Vote by Mail: A study on the effects of voting by mail on election turnout in Utah

Abstract

The paper aims to replicate the 2018 Showalter report on the impact of Voting at Home (VAH) on election turnout in the Utah 2016 general election and extend the analysis to the 2014 midterm elections in Utah. We were able to successfully verify that the VAH outperformed traditional voting counties by 4-10%. We attempted to reverse-engineer the TargetSmart propensity scores for 2016 in order to assign propensity scores to the voters in the 2014 election. We found that VAH positively impacted voter turnout in 2014. This suggests that the increased turnout in the 2016 general elections cannot solely be attributed to the circumstances of that election. The models we built utilized general election information which we then applied to midterm elections. This meant that we had to assume the circumstances between these election types are not significantly different from one another, however, we know this to be false as more people come out to vote for presidential elections than in midterms. Moreover, the 2016 presidential election was a unique election that included many factors that impacted voter turnout.

Keywords: Vote-at-home, random forest, regression, ordinary least squares, boosted forest

1 Introduction

1.1 History of Vote-at-Home

Questions regarding voter turnout have repeatedly come up in political discourse. This makes sense as the U.S. falls behind many modern countries with a 55% turnout, where only 64% of the voting-age population are registered to vote (Desilver n.d.). The U.S. has made a few attempts at increasing voter turnout through federal regulations, including the Voter's Rights Act of 1967 and the National Voter Registration Act of 1993 (Berinsky, Burns, and Traugott 2001).

In the United States, the traditional method of voting involved waiting in line at your local polling place and filling out a paper ballot. But during World War II, a large proportion of eligible voters were overseas. So the United States passed laws providing absentee ballots to soldiers ("Voting by Mail and Absentee Voting" n.d.). Since then states have continued absentee voting or voting by mail to varying extents. But in 1981, the state of Oregon had to conduct a special election, which they carried out by mail. This method of conducting an election entirely with mail-in ballots is called vote-at-home(VAH)[1]. Ballots are delivered to citizens prior to election day, who could then fill out the ballots and return them by mail or in person. Following the success of this election, Oregon continued to use VAH methods to varying degrees until 1998 when Oregon passed an initiative to conduct all further elections by mail (Berinsky, Burns, and Traugott 2001). As of March 2020, 21 states that have some sort of conditional VAH regulation in place and five states that administer all elections by mail (Underhill 2020).

Today some barriers of voting include access to transportation, time off from work, and childcare (Southwell and Burchett 2000). VAH is a promising method of increasing voter turnout because it is thought to decrease some of these costs of voting. However, there are also some criticisms to VAH. Critics argue that VAH would increase voter fraud and voter bias. While VAH is meant to increase accessibility, there is concern about eligible voters failing to receive a ballot, possibly by getting sent to the wrong address (Southwell and Burchett 2000). Some areas, like Native American reservations, lack proper infrastructure for postal services and may be disproportionately impacted. Interpretability of election material and the lack of a home address or P.O. box could also hinder voting (Underhill 2020). There is also a loss of the tradition of experiencing the neighborhood polling place. Mail-in ballots take longer to process which delays the outcome of an election (Underhill 2020).

There are other benefits that result from VAH beyond an increase in turnout. There is a decrease in the need for polling stations, which translates to a decrease in the cost of both staffing and machinery. Some voters find the extra time to research candidates to be helpful and convenient, increasing their satisfaction (Underhill 2020).

1.2 Recent Research

There are mixed conclusions regarding the effect of VAH on voter turnout. Usually, when a state switches to VAH, the entire state makes the switch at one time. This makes it difficult to investigate the effect VAH has on voter turnout because it is difficult to account for other variables that affect turnout. It is even more difficult to distill individual impacts of VAH from aggregate impacts, which don't necessarily account for changes in the electorate (Berinsky, Burns, and Traugott 2001).

When Utah opted to switch to VAH, the entire state did not switch at once. It instead allowed individual counties to decide whether to switch to VAH. This naturally provided a test group of VAH counties, and a control group of non-VAH counties, allowing researchers to compare the two groups and more accurately calculate how VAH affects voter turnout. The Showalter 2018 Report (**???**) chose to analyze Utah for this reason and found that VAH had a positive effect on voter turnout in the 2016 presidential election. The 2016 election was highly unusual, however, and this could have caused flaws in the results of that paper. For this report, the first aim is to replicate the results seen in the Showalter 2018 analysis, and then attempt to extend those results to the 2014 midterm election.

[1]VAH is also referred to as vote-by-mail or all-mail elections

Ordinary least squares regression is a common method used when analyzing voter turnout and we see this method used in several papers.[2] A feasible generalized least squares model was used in Southwell & Burchett (2000) on Oregon elections from 1960-1996. This included fixed effects on election type, election competitiveness, and a time variable. This study showed that voting by mail had a strong impact on turnout (Southwell and Burchett 2000). In the paper "Who Votes by Mail?", Berinsky et al. implemented a 'duration model' to capture the volatility of individual voters and the underlying likelihoods of mobilization and retention. The study was on election data from the state of Oregon. They concluded that although VAH had an overall positive impact on turnout, there was a varying effect on voters and nonvoters. More specifically, VAH was successful in increasing retention rates but had little impact on pulling non-voters out (Berinsky, Burns, and Traugott 2001). This was a suitable choice to apply to election data for the state of Oregon. However, this was not a suitable choice for our analysis due to the proximity of the implementation of VAH and the resulting lack of data.

2 Data and Methods

Section 2.1 introduces the data that we used for the entirety of this analysis. It then looks into the distribution of different demographics in Utah. The methods used in this analysis are discussed in section 2.2, starting with a description of the difference-in-differences analysis. It then discusses the different statistical models used in this report along with how they were used.

2.1 The Data Set

The data set used in this analysis was a 'snapshot' taken of Utah's voter file from 2016, curated by TargetSmart, a data organizing, managing, and analysis company. This voter-file is a combination of consumer information and public voting records and it captured the entire population of registered voters. That data consists of 2,006,786 observations with 256 variables. In our analysis, we utilized 38 of these variables. Of these 38, 6 of these variables had missing data, the worst of which was voter registration date. In 2016, 77% of voters in the data set resided in a VAH county, 15% also resided in a 2014 VAH county, and 8% of residents voted by mail in 2012. There are 29 counties in the state of Utah. It is important to consider the distribution of the population among these counties because it could influence the effect that we see. In 2014, 8 counties had implemented VAH elections and in 2016, 21 counties conducted VAH elections as shown in Figure 1.

2.1.1 Utah Demographics

The state of Utah has a population of 3.2 million people as of July 2019, with 33.6 people per square mile. The median household income is $68,374, with the percentage of people in poverty at 9.0%. Approximately 70.5% of the population is of eligible voting age. Utah's racial distribution is shown in table 1. Those with a high school degree or higher is 92%, followed by 33.3% with a bachelor's degree or higher (**???**). The state of Utah is fairly rural outside of the Great Salt Lake region, predominately white, and of middle-class socioeconomic status. These characteristics make Utah a good candidate to represent Western and Midwestern United States, but not for the United States as a whole. To get a better understanding of how VAH would impact the United States, we would have to consider voters that more accurately represent the race, wealth, and education of the US.

In 2005, the state of Washington permitted counties to switch to VAH. There was a strong rural/urban divide between those that opted to switch and those that did not (**???**). Figure 2 plots the population of each county, colored by the county type. We see an even spread of the traditional voting counties among the VAH counties, which shows the absence of this divide in Utah, illustrating the lack of a relationship between county type and population.

When investigating the distribution of political parties among VAH and non-VAH counties, we see that the bulk of Utah's Democrats live in counties that switched to VAH ballots in 2016. Figure 3 shows that across all three county types, the Republican Party is the most popular. These counties may be more competitive

[2] OLS is used in the analysis of the effect of VAH in Gronke, P. & Toffey, D. (2008), Showalter (2018), Gronke, P., & Miller, P. (2012).

