

# **Voting Habits**

**And What Missing an Election Tells Us about Them**

**Benjamin A. Crumpler**

**Columbia University**

**Quantitative Methods in the Social Sciences**

**May 2013**

## **Abstract**

“Who votes and why?” has been an enduring question for those working in politics and political science. One answer to the question is that voting is a habit. In 2003, making the case for voting habits, Gerber, Green, and Schafer showed that voting in one election causes individuals to be more likely to vote in the next. However, a habit involves more than previous actions’ causing the next action. A habit is formed by repeatedly performing the same behavior. However, once a habit is formed the habit itself and not previous actions causes the next. Therefore, for the habitual voter, the causal connection between one election and the next should disappear. In this thesis, I used logistic regression to examine the voting history of all voters in Tennessee and found that voting behavior has the nature of a habit. In the process, I found evidence of distinctive kinds of voting habits.

## 1. Introduction

As usual, a swarm of post-election analyses popped up in the news right after the 2012 election. These post-mortem analyses attempted to explain what happened in each campaign and how it came to be that Obama won while Romney lost. While the reports pointed to different tactics and pivotal moments in the campaign, one of the more interesting comments was about the divergence in polling results from the two campaigns. For both campaigns, the results from internal polling showed their candidate winning. While it is not surprising that internal polls were biased toward the favored candidate, it is surprising that they diverged enough to generate several news stories. The reason for the divergence revolved around "whether Newhouse's internal forecasts and assumptions about the composition of the 2012 electorate are correct versus the ones made by the Obama campaign" (Haberman & Shultheis, 2012). In other words, the Obama and Romney campaigns had different expectations about who would vote in the election. While the answer to the question "Who votes?" would seem to have been settled, as the diverging polls show, the question still seems up for debate.

"Voting is a habit" is a partial answer to the question "Who votes?" An individual goes and votes simply because he or she has been going and voting in the past. The idea of a voting habit has been around for a long time, at least since 1972, where Verba and Nigh (1972) assumed that voting was a habit. It more than likely goes back farther, since persistence in voting patterns was observed as early as 1960 (Campbell, Converse, Miller & Stokes). Despite early assumptions and observations, the idea of voting as a habit has been the object of research only in this century, starting with attempts to identify a causal connection between one election and the next (Cutts, Fieldhouse & John, 2009; Gerber, Green & Schachar, 2003; Green & Schachar, 2000; Fowler, 2006), and then to identify what characteristics a habit of voting might have

(Aldrich, Montgomery & Wood, 2011; Denny & Doyle, 2009; Franklen, 2004; Plutzer, 2002).

In this thesis, I will follow in the footsteps of these later researchers and attempt to identify another trait of voting habits. In particular, I will be look at how missing an election differs for an individual who is a habitual voter, as opposed to an individual who is just starting to vote. An example may be best. Suppose John is a habitual voter. John votes in every election possible; presidential, midterm, primary, and municipal election--he votes in them all. Joe is a new voter, and he has just registered. The presidential election comes up; and for one reason or another, John, and Joe fail to vote. We know that failing to vote decreases the likelihood that an individual will vote again in the future; however, the effect of not voting will more than likely be different for John and Joe.

In this thesis, I attempt to find out if the effect that voting in one election has on the next differs according to an individual's voting habits. I do this by examining the records of the Tennessee voter file. The voter file is a list of all registered voters in the state and contains their voting history and a record of whether or not they voted in any given election.

This thesis is broken up into seven parts. After the introduction is the Literature Review, in which I briefly cover work in political science about voting behavior, with a particular focus is on voting as a habit. In the Literature Review, I also look at work on what a habit is, how habits are formed, and what happens to the habit when an individual fails to perform the habitual act. After the Literature Review, I formally state the hypothesis in light of previous work. In the section devoted to data, I describe the source and formatting of the data. I briefly describe the data itself, as well as the strengths and weakness of using a voter file. In the methodology section, I describe the regression analysis I use; and in the results section, I show the evidence of

habit discovered in the voter file, as well as interesting discoveries I made about voting habits. Finally, I present my conclusions.

## **2. Literature Review**

### **2.1 An Overview of Voting Behavior**

For quite some time, political scientists have been trying to figure out why people vote or do not vote.. Two major observations about individuals and their voting habits seem to be the key to answering the question, “Why do people vote?” The first is that turnout in the United States is not uniform across various demographic groups. Whites turn out in higher numbers than Blacks; females vote more often than males (Gosnell, 1927); people of different ages vote at different rates (Miller, 1996); and partisans vote more often than non-partisans (Campbell et al., 1960). In fact, most factors related to which candidate a person votes for also relate to whether the person votes at all (Aldrich et al., 2011). However, \

those factors are less strongly connected to whether or not the person votes. The second observation is that voters are amazingly consistent. Voters tend to vote (or not) in the same pattern throughout their lives (Campbell et al., 1960).

