?** which constructs binary trees by recursively partitioning the covariate space to produce accurate predictions. BART rests on a complete Bayesian probability model by using different regularizing prior distributions such that the overall model fit dominates fits of single trees. Distinct prior distributions are used for the complexity of the tree structure, data shrinkage within the nodes and

the variance of the error term. The same idea is now transferred to the setup of an irregular assignment mechanism. We start with modeling the numerator of Definition 2.6 and restrict our dataset to observations within $\mathcal{I}_{disc}$. Let the outcome variable Y_i be modelled semi-parametrically as

$$Y_i = f(Z_i, X_i) + \varepsilon_i$$

with $\varepsilon_i \sim \left(0, \sigma_\varepsilon^2\right)$. The conditional expected value of the outcome be defined similar to Hahn et al. (2020) as

$$\mathbb{E}\left[Y_i | Z_i = z, X_i = x\right] = \mu\left(\pi(x), x\right) + ITT_y(x)z.$$

Therefore, the conditional expectation is mainly governed by two additive functions. The first additive term $\mu\left(\pi(x), x\right)$ incorporates the IV's propensity score $\pi(x) = \mathbb{E}[Z_i = 1 | X_i = x]$ and accounts for the direct, treatment-independent influence of the control variables on the outcome variable. The usage of the propensity score $\pi(x)$ prevents targeted selection and regularization-induced confounding. The second additive component $ITT_y(x)$ accounts for the direct, possibly heterogeneous, intention-to-treat effect. While the first term follows the standard prior specifications as in **?**, the second term uses alternative tree depth penalty parameters ($\eta = 3, \beta = 0.25$) that encourage rather simplistic trees Hahn et al. (2020). Consequently, we can get estimates of $\widehat{ITT}_y(x)$ as required in the discovery step 2(a) of Algorithm 1. Analogously, we can model the denominator of Definition 2.6 by still restricting our dataset to observations within $\mathcal{I}_{disc}$ and using

$$W_i = f(Z_i, X_i) + \varphi_i$$

with $\varphi_i \sim \left(0, \sigma_\varphi^2\right)$. The conditional expected value of the treatment indicator is now defined as

$$\mathbb{E}\left[W_i | Z_i = z, X_i = x\right] = \delta(z, x)$$

This conditional expectation is governed by a standard BART prior such that the denominator of Definition 2.6 is estimated using BART in the sense of **?**. This gives us estimates $\widehat{\pi}_C(x)$ as required in discovery step 2 (b) of Algorithm 1 such that, together with $\widehat{ITT}_y(x)$, we can compute $\widehat{\tau}_{CACE}(x)$ straightforward. Finally, a CART algorithm is used on the estimates $\widehat{\tau}_{CACE}(x)$ to discover heterogeneity in CACE while retaining interpretability by providing transperancy on which variables have been used to construct the relevant subgroups.

3.1.3 Inference of conditional CACE

Finally, Definition 2.6 is used in all subgroups independently using the tree structure learned in Subsection 3.1.2 with unseen data from $\mathcal{I}_{inf}$ to infer conditional CACE by exploiting honest sample splitting outlined in Subsection 3.1.1.

3.2 Shrinkage 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

Further ideas

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)?

Theorems of conditional 2SLS

Moreover, **?** provide theorems for consistency and asymptotic normality for the unconditional TSLS estimator for the true population parameter τ_{CACE}. Bargagli-Stoffi et al. (2022) show that those theorems can be transferred to the conditional TSLS estimator if there are sufficient number of observations for every subgroup in which the cCACE is estimated.

Theorem 6.1 (Consistency and Asymptotic Normality of the Conditional 2SLS Estimator)**.** *Let Assumptions 1, 2, and 3 hold, i.e.,* $E(Z_{i,X_j}^2) \neq 0$ *(Assumption 1),* $E(Z_{i,X_j}\varepsilon_{i,X_j}) = 0$ *(Assumption 2), and* $\pi_{C,X_j} \neq 0$ *(Assumption 3). Then:*

1. *(Consistency)* $\hat{\tau}_{X_j}^{2SLS} - \tau_{X_j} \xrightarrow{p} 0$ *as* $N_{X_j} \to \infty$*, where* $\xrightarrow{p}$ *denotes convergence in probability, and* N_{X_j} *is the number of observations within the node* X_j*.*
2. *(Asymptotic Normality) If, in addition,* $E(Z_{i,X_j}^2\varepsilon_{i,X_j}^2)$ *is finite (Assumption 4), then:*

$$\sqrt{N_{X_j}}(\hat{\tau}_{X_j}^{2SLS} - \tau_{X_j}) \xrightarrow{d} \mathcal{N}(0, N_{X_j} \cdot avar(\hat{\tau}_{X_j}^{2SLS}))$$

as $N_{X_j} \to \infty$*, where* $\xrightarrow{d}$ *denotes convergence in distribution,* $\mathcal{N}(0, N_{X_j} \cdot avar(\hat{\tau}_{X_j}^{2SLS}))$ *stands for the normal distribution, and* $avar(\hat{\tau}_{X_j}^{2SLS})$ *is the asymptotic variance of the 2SLS estimator that can be approximated as in Chapter 15 of Wooldridge (2015).*

Bibliography

Bargagli Stoffi, F. J. & Gnecco, G. (2020), 'Causal tree with instrumental variable: an extension of the causal tree framework to irregular assignment mechanisms', *International Journal of Data Science and Analytics* **9**(3), 315–337.
URL: *https://doi.org/10.1007/s41060-019-00187-z*

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*