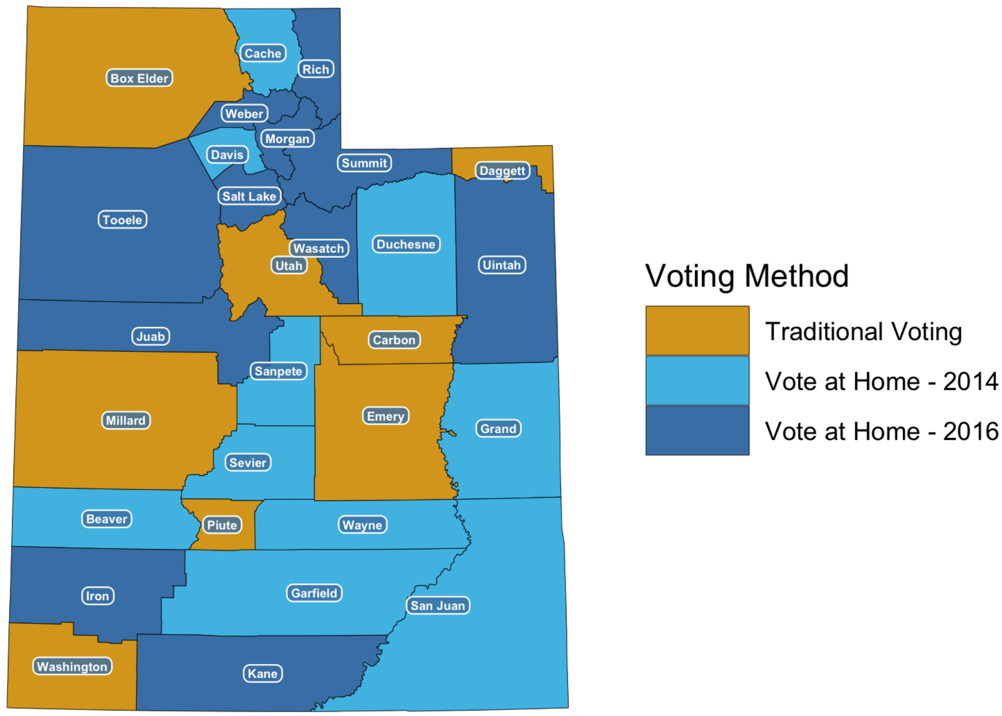

Figure 1: A county map of Utah, where the counties are colord by the election method conducted in each county. The three methods visualized are traditional voting, VAH implemented in 2016, and VAH implemented in 2014.

Racial Distribution in Utah

Race	Percentage of Population
Caucasian	90.7%
Hispanic or Latino	14.2%
Asian	2.7%
Two or More Races	2.6%
Indigenous	1.5%
Black or African American	1.4%
Hawaiian and Pacific Islander	1.1%

Table 1: This table shows the racial distribution of registered voters in Utah in 2016.

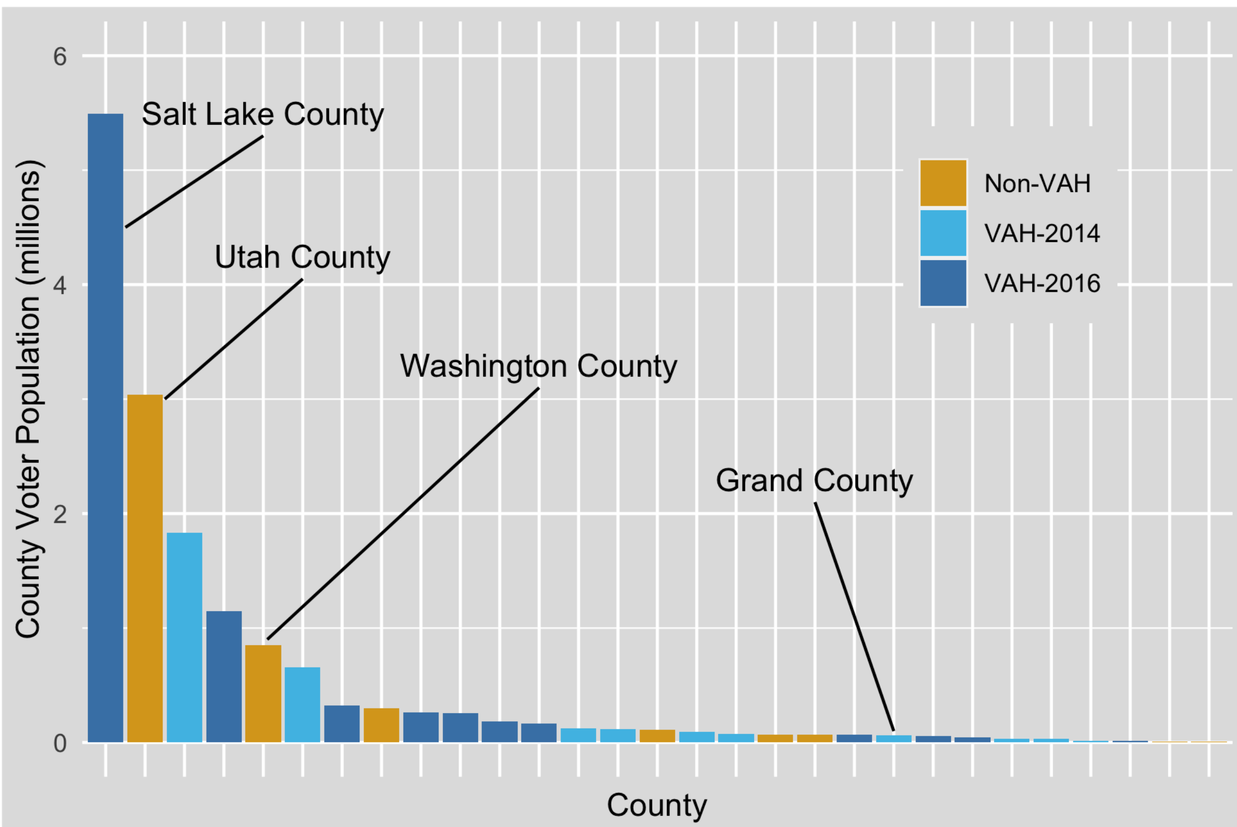

Figure 2: 2016 County population colored with regard to VAH county type. The counties are ordered by population. We can see there is an even distribution of VAH counties across counties of all populations.

during elections. In a state with a Republican majority, voters registered to other parties may feel a more urgent need to vote as compared to Republican voters who might feel safe about the outcome of the elections. This could partially explain the increased turnout, particularly for the 2016 presidential election. However, we believe this should not have had a large effect on our analysis for 2014 because this dynamic would have been present before the introduction of VAH ballots, therefore the propensity models would have been able to account for this effect.

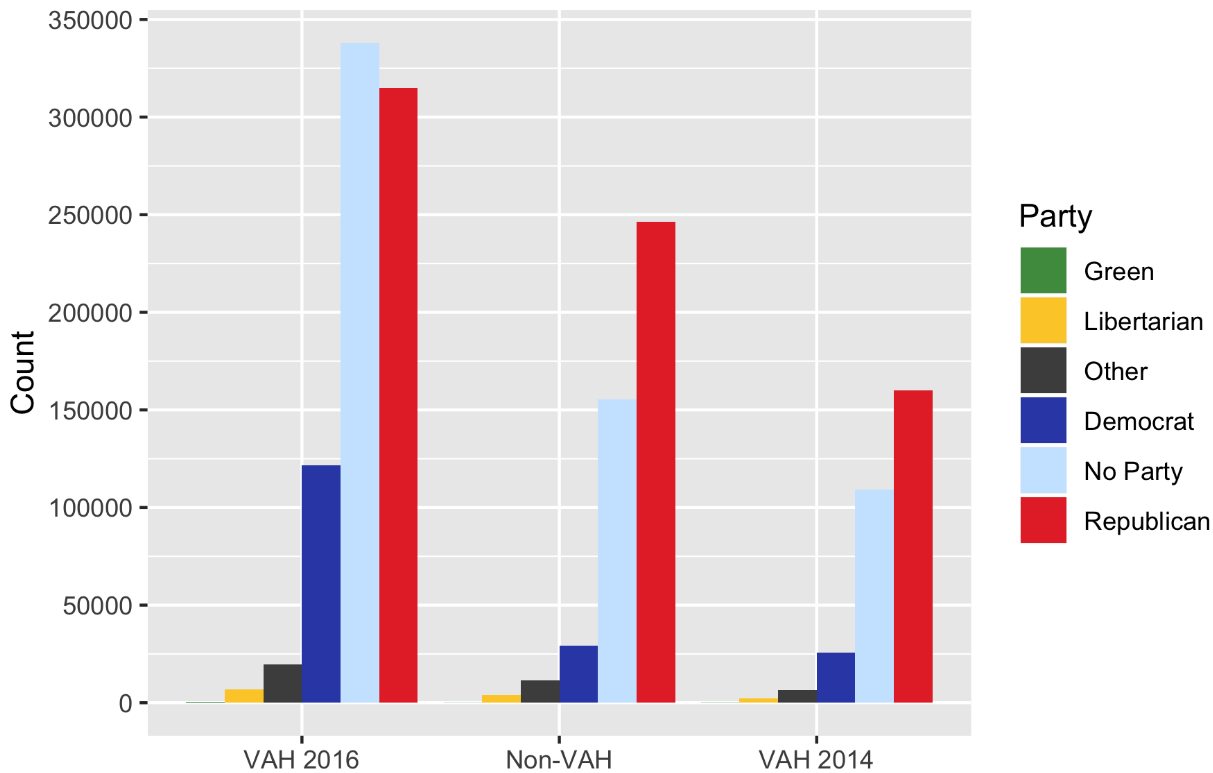

Figure 3: This plot shows the population of each major political party by county type for 2016