There are at least two ways to view an election, and these views lead to two general explanations for why people vote. The first way to see an election is as an independent event, in which individuals, based on some criteria, choose to vote or not (Gerber et al., 2003). This basic understanding leads to many theories about voter choice, mainly focusing on the different criteria people use when they make a choice to vote. Some authors examine the psychological state of the potential voter, and show that an individual's sense of civic duty (Verba, 1972), partisan fervor (Campbell et al., 1960), or sense of efficacy (Finkel, 1985) affect whether or not the

individual votes. Others frame the choice as a cost-benefit analysis and argue that lowering the cost of voting—whether by easing voter registration laws (Highton, 1997) or by making election day less hectic (Brody & Sniderman, 1977)—or raising the benefits—by making a race close (Grofman, B., Collet, C., & Griffin, R. 1998)— translates into a higher voter turnout. Still another strand of thought has explained the effect of social factors on voting behavior. A person has a network of friends, family, and coworkers and votes because the individuals around him or her are voting (Lake & Huckfeldt, 1998). A person may also live in a contested state or district, so his or her name shows up repeatedly on the call list of a campaign; the person votes because of phone calls reminding him or her to vote (Huckfeldt & Sprague, 1992). However, the core of all of these theories explains that demographic differences in voting are the result of different kinds of people in different kinds of situations making different decisions about whether or not to vote. Voting is consistent because any given person's situation does not change much from one election to the next and the individual will make similar decisions in similar situations.

Another way of viewing an election is to see it as an event related to past elections. In this way of understanding an election, an individual may choose to vote for all the same reasons suggested before (civic duty, ease of voting, and so on). However, a person will also choose whether or not to vote because of what has occurred in past elections. The difference in turnout among demographic groups may be due to different starting points; some groups are more likely to vote than others. However people consistently vote because they have voted before (Gerber et al., 2003). In this view of elections, the most common explanation of voting consistency is that voting is a habit. The idea of a voting habit is not a new one; it probably goes as far back as the 1960s (Campbell et al., 1960). However, it was not until 2000 that Green and Schacar (2000) began to test the habitual nature of voting.

## 2.2 Voting as Habit

The first step in establishing the theory that voting is a habit is to establish it as an empirical possibility. Using a variety of regression techniques, Green and Schacar (2000) showed that the relationship between voting in one election and likelihood of voting in the next was both robust and large. Later, Fowler (2006) showed that a behavioral model of voting as habit was consistent with observations made about national turnout in the United States. The most important step in the establishment of voting a habit was when Gerber et al. (2003) conducted a field experiment that showed a large causal effect between voting in one election and the next. Cutts et al. (2009) later replicated the experiment in the United Kingdom with similar results.

The Gerber study established voting as a habit, and many others began to expand on the idea of habitual turnout. The next logical step was to describe the habit. We now know that it typically takes three elections for an individual to become a consistent voter or non-voter (Franklin, 2004). Through inventive use of growth curves, Plutzer (2002) discovered that voting in one election has a larger effect on the future action of the individual than does the act of non-voting. Using elaborate time-series analysis, Denny and Dolye (2009) found that the effect of voting in one election could increase the likelihood of voting in the next by as much as 13%.

While much of the work on voting habits since Green and Schacar is both interesting and informative, it simply assumed that voting is a habit and moved on to describe the form and behavior of the voting habit. In 2011, Aldrich et al. approached habitual voting in a new way. Instead of assuming that voting was a habit and then looking at its shape and form, they argued that if voting is a habit, it needs to meet the criteria of what a habit is. Then they showed that voting does in fact meet those qualifications. Aldrich et al. found that notions of civic duty or

partisan fervor, classic indicators of voting turnout, were not as strong with habitual voters as with new voters, indicating that values that may initially lead to voting do not need to remain once a voting habit is established.

From here, I will follow the example of Aldrich et al. and look at literature about what makes a habit a habit.

### **2.3 A Brief Look at Habits**

According to Wood and Neal (2007), a habit is a learned inclination to repeat past responses. A habitual act is triggered by the context in which the response occurred. Contextual cues may be, for example, a location, the presence or absence of other people, a series of actions that preceded the response; they need to have occurred at the same time as habitual response and as the habit was forming. As Aldrich et al. (2011) puts it, a habit forms by repeatedly performing the habitual action under that same set of conditions. Eventually, the action performed becomes automatic whenever the same conditions exist.

