Econ 293 Final Project

Austin Murphy, Fred Lu

June 2020

1 Introduction

The General Social Survey (GSS) aims to monitor social change in the United States of America by gathering data on contemporary American trends, behaviors, and attitudes. Since 1972, it has conducted roughly biannual surveys on various topics ranging from science and health, to politics and discrimination. The GSS data are among the most widely studied United States survey datasets by researchers in the social sciences. In our project, we analyze the Trends in National Spending Priorities Survey which evaluates the attitude of Americans toward government spending on a number of programs. In our analysis, we analyze a modified version of the 'welfare' dataset.

The 'welfare' dataset analyzes the attitude of Americans towards programs that involve the name 'welfare'. In addition, we merged additional data from the same survey to analyze the attitude of Americans towards programs that offer assistance to African Americans. The survey ran a question wording experiment in which respondents were randomly assigned to one of two questions regarding spending for both welfare programs and African American assistance programs. The questions have the same introduction and possible responses, but differ in the name or description of the spending program. The questions and possible responses are given as follows:

> "We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount... Are we spending too much, too little, or about the right amount on (ITEM)"
>
> - Too little
> - About right
> - Too much

In the welfare related question, ITEM is randomly set as "welfare" or "assisting the poor". In the second question we analyzed, ITEM is randomly set as "improving the conditions of blacks" or "assistance to blacks". We analyze the conditional average treatment effect of several methods on these two responses.

2 Methods

2.1 Data

The welfare dataset was downloaded from the the GSB Digital Business Initiative causal dataset GitHub[1]. The response variable in question was downloaded using the GSS Data Explorer tool[2], and was joined with

[1] `https://github.com/gsbDBI/ExperimentData`
[2] `https://gssdataexplorer.norc.org`

the welfare dataset using the year and respondent identifier. We limited the data to the same years as used in Green and Kern [1], limiting to the surveys distributed from 1986-2010, with few exceptions.

The data included demographic information on the respondent, including the age and high education level attained, along with metrics on party identification and political view (both on a 7-point scale). Lastly, a metric attempting to gauge the respondents' attitude towards African Americans was created by the average response to four questions regarding the economic differences between races and their origins. Finally, a treatment variable indicates which wording was presented to the respondent when asked about government assistance to African Americans.

2.2 Models

We investigated two causal relationships. The first is a re-implementation of the 'welfare' question from the same survey, which was studied in Green and Kern [1]. This question essentially measured two outcomes: The first asked whether the government is spending too much money on 'welfare' while the second asked if the government is spending too much money on 'assisting the poor'. Since each respondent received one of the two questions, the treatment effect is measuring the rate of responding 'too much' based on which question was assigned.

The second and primary focus of our analysis is on the 'assistance' question, which is described in the Introduction. The questions are analogous in structure, so we apply the same methods to study both. In particular, we compute heterogeneous treatment effects using three models:

1. Bayesian Additive Regression Trees (BART): This approach was used in the original Green and Karn study to estimate CATEs [1]. This is a nonparametric Bayesian approach using an iterative ensemble method with MCMC sampling.
2. Causal Forest: This approach is based on the Random Forest but specifically developed to estimate conditional treatment effects.
3. X-learner: This is a meta-algorithm that builds on base learners for estimating conditional treatment effects. We use regression trees as base learners.

As baselines we implement the S-learner and T-learner methods.

2.3 Evaluation

As discussed on previous homework, we wish to evaluate the performance of our methods in estimating the true treatment effects. To obtain an estimator of this performance, we use the Transformed Outcome method. That is, we compute the mean square error (MSE) on the test set against y^*, where

$$y_i^* = \frac{W_i - p_i}{p_i(1 - p_i)} \cdot y_i$$

As a second evaluation metric, we compute the R-loss, which gives a more stable (though biased) estimate of the overall MSE than the Transformed Outcomes method. This method minimizes

$$\mathbb{E}[e(X_i)(1 - e(X_i))(\tau(X_i) - \hat{\tau}(X_i))^2]$$

that is, a weighted MSE with more weight to datapoints with good overlap.

We use a random split of 5000 responses (14% of the data) as the holdout set for evaluation.