The most important variables in our analysis were `uvbm_county`,`binary_vf_g2016`, and `tsmart_general_turnout_score`. The TargetSmart propensity scores were derived using a proprietary model that uses a variety of information to predict the probability that someone will vote. These scores take on values between 0 and 100. While we do not know how TargetSmart's propensity model functions, we presume it to be a multi-level regression that weights each variable differently depending on interactions at each level. The propensity score is useful because it tries to account for all relevant variables excluding the VAH indicator, so the differences between actual and predicted turnout between the two counties can be better attributed to VAH alone.

2.2 Methods

There are three main portions of our analysis. The first compares the difference between predicted and actual turnout between the two county types in 2016. The second portion examines the effect of VAH on voter turnout using regression on the 2016 voter file. Finally, we combine these two methods to analyze the 2014 data set and compare the effects of the relevant variables.

2.2.1 Replication

Our first objective was to verify the results seen in the Showalter 2018 Utah Report. We tested the replicability of the Showalter report in order to ensure that the extrapolated results came from valid data. The analysis relied heavily on TargetSmart's voter propensity scores, which allow us to calculate predicted voter turnout for different groups of voters. These predicted turnouts allow for difference-in-difference analysis to be conducted. It allows us to measure the impacts of VAH on the turnout of various groups of voters. Along with the difference-in-differences analysis, we also replicated the ordinary least squares and logistic regression analysis.

2.2.2 Creating Propensity Scores for the 2014 Voter File

The second objective of this paper is to extend the results of the Showalter 2018 Report to the 2014 midterm election. The 2016 presidential election was largely regarded as unusual, so it is necessary to see that the findings from that election agree with findings from another election in Utah. The 2014 snapshot of Utah's voter file was formatted very similarly to the 2016 voter file, although it did not include propensity scores. TargetSmart provided figure 4 illustrating the relative importance of different variables in their propensity scores model. We attempted to use this plot to reverse engineer their model and produce our own propensity scores. We also generated propensity scores using the random forest and boosted tree models created from the 2016 data. Note that these models are set to predict a categorical variable (whether a person voted in an election or not). In order to generate propensity scores, we convert this categorical variable into a binary variable and have the models treat it like a numerical variable. This leads the models to output values that are between zero and one, which can be interpreted as an estimated probability that a person voted. We used the generated propensity scores in place of the TargetSmart propensity scores in the regression analysis.

2.2.2.1 Weighted Variables Model

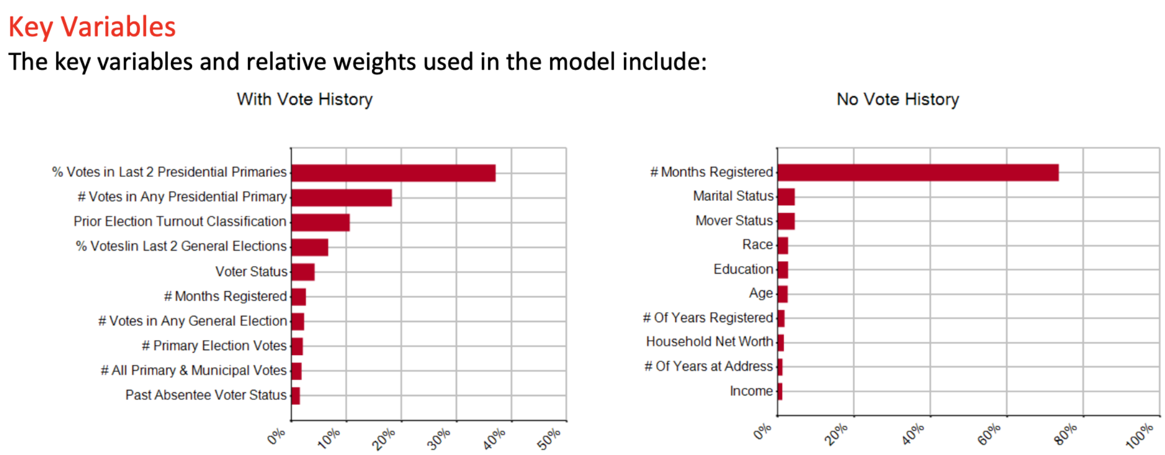

Figure 4: TargetSmart provided a graphic demonstrating the relative weights that they assigned each the strongest variables in their 2014 propensity model.

According to figure 4, TargetSmart produced two propensity models, one that compared the relative weights of variables when modeling data with vote history and another that compared the relative weights of variables when modeling data with no voting history. Both of the proposed propensity models included at least 10 predictive variables.[3] We list the estimated weights of each predictor in table 2.

2.2.2.2 Decision Trees

Alongside the weighted variables model, we used two decision tree ensembles to predict the likelihood that an individual would cast a vote depending on whether or not they lived in a VAH county. One downside of this basic decision tree is that it is 'greedy'. The data is split by how the current split minimizes the RSS, not how future splits will minimize the RSS. This means that the model can make poor decisions on how to split. Because of this greediness, simple regression trees are often not especially good at making predictions.

[3] While 4 only shows 10 variables, it is likely that TargetSmart utilized many more predictors when building the model, but only showcased the 10 'strongest' variables

Relative Variable Weights			
With Vote History		No Vote History	
% Votes in Last 2 Presidential Primaries	0.3723	# Months Registered	0.7347
Votes in Any Presidential Primary	0.1844	Marital Status	0.0459
Prior Election Turnout Classification	0.1064	Mover Status	0.0434
% Votes In Last 2 General Elections	0.0674	Race	0.0281
Voter Status	0.0461	Education	0.0281
# Months Registered	0.0255	Age	0.0255
# Votes in Any General Election	0.0248	# Of Years Registered	0.0179
# Primary Election Votes	0.0213	Household Net Worth	0.0153
# All Primary & Municipal Votes	0.0200	# Of Years at Address	0.0128
Past Absentee Voter Status	0.0102	Income	0.0102

Table 2: The implied variable weights generated from the TargetSmart graphic.

To improve upon the simple tree, two other model types have been created that grow many trees and then aggregate them in order to make more accurate predictions. They are known as the boosted tree and the random forest.

2.2.2.3 Boosted Tree

The idea of boosting is that you fit a decision tree to data and craft a second tree using the residuals of the original tree. You continue this until you have B trees. There are three parameters used in growing a boosted tree. The total number of trees to be grown is B. The shrinkage parameter is denoted as λ, often taking values of 0.01 or 0.001. Lastly, we have d, the number of splits to make in each tree. A boosted tree model sequentially creates decision trees with each following tree 'grown' from information produced from the previous tree.

The result of this process is your boosted regression tree:

$$\hat{f}(x) = \sum_{b=1}^{B} \lambda \hat{f}_b(x).$$

(DeGroot and Schervish 2012)

For our boosted tree model, we set the number of trees $B = 100$, the number of splits $d = 10$, and our shrinkage parameter $\lambda = 0.1$. We randomly split our data into the training data and the test data.

The predictor space consisted of variables that represented voting history, demographics, and indicators on the type of counties. The variables relating to voter history were binary variables that denoted whether a person voted or not in each major election since the year 2000. The model also used each person's date of registration, along with a binary variable that denoted whether or not the person voted using an absentee ballot in 2012. Demographic variables that were used to train this model were age, race, gender, income, and a variable denoting which political party each person was registered as. Variables such as a person's address, county, and whether they lived in a VAH county or not, were not used as predictors for this model so that it would not account for differences between VAH and non-VAH counties when making predictions. This made it possible to assert that any differences between county types were caused by VAH ballots.

