

**PAPER TITLE Blah Blah**  
**Estimating Reaction to Welfare**  
**Using Causal Forests**

ECMA 31330 Final Project

Spring 2021

**Max Bronckers**  
**Veronica Song**  
**Dustin Zhang**

# 1 Introduction

There has been an increasing adoption of Machine Learning (ML) methods in economics for algorithmic approaches to causal inference. Whereas initially the ML methods were avoided due to uncertainty in its consistency, normality, and efficiency, major developments in methodology has allowed a stable large-sample confidence interval to be constructed around treatment effect estimates conditional on multiple covariates.[1] One such method is the Bayesian Additive Regression Trees (BART) used by Green and Kern (2012), which allows heterogeneous treatment effects to be modeled flexibly. Using a Markov Chain Monte Carlo (MCMC) algorithm that derives effects from the posterior mean and interval instead of pre-specified tree parameters, BART has a much smoother and adaptive structure than traditional OLS or single tree models popular in economics, and is also resilient to problems with overfitting.[3] However, despite its excellent predictive capacity, BART lacks an asymptotic explanation of the estimates and are thus difficult for causal inference and application. Though recent work by Ročková and Saha (2018) suggests modifications to BART that may allow concentration of the posterior mean around the true mean, the construction of an asymptotic theory of BART estimates is still an ongoing effort. [4]. Another such method is the causal forests proposed by Wager and Athey (2017), which estimates are asymptotically gaussian and unbiased to allow proper confidence intervals to be constructed around the treatment effect. The construction of adequate confidence intervals is especially relevant in policy applications, as consistent estimates can be produced for the treatment group.[2]

In this paper, we compare the aforementioned two popular ML tools in heterogeneous treatment effects estimation among survey responders using an empirical dataset. Specifically, we aim to compare uses of BART versus Causal Forests, using OLS estimates of heterogeneous treatment effects as a benchmark. To measure the stigma around the word "welfare", we take a survey data of individual perception on public spending and find the impact of the question phrasing on the responses. In order to uncover the heterogeneous effects of the question phrasing based on a suite of socioeconomic backgrounds of the respondent, Green and Kern utilizes BART. Given the theoretical shortcomings of BART, we aim to make causal forest estimates and compare it to the authors' findings. The first section of this paper examines the causal forest under different parameters of a synthetic DGP that resembles our empirical dataset. Here, taking our DGP as a ground truth, we aim to assess conditions under which causal forest performs well or poorly and thus evaluate it against our empirical dataset. In the second section, we fit a causal forest model on the empirical data and compare our estimates against BART estimates of conditional average treatment effect (CATE) obtained by Green and Kern.

## 2 Discussion of Data?

## References

- [1] S. Athey and G. Imbens. Machine learning methods economists should know about, 2019.
- [2] S. Athey and S. Wager. Estimating treatment effects with causal forests: An application, 2019.
- [3] D. P. GREEN and H. L. KERN. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *The Public Opinion Quarterly*, 76(3):491–511, 2012.
- [4] V. Rockova and E. Saha. On theory for bart, 2018.