3 Results

3.1 Welfare question treatment effect

We first check our methods by analyzing the question studied in [1]. Specifically, the question involves the change in support between government spending on 'welfare' versus 'assisting the poor'. For this question, the Average Treatment Effect (ATE) computed over the entire dataset is 0.33. In other words, 33% more respondents who were asked the question with the word 'welfare' as opposed to 'assisting the poor' indicated that the government was spending too much on that item. This suggests that the word 'welfare' has negative connotations across the United States population. This average finding is similar to published results, and has been fairly consistent across years.

Our heterogeneous treatment effects also obtained using BART match fairly closely with previously reported findings. For example, decreasing support for 'welfare', as measured by conditional average treatment effect, is associated with increasing levels of conservatism and negative attitudes toward Blacks, as measured using survey apparatus. Having implemented these results, we then implement the same methods to investigate the causal effects of the assistance question.

To assess whether BART is an appropriate method to use for the survey data, we conduct an evaluation experiment with two alternative powerful methods for estimating heterogeneous treatment effects. The results are displayed in Table 1. BART, Causal Forest, and X-learner perform identically well on both metrics. As a baseline, we implemented the S-learner and T-learner methods for estimating heterogeneous treatment effect. We note that BART, Causal Forest, and X-learner metrics produce far lower MSE as well as lower variance than the baselines. This suggests that our choice of method is appropriate.

	Transformed Outcomes		R-loss	
Method	mean	sd	mean	sd
BART	0.989	1.547	0.159	0.197
Causal Forest	0.991	1.544	0.159	0.195
X-learner	0.990	1.545	0.159	0.196
S-learner	1.627	2.603	0.266	0.344
T-learner	1.208	1.956	0.159	0.196

Table 1: Performance of multiple methods in estimating heterogeneous treatment effects on the 'welfare' question, as measured on a holdout set using two methods: Transformed Outcomes and R-loss

3.2 Race assistance question treatment effect

We find that the average treatment effect over the survey dataset is 0.062. In the welfare question analysis, we observed that BART, Causal Forest, and X-learner perform similarly, while achieving significantly better metrics than S-learner and T-learner (Table 1). However, on the assistance question, all methods are observed to perform more comparably, as seen in Table 2. We still note that BART, Causal Forest, and X-learner perform slightly better than the alternatives. Overall, this change in performance may be attributed to the lack of a strong treatment effect in the dataset, compared to the previous ATE of 0.33. If there is not a strong or consistent pattern of heterogeneous treatment effect, any improved methods will likely not outperform some baseline treatment effect estimate. For example, the treatment effect may be so small as to be homogeneous, or just simply due to noise.

3.3 Analyzing conditional treatment effects by covariate

Using the BART posterior estimates, we can construct conditional estimates for each covariate by fixing the value of that covariate over the entire dataset. Then, by modifying the conditioned value, we can estimate the varying conditional average treatment effects over the domain of the covariate, as shown in Figure 1.

	Transformed Outcomes		R-loss	
Method	mean	sd	mean	sd
BART	0.748	1.547	0.145	0.228
Causal Forest	0.749	0.145	0.159	0.227
X-learner	0.749	1.553	0.145	0.227
S-learner	0.751	1.564	0.149	0.231
T-learner	0.773	1.603	0.145	0.227

Table 2: Performance of multiple methods in estimating heterogeneous treatment effects on the 'assistance' question, as measured on a holdout set using two methods: Transformed Outcomes and R-loss

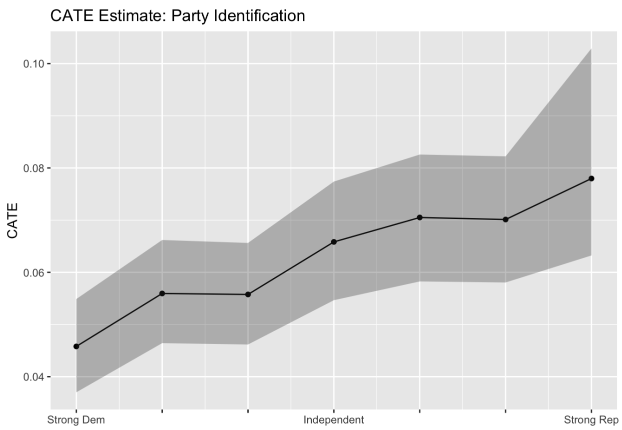
We observe a consistent trend of increasing Republican affiliation with increasing CATE. We can interpret a value of the CATE as the percent of survey responders who think too much money is being spent on "assisting Blacks" but not on "improving the conditions of Blacks". The latter response is worded such that it is less about "giving out money" and thus may be supported by a higher percentage of people. The estimate of the difference in treatment effects for strong Republicans vs strong Democrats is 3.5%. Interestingly, being more politically conservative seems to have less of an effect than party affiliation.