2.2.2.4 Random Forest

The other type of decision tree ensemble that was used was a random forest. For our propensity model, we opted to set the size of the forest to 50 trees, with each tree having access to 8 predictors. Each tree was allowed a maximum of 256 nodes. Like the boosted tree, we used half of the data set to train the random forest, and the other half made up our test data. The model was set to predict whether or not a person voted, and the predictors included the same variables that related to voter history, along with demographic information. Because the model produced outputs that minimized error, the predictions ranged between zero and one, which denoted the probability of each person voting.

To assess these propensity models, we calculated the mean squared error (MSE) of each model's predictions, using the 2016 voter file. To do this, we first found the difference between each person's turnout score and the binary variable for whether or not they voted in 2016. This difference was then squared, and the mean of these squared differences was calculated for each model type. A smaller MSE means that the model more accurately predicted whether a person would vote or not.

Comparing MSEs allowed us to compare the performance of our models to that of TargetSmart's propensity scores. The weighted variables model used predictors that were only available in the 2014 voter file, meaning it could not be fit to the 2016 voter file, so it was not used in this comparison.

After evaluating these models, we had to use one to assign our own propensity scores to the 2014 voter file. Earlier, we split the voter file into test and training data sets. This meant half of the people in the voter file could not be used for further analysis, as they were used to train the model. In order to fix this issue, we implemented k-fold cross-validation. This method creates k estimates of the MSE by splitting the data into k groups. iteratively training the model on $(k-1)$ groups and testing it with the last group, altering which group you leave out until all groups have been left out. You then take the average of the MSEs. This will produce a more accurate estimate of the error (James et al. 2013).

We used 5-fold cross-validation to generate the general turnout scores for both the 2016 and 2014 voter files. Once propensity scores were created for everyone in the 2014 voter file, the analysis was completed in largely the same manner as the 2016 analysis. For the 2014 data set, there were no binary columns denoting whether a person had voted in each election, but there were categorical variables denoting the method by which each person voted, if they voted. New binary variables were calculated that represented purely whether or not each person in the file voted so that we could work with the same variables as the 2016 file.

3 Results

Section 3.1 discusses our results from attempting to replicate the Showalter 2018 Report. Section 3.2 compares the different propensity models that we created and evaluates their ability to assign turnout scores. The 2014 results, using scores generated by our model, are shown in section 3.3. Sections 3.4 and 3.5 contain a discussion and conclusion.

3.1 2016 Replication Results

In the 2016 presidential election, we see a significant increase in voter turnout in counties that conducted vote at home elections as compared to counties that used traditional methods.

We compared the turnout performance from those in VAH and non-VAH counties by plotting the predicted turnout against the actual proportion of people that voted. In Figure 5, we can see that VAH counties outperformed non-VAH counties, especially with regard to low propensity voters. Figure 6 illustrates that counties that switched to VAH in 2016 outperformed counties that made the switch in 2014. We consider this to be evidence of a 'burn-in' effect with TargetSmart's propensity model. The propensity scores for voters who lived in VAH 2014 counties would have taken into account an increase in turnout due to the previous implementation of VAH. Because voter history plays a large part in building propensity scores, any increase in participation would have resulted in an increased propensity score for following elections. This disguises some of the effects that VAH may have had in voter turnout. Although the increase in the 2014 counties had been partially accounted for in the propensity scores, Figure 6 demonstrates that those counties still out-perform non-VAH counties. This suggests a continued effect of the change to VAH from 2014 to 2016.

3.1.1 Replication Analysis

The overall difference-in-difference between VAH and non-VAH counties in 2016 was found to be 6.3%, which differs slightly from the 7.0% found in the Showalter 2018 Report. The difference-in-differences analysis was

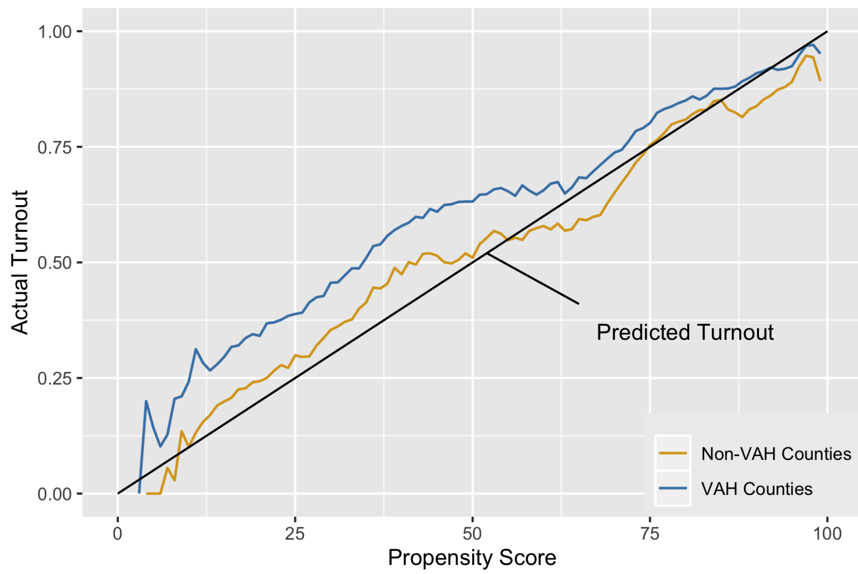

Figure 5: This graphic plots predicted voter turnout against actual voter turnout for the 2016 presidential election. The black line indicates what we would see if the actual turnout was exactly equal to predicted turnout. The VAH counties line is above the line for non-VAH counties which implies that voters in VAH counties outperformed voters in non-VAH counties. This is especially true for lower propensity voters, where the difference between county types is maximized.

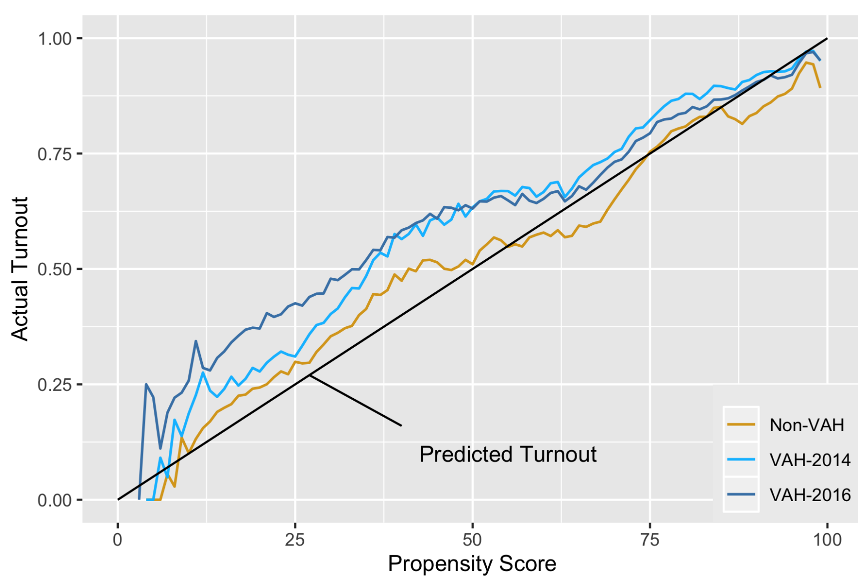

Figure 6: This plot calculates the difference between actual and predicted turnout for registered voters in the 2016 election. This time we separate the voters into three county types.

2016 OLS Regression

	Coefficients		
VAH County	0.06709 (0.0007005)	0.0498 (0.006771)	0.05124 (0.006770)
TargetSmart Turnout Score	0.007816 (0.00001288)	0.007802 (0.00001289)	0.007748 (0.00001317)
County Level Fixed Effects	Not Included	Included	Included
Voted Absentee in 2012	Not Included	Not Included	0.02154 (0.001080)

Table 3: The values indicated in the columns represent the effect size of each variable used in the ordinary least squares regression model. Each column represents a separate model. The standard errors for each estimate are given in parentheses below each coefficient.

2016 Logistic Regression

	Coefficients	Odds	Coefficients	Odds
VAH County	0.4108 (0.004376)	1.51	0.4017 (0.004381)	1.49
TargetSmart Turnout Score	0.0423 (0.000085)	1.04	0.0417 (0.00008632)	1.04
Voted Absentee in 2012	Not Included	NA	0.3111 (0.008433)	1.36

Table 4: The estimates for each coefficient and their odds are given in the columns above. The standard error for each coefficient estimate is given in parentheses.

also conducted across various demographics, to see if specific groups of people had a stronger response to vote at home ballots.