Lally et al.(2011) discovered that the important aspect of habit formation, namely, creating the link between cues and action, initially takes quite a bit of effort, but over time becomes much easier. This means that for a new voter, who is just beginning to form the habit of voting, a lot of effort may be required to get the him or her to vote. Joe, the new voter from the earlier example, may need to set up a plan as to when and where to vote or he may need a phone call, reminding him to vote, because for him the mental link between election environment and voting has not yet been formed. John, the habitual voter, sees somebody wearing an “I voted” sticker and almost without thinking, goes to the polls on the way home from work.



What happens to a habit when a person performs a counter-habitual response? In the case of voting, what happens when a person fails to vote? Wood and Neal (2007) state that once a habit is formed, an infrequent response other than the one expected from the habit has little effect on the habit. The picture is not so clear when a habit is forming. Lally et al. (2010) performed a study in which participants chose to form a new healthy daily habit, which ranged from drinking eight glasses of water a day to doing daily sit-ups to eating healthy snacks. The study showed that when an individual missed an opportunity to perform the expected act, there was a slight decrease in how automatic the habitual act was the next time, but it did not drastically slow the formation of the habit. Armitage (2005), however, did a study in which several individuals attempted to take up the habit of going to the gym. Armitage found that if in the first five weeks, an individual failed to go to the gym for a week, the individual was much less likely to form a habit of going to the gym.

### **3. Hypothesis**

In proposing an answer to “What happens to a voter’s voting habits when the individual fails to vote?” I follow Aldrich’s lead and turn to the literature on habits. For an individual who is already in the habit of voting, failing to vote in an election should have a little effect on his or her voting habit. Since infrequent counter-habitual responses have little effect on a habit, failing to vote in an election should have little effect on the likelihood that the individual will vote in the next election.

For the new voter or the voter who is beginning to vote, the expected response is a little more complicated. Elections occur infrequently; so at first glance, failing to vote seems to have more in common with Armitage’s participant, who failed to go to the gym for a week, than it does with Lally et al.’s participant, who occasionally, on a given day, missed an opportunity to

make a healthy choice. It is probable that if an individual who is just starting to vote fails to vote in an election, the likelihood that he or she will vote in the next election should drastically decrease.

## 4. Data

### 4.1 Description of the Voter File

Looking to answer the question, "What effect does missing an election have on voting habits?" I will use the Tennessee voter file. I received the voter file from the Tennessee Secretary of State, through the Tennessee Democratic Party. It is a list of all the registered voters in Tennessee as of May 2012. It contains limited demographic information, including the gender, date of birth, date of registration, and limited racial information of the voter. It also includes a variety of geographic information, such as the district and precinct in which the individual is registered. Most important, the voter file contains each person's voting history for state level elections from 1992 until 2010. To be clear, voter history is not *how* an individual voted but only *if* he or she voted in each election. Four elections are considered state level elections: the November general election, both presidential and midterm; a primary held in August for state (governor) and federal (United States House of Representatives) offices, and a local primary in May for county (sheriff) offices. The presidential primary occurs at different times, but that election replaces the local primary. The voter file also includes information about whether an individual voted in the Democratic or Republican primary elections..

2,875,402 voters are in the voter file, after I dropped 1,052, 071 records. 1,047,997 of the dropped records were individuals who registered to vote after the 2006 general election; for this analysis I needed at least two midterm elections on file. The remaining 4,074 records had

missing or clearly incorrect dates of registration; a correct date of registration was necessary to transform the file into its panel-data format.<sup>1</sup> The panel-data format produced a data set with 60,509,051 voter-elections, each representing an individual's voting behavior for an election from the 1992 presidential primary to the 2010 general election. Most of the analysis was performed only on midterm elections, which included 10,772,046 voter elections.

A good starting point for understanding longitudinal data is to understand who is involved, when they became involved, and how long they are being followed. It is not surprising that the most common age to register to vote is eighteen, followed by nineteen. The number of people who are over the age of nineteen and register to vote is about the same from year to year, but the number of new registrants declines as people get older (see Figure 1). About half of the registrants, are over the age of thirty-one. The fact that many individuals register relatively late in life and may have voted elsewhere, makes defining a person's voting habit somewhat tricky; I will deal with this issue more thoroughly in the methodology section.

Table 1 shows that voters in Tennessee are not all that different from voters elsewhere. Whites vote at a higher rate than Blacks do. Older voters vote more often than younger voters do. Partisans, defined as individuals who have voted in a primary, vote at higher rates than non-partisans do. One statistic that does buck the national trend is that men seem slightly more likely to vote in midterm elections than women are. But overall, the voting patterns appear unexceptional.

---

<sup>1</sup> Panel-data creates several observations for a single person over time. This form allows for many kinds of analysis related to changes over time.