The CATE estimates for both age and the estimate for a negative attitude towards blacks do not have as significant a difference as party identification.

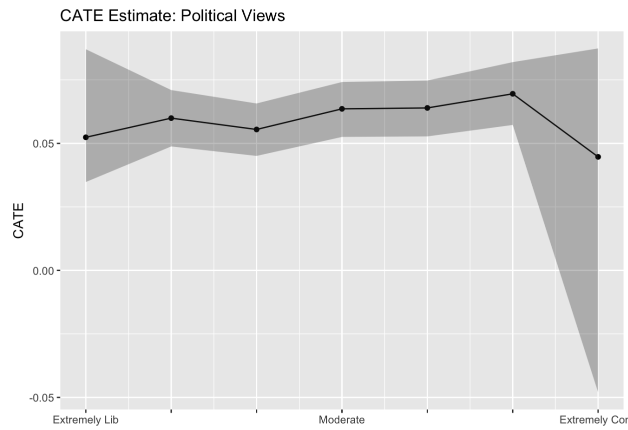
Even though the ATE is very small, we nonetheless observe interesting patterns in the conditional treatment effects due to heterogeneity in the surveyed population. The trends are more pronounced in the analysis in Green and Kern [1], but this analysis of the assisting of blacks offers additional information in the perception of government spending and race relations in the United States.

References

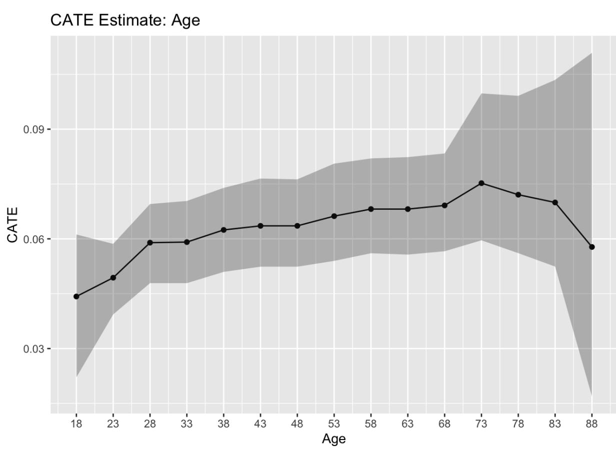
[1] Donald P Green and Holger L Kern. Modeling heterogeneous treatment effects in survey experiments with bayesian additive regression trees. *Public opinion quarterly*, 76(3):491–511, 2012.



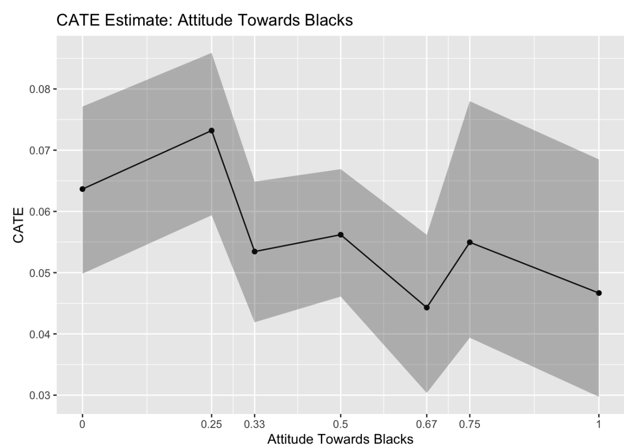
(a) Party Identification



(b) Political View



(c) Age



(d) Attitude towards Blacks

Figure 1: CATE Estimates