We ran three iterations of ordinary least squares regression to find an estimate of the effect size for each variable. For these tables, the value at the top of each cell is the estimate of the coefficient, and the number in parentheses at the bottom is the standard error. As seen in Table 3, the estimate for the VAH coefficient in the first model was 0.06709, which differs from the Showalter 2018 Report by 0.0064. The parameter estimate for the turnout score was 0.007816, which differed from the original analysis by 0.000083. This difference is not significant, moreover, the results are practically equivalent. The second model used county-level fixed effects to account for any inter-county correlation that might inflate the turnout difference. The third model added a variable indicating whether the individual voted by mail in 2012, to account for the increase in turnout being 'baked in' to the new turnout score. All three of these models resulted in parameter estimates that are within 0.01 of the numbers reported in the Showalter analysis.

We also ran two logistic regression models. The estimates for each coefficient can be seen in Table 4. The odds that an individual in a VAH county would participate in an election was 1.51 as compared to the TargetSmart propensity score odds of 1.04 in both models. When we account for the voting method in 2012, the odds of voting in a VAH county decrease to 1.49, with the odds of voting if the individual voted by mail in 2012 being 1.36. This shows a substantial increase in the likelihood of an individual voting if they reside in a VAH county.

3.2 Propensity Models Results

Once we finished verifying the results of the Showalter 2018 report, we attempted to create propensity scores for the 2014 voter file. After fitting the weighted variables model, the boosted tree, and the random forest, we created propensity plots for each model. This allowed us to compare the models and determine which method generated the most accurate propensity scores when compared to actual turnout. Figure 7 shows the propensity plot with each model type for 2016. In this plot, the random forest appears to be best at

Model Type	MSE
TargetSmart Turnout Score	0.2072826
Boosted Tree	0.1681711
Random Forest	0.1365293

2016 Model Comparison

Table 5: This table presents the accuracy of the TargetSmart, boosted tree, and random forest models for the 2016 election. We determine accuracy by comparing the mean squared error of the model when compared to the actual turnout. A lower MSE implies the model was more accurate.

generating propensity scores. Figure 8 shows the propensity plot with each model type for the 2014 data. Again, the random forest appears to outdo the other two models. The model types are separated by year because we only have TargetSmart propensity scores for 2016. Additionally, the weighted variables model can only be generated for the 2014 voter file because some of the variables that were utilized were not available in the 2016 voter file.

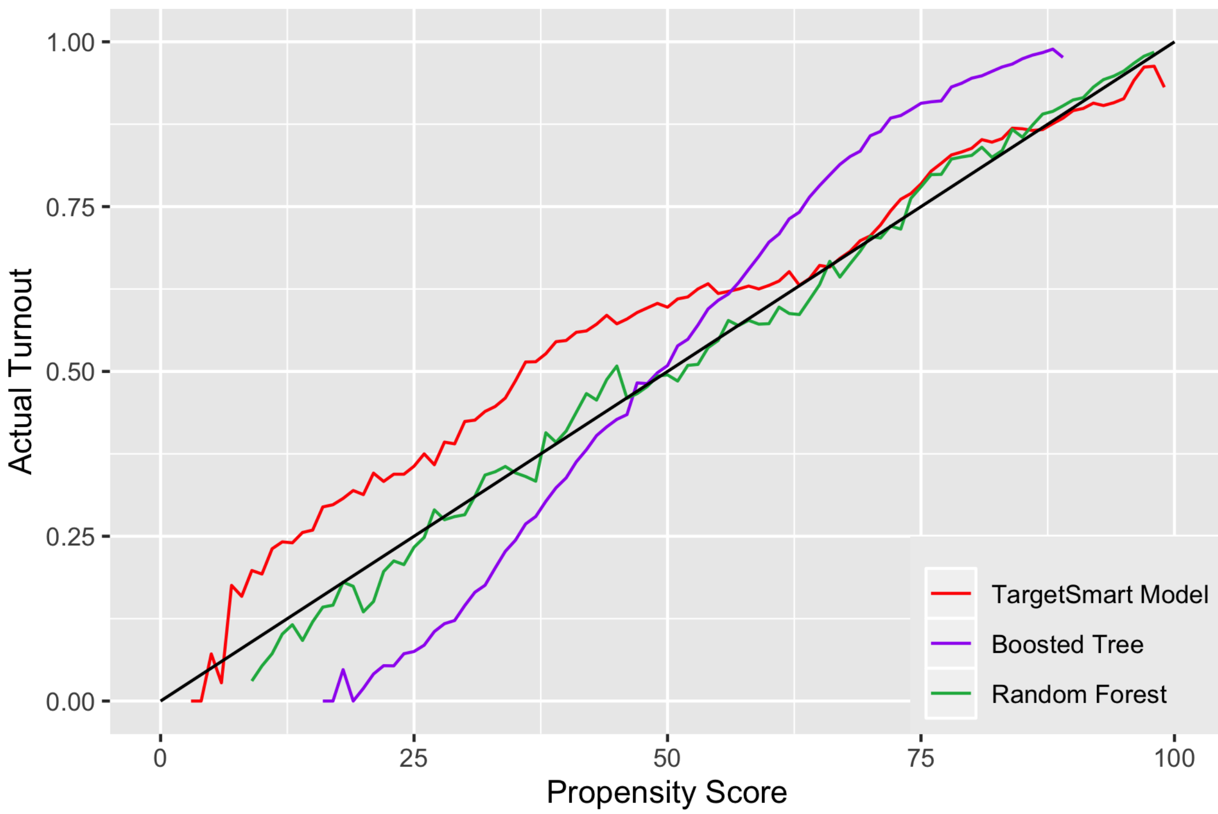

Figure 7: Comparing 2016 propensity models: The random forest model provides the most accurate prediction of voter turnout.

After this, the MSEs were calculated for each model. In Table 5, we can see that the random forest model had the lowest MSE, meaning it performed better than the boosted tree model and TargetSmart's propensity model. It is worth noting that TargetSmart's propensity model likely used training data from previous years instead of the 2016 voter file. This would have made it more difficult to accurately predict whether a person would vote or not. In order to compare the boosted tree model and the random forest to the weighted variables model, we calculated each model's MSE when fit to the 2014 data set, as seen in Table 6. Again, the random forest model outperformed both of the other two models.

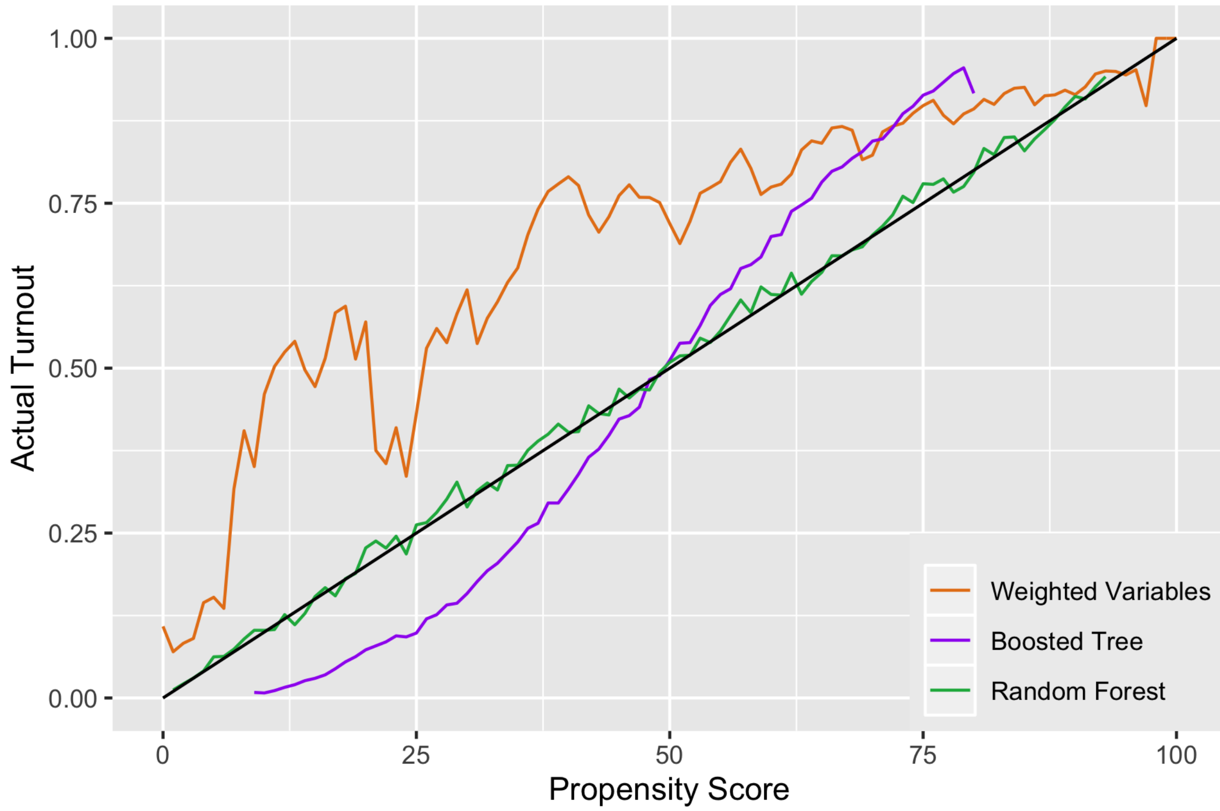

Figure 8: Comparing 2014 propensity models: Like the 2016 models, the random forest again provides the most accurate prediction of turnout.

2014 Model Comparison

Model Type	MSE
Weighted Variables Model	0.2068517
Boosted Tree	0.1557679
Random Forest	0.1435021

Table 6: This table compares the accuracy of the weighted variables model, the boosted tree, and the random forest models for the 2014 election. Like the 2016 models, the random forest had the smallest MSE.

The greatest variability occurs for low propensity voters. The scores are generally more accurate for high propensity voters. The TargetSmart scores and the analogous weighted variables model that was built from their variable importance figure both underestimated voter turnout. The boosted tree models did not generate scores less than 10 and greater than 90. They also overestimated turnout for low propensity voters and underestimated turnout for high propensity voters. This may be a product of the boosted trees being under-fit to the data. Each prediction is the result of finding the average of all points at a particular node of the tree. If there were fewer nodes in the model, the average number of data points per node would be greater, so predictions may be drawn closer to 50.

We chose to use the random forest model to generate 2014 propensity scores because it had the smallest MSE and appeared to be most accurate according to Figures 7 and 8. Before doing this, we used the random forest model to generate 2016 propensity scores. This allowed us to see if scores generated from the random forest model would have yielded similar conclusions to the ones using TargetSmart's turnout scores.

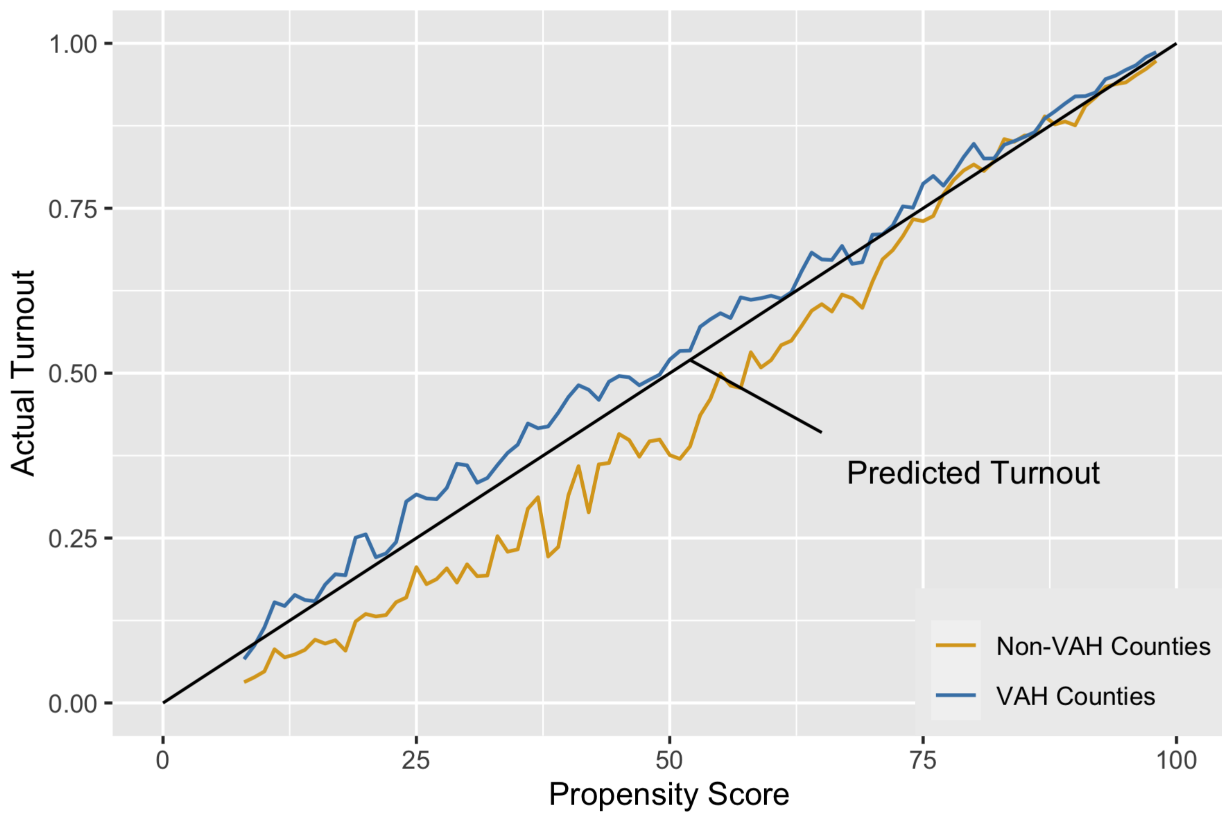

Figure 9: Propensity plot for the 2016 voter file using turnout scores generated by the random forest model.

Figure 9 shows the propensity plot for 2016 using the random forest turnout scores. We can see that VAH counties still outperformed non-VAH counties to a similar degree as when we used TargetSmart's turnout scores (Figure 5). The similarities between Figures 5 and 9 suggest that our findings for the 2014 analysis using random forest scores would agree with the results we would have had if we had access to TargetSmart's propensity scores for 2014.

3.3 2014 Results

Now that propensity scores had been created for the 2014 voter file, we could analyze the 2014 voter file using the same methods that were used on the 2016 voter file. Figure 10 plots the predicted turnout against the actual voter turnout in 2014.

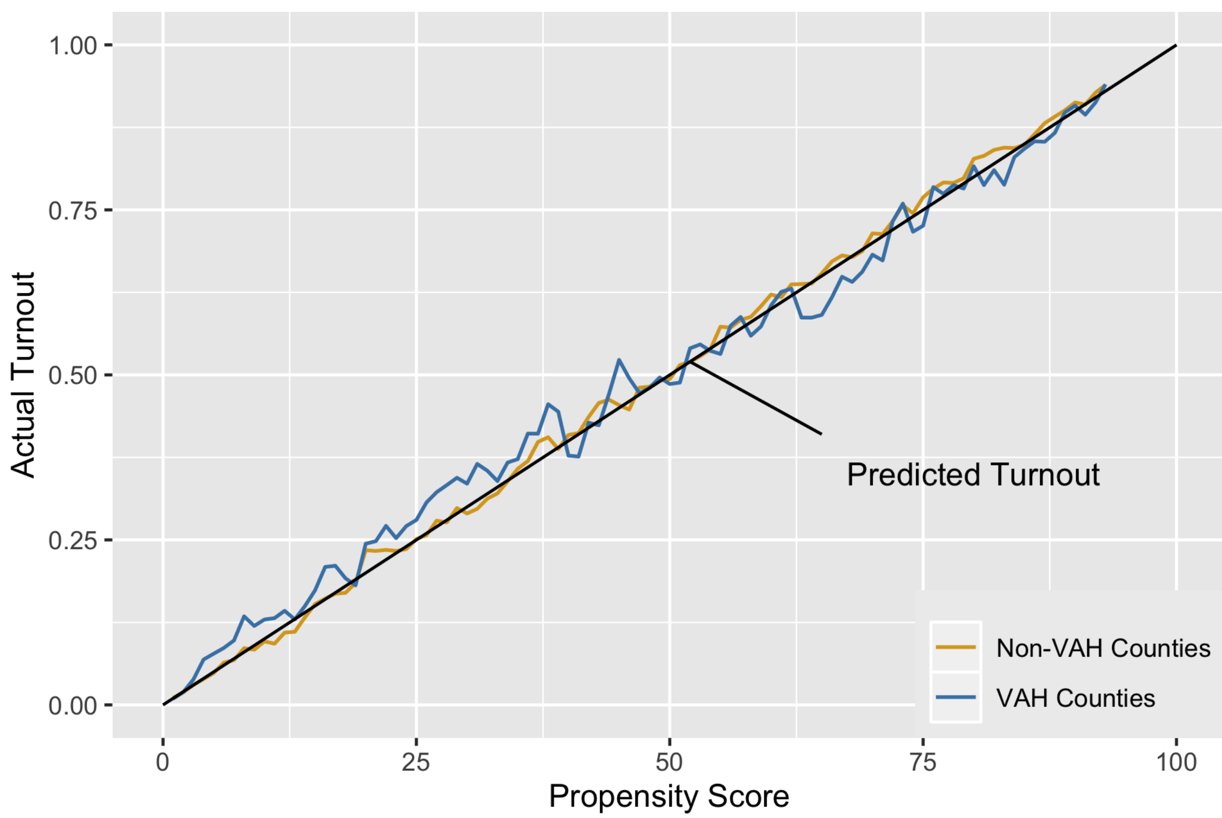

Figure 10: This figure plots predicted voter turnout against actual voter turnout for different county types in the 2014 midterm election. Low-propensity voters in VAH counties outperformed those in non-VAH counties, but by a smaller margin than in 2016.

The difference between VAH and non-VAH counties is much less obvious in this plot than in 2016, although it is still present, specifically with respect to low propensity voters.

3.3.1 2014 Difference-in-Difference Analysis

The overall difference-in-difference between VAH and non-VAH counties in 2014 was 0.38%, which is notably smaller than the difference-in-difference seen in 2016. Next, we investigate difference-in-differences for various demographics. This allows us to see how these demographics responded differently to VAH. A larger difference-in-difference means that voters of that demographic had a stronger response to VAH ballots.

3.3.1.1 Age

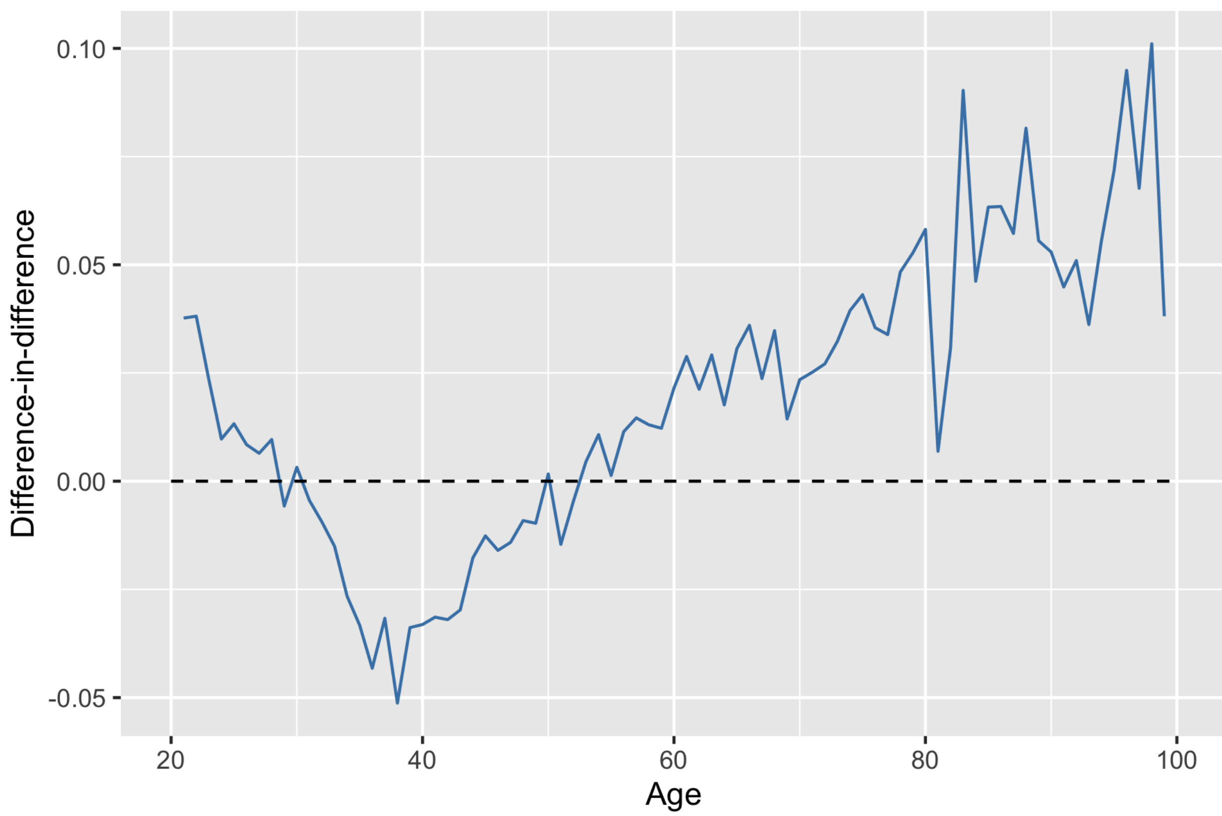

Figure 11: This figure illustrates how the difference-in-difference between VAH and non-VAH counties changed with respect to age in the 2014 election. Older voters have a stronger positive response to VAH than younger voters.

In figure 11, we see that very young voters had a strong positive response to VAH ballots. The size of this response decreases with age until about 38, where voters seemed to actually have a negative response to VAH ballots. After age 38, the response to VAH ballots increases again, with elderly voters having the strongest response to the new ballot type.

3.3.1.2 Gender

Table 7 indicates that female voters had almost a 5x stronger response to VAH than male voters. This does not agree with what was observed in 2016, where we found both genders to have very similar responses. #### Race

Difference-in-differences for each race are shown in Table 8. Native American voters had the strongest

Difference-in-difference by Gender in 2014	
Female	0.0055098742
Male	0.0009873571
Unknown	0.0233101675

Table 7: This table shows the difference-in-difference for each gender listed in the 2014 voter file. Female voters seem to have had a much stronger response to VAH than male voters.

Difference-in-difference by Race	
Caucasian	0.002428829
Asian	0.040478093
Hispanic	-0.009663941
African-American	0.047601594
Native American	0.110518597

Table 8: This table shows the difference-in-difference for each race listed in the 2014 voter file.

response to VAH, with a difference-in-difference of 11.1%. Hispanic and Caucasian voters had very small responses to VAH.

3.3.1.3 Income

Figure 12 shows that for the most part, the voter's response to VAH is not affected by income. After around $125,000, this plot becomes noisy and it is believed that this is due to the small sample size of voters earning that much money per year.

3.3.1.4 Political Party

Table 9 shows that democrats may have had a negative response to VAH ballots, while Republicans had a difference-in-difference of 0.6%. For both parties, this difference-in-difference is quite small.

3.3.2 2014 Regression Results

To calculate the effect size of VAH ballots, we fit an ordinary least squares model, as well as a logistic regression model, to the data set.

Table 10 shows the coefficients for the different OLS models used for the 2014 voter file. For the first OLS model, the VAH coefficient came out to be 0.003647. This is much smaller than the coefficient from 2016. For the OLS models that accounted for possible confounding variables, the VAH coefficient came out to a much larger value at around 0.03. These coefficients are smaller but very comparable to the values of 0.0498 and 0.0512 seen in 2016. Across all three OLS models, the coefficient for VAH came out to be greater than zero, supporting the claim that VAH ballots have a positive effect on voter turnout.

Table 11 shows the resulting coefficients and log odds for the different logistic regression models used in the 2014 analysis. The coefficients for VAH in each model came out to be 0.032 and 0.043. While still greater than zero, these values are around one-tenth of the magnitude of the coefficients seen in the 2016 logistic regression models, which were 0.412 and 0.402, respectively. These OLS and logistic regression models rely mainly on the random forest propensity scores, as opposed to a variety of other variables. We assume the

Difference-in-difference by Party	
Democrat	-0.0024246157
Republican	0.0065518206
No Party	0.0010682965

Table 9: This table shows the difference-in-differences for the two main parties in 2014.

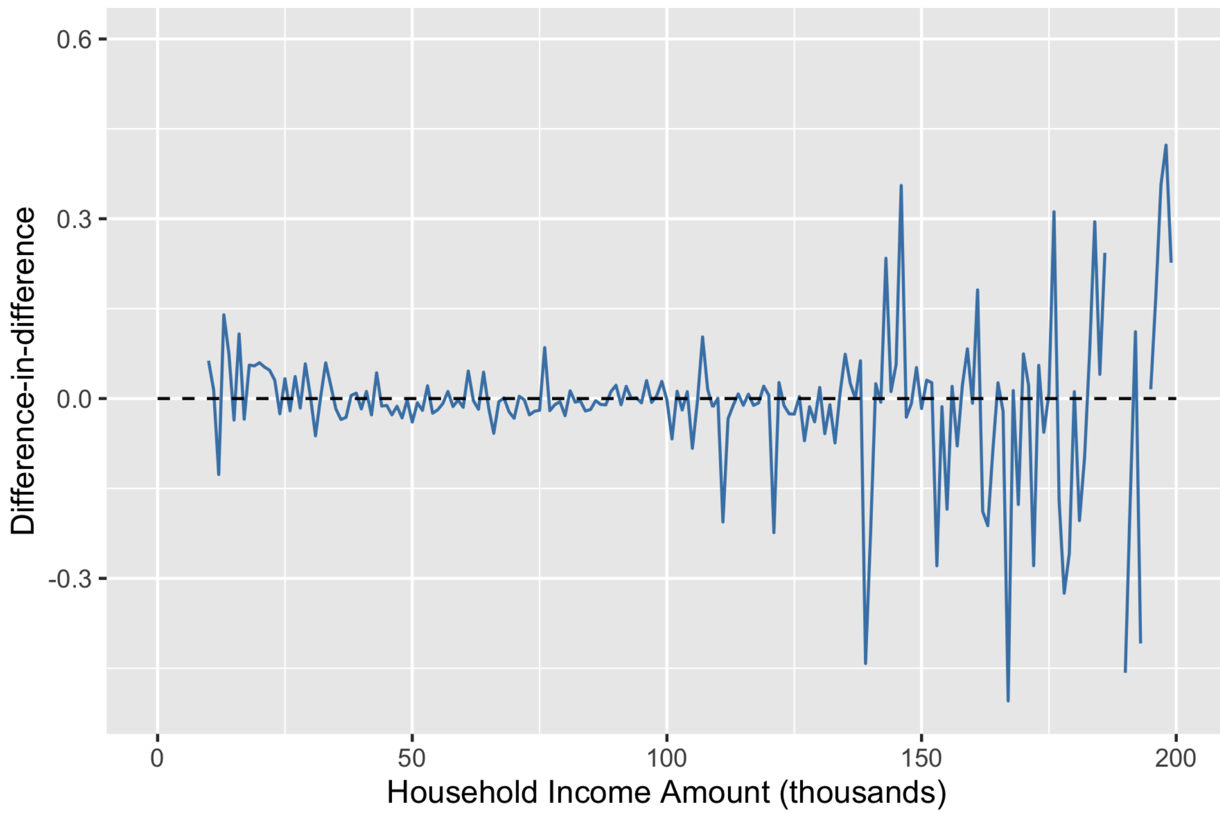

Figure 12: This plot illustrates how the difference-in-difference between county types changes with respect to voter income.

2014 OLS Regression

	Coefficients		
VAH County	0.003647 (0.0007918)	0.03205 (0.006505)	0.03411 (0.006499)
Random Forest Turnout Score	0.01009 (0.00001037)	0.01008 (0.00001036)	0.009930 (0.00001079)
County Level Fixed Effects	Not Included	Included	Included
Voted Absentee in 2012	Not Included	Not Included	0.05188 (0.001020)

Table 10: The relative effect size of each variable determined by ordinary least squares regression for the 2014 election.

2014 Logistic Regression

	Coefficients	Odds	Coefficients	Odds
VAH County	0.03202 (0.005536)	1.03	0.04303 (0.005535)	1.04
Random Forest Turnout Score	0.05399 (0.00008313)	1.06	0.05303 (0.00008435)	1.05
Voted Absentee in 2012	Not Included	NA	0.3954 (0.006569)	1.48

Table 11: This model compares the coefficients and log odds of each variable used in the logistic regression models for the 2014 election.

random forest model took relevant variables into account when generating turnout scores, so there was no need to include the covariates directly.

4 Conclusion

4.1 Discussion

In this paper, we were able to determine that VAH ballots increased voter turnout in the 2016 presidential election, by using TargetSmart's propensity scores. We were also able to evaluate the effect VAH had on different demographics. In particular, we found that voter response to VAH decreases with age until about 75, where it begins to rapidly increase again. We also found that political parties did not seem to influence voters' responses to VAH. We were able to build propensity scores for voters in the 2014 election. This allowed us to extend our findings to a second election, providing stronger evidence that VAH ballots increase voter turnout.

Our results for the 2016 presidential election agree very strongly with the results of the Showalter 2018 Report. There were only minor differences between our findings and that of the original report, and these differences are believed to simply be the result of minor differences in data wrangling. In 2014 we find an increase in voter turnout as a result of VAH. However, the influence on turnout is smaller in 2014 than in 2016. This difference could be the result of a difference between turnout for midterm and presidential elections. It could stem from the smaller amount of VAH counties in 2014.

For the propensity scores that were generated by the random forest, the training data came from the same year as the output, meaning that the training would have included people that voted due to VAH. While the model did not have access to which county each voter was in, it did have access to other variables that may have suggested which county a person lived in. This would have led the model to account for the difference between county types, which would have led us to underestimate the effect size of VAH ballots. Also, for 2016, both the TargetSmart and random forest models were able to see whether or not a person voted in 2014. This would have led to higher propensity scores for people in VAH-2014 counties, which would have led to an underestimation of the effect size of VAH ballots.

One potential solution to these problems posed by A. Showalter would be to use voter files from 2010 and 2012 (both before the introduction of VAH) to generate propensity scores for 2016, and use voter files from 2008 and 2010 to generate scores for 2014. The training data would predate the introduction of VAH ballots so the generated scores would not be under any influence of VAH ballots. These propensity scores would likely be less accurate due to the large time span between the training data and the data receiving scores. However, the lack of influence from VAH may lead to a more accurate estimate of the effect size of VAH ballots.

In order to further strengthen our findings, this sort of analysis should be conducted on other states as they switch to VAH. This would also allow for a more reliable investigation into different demographics and their response to VAH. Only a very small amount of Utah's voters are registered as racial minorities, which limited our ability to find differences in response due to small sample size.

Another area to build off of this paper would be to conduct further analysis in Utah in order to determine if the VAH effect size diminishes or remains constant. When Oregon switched to VAH, they initially enjoyed an increase in turnout. Over time, this increase diminished and there was no noticeable effect in all but special elections (Gronke and Miller 2012). Further analysis in Utah could provide evidence for or against the existence of this novelty effect.

4.2 Implications

Strong voter turnout is paramount to having elected officials that represent the opinions of the majority of the population. The Showalter 2018 Report claimed that the introduction of VAH ballots increased voter turnout in Utah for the 2016 presidential election. Our findings agreed with this result, and extended it to the 2014 midterm election, further strengthening the claim that VAH increases voter turnout. Given this evidence, we hope that policymakers in other states may choose to adopt VAH ballots in order to increase voter turnout in future elections.

Bibliography

Berinsky, A. J., N. Burns, and M. W. Traugott. 2001. "Who Votes by Mail?: A Dynamic Model of the Individual-Level Consequences of Voting-by-Mail Systems." *Public Opinion Quarterly* 65 (2): 178–97. https://doi.org/10.1086/322196.

DeGroot, Morris, and Mark Schervish. 2012. *Probability and Statistics, 4th Edition.* 4th ed. Pearson.

Desilver, Drew. n.d. "U.S. Trails Most Developed Countries in Voter Turnout." *Pew Research Center.* Accessed April 29, 2020. https://www.pewresearch.org/fact-tank/2018/05/21/u-s-voter-turnout-trails-most-developed-countries/.

Gronke, Paul, and Peter Miller. 2012. "Voting by Mail and Turnout in Oregon: Revisiting Southwell and Burchett." *American Politics Research* 40 (6): 976–97. https://doi.org/10.1177/1532673X12457809.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R.* New York: Springer.

Southwell, Priscilla L., and Justin I. Burchett. 2000. "The Effect of All-Mail Elections on Voter Turnout." *American Politics Quarterly* 28 (1): 72–79. https://doi.org/10.1177/1532673X00028001004.

Underhill, Wendy. 2020. "All-Mail Elections (Aka Vote-by-Mail)." *National Conference of State Legislatures.* https://www.ncsl.org/research/elections-and-campaigns/all-mail-elections.aspx.

"Voting by Mail and Absentee Voting." n.d. *MIT Election Data + Science Lab.* Accessed April 29, 2020. https://electionlab.mit.edu/research/voting-mail-and-absentee-voting.