## 4.2 Strengths and Limitations

The real strength of using the Tennessee voter file in this research is that the file lets us look at voting behavior since 1992; since most individuals in the file were registered after 1992, the voter file allows us to see the entirety of their Tennessee voting life. Much of the previous regression analysis used longitudinal studies that typically lasted only a few years and only included data on voting behavior for a single kind of election. The time frame of these studies limited the degree that voting patterns could be discussed. With eighteen years of voting history, I am able to capture more accurately the scope of a person's voting behavior prior to an election.

The other major advantage to the Tennessee voter file is its size. When the data set includes millions of observations, assuming a functional form other than linear is possible without losing degrees of freedom. In a similar vein, the size allows for the detection of more granular patterns..

However, the limitations of the voter file offset its strengths. The voter file is not a random sample. To be on the file, an individual must register to vote, which is a result of the voter's actions. Furthermore, Erikson shows that individuals most likely to register, are also those who are most likely to vote (1981). The voter file, therefore, is limited by the possibility of selection bias. While this aspect of the selection problem can be controlled, the way in which individuals fall off the rolls cannot be.

The election commission may remove someone from the voter file for several reasons. If an individual moves out of the county, registers elsewhere, changes his or her name, commits an "infamous" crime, requests to be purged from the rolls, or dies, then the individual can be removed from the voter file. How these methods might bias results is not clear since most of these issues are conceivably independent of voting behavior. Any bias they create is more than

likely similar to individuals' being dropped from a longitudinal study because they did not finish the study. For this study, the effect of having voters removed from the file for these reasons is not immediately clear. However, someone who does not vote for more than one election cycle (starting and ending with the presidential election in November) becomes inactive, and may also be removed from the voter files; so a person who consistently does not vote is less likely to be on the rolls.

The effect of having voters dropped from the voter file is shown in Figure 2. While the turnout in 2010 estimated from the voter file is similar to the true turnout reported by the Secretary of State, the difference becomes big quickly, ballooning to a fifteen percentage point difference in 2000. Figure 2 shows clearly that the voter file has a bias toward people who vote.

## 5. Methodology

The hypothesis—failing to vote should have less of an effect on habitual voters than on non-voters—assumes that the probability of voting in a given election is a function of voting in the previous election of that type, habit, and some controls. Equation 1 formally states the parameters of this function:

$$p(\text{voting this election}) = f(\text{voting in last election}, \text{habit}, \text{voting in last election} * \text{habit}, \text{controls}, \text{error})$$

The hypothesis also makes a difference in difference claim. In particular, it claims that

$$p(\text{vote} | \text{voted last time}, \text{high habit}) - p(\text{vote} | \text{no vote last time}, \text{high habit}) \ll p(\text{vote} | \text{voted last time}, \text{low habit}) - p(\text{vote} | \text{no vote last time}, \text{low habit})$$

In the methodology section, I will begin by discussing how I will measure this election, last election, and habit; and I will discuss why I chose to measure these concepts in the way I did. I will describe the process I used to build a model approximating Equation 1. Finally, I will

describe the process I used to measure and to account for possible bias introduced from the dataset.

## **5.1 Operationalization of Variables**

### ***Voting in This Election and the Last***

In this analysis, three important variables need to be described and measured: *voting in this election*; *voting in last election*; and most important, *habit*.

*Voting in this election* is a relatively simple concept: An individual casts a ballot in an election that occurs at a time designated as now. On the voter file are six kinds of elections, dating from 1992 to 2010. Any or all of these elections could be selected as “this election” for the purposes of measuring. In this study, I chose to use midterm general elections as “this election.” Incorporating more than a single election allows for discussion of the effect of habit on elections in general, since the idiosyncrasies of a particular election will be averaged out across all the elections.

I chose midterm elections for both practical and theoretical reasons. I chose a single type of election instead of all elections because each kind of election creates a different contextual environment. (Presidential elections create an all day, every day stream of news for months on end, whereas a local primary might be mentioned on the radio on election day.) Since habits are trained responses to contextual cues, it is reasonable to assume that slightly different habits are in play for each type of election. Limiting this study to a single type of election ensures that measures of habit are more likely to be measuring the same habit rather than six different kinds of habit. I selected midterm elections, as opposed to presidential elections, because the most

recent election on file was a midterm election; using the most recent elections should reduce any bias related to time.

Any particular election in which voting is counter-habitual should not make much of a difference; the voter's likelihood of *voting in this election* should be similar to what it would be had he or she not missed that election. I selected the last midterm election as the missed election for several reasons. Since the previous midterm election is of the same type as *this election*, with the same contextual cues, missing it would be a counter-habitual act. In addition, the past election of the same type is used in much of the literature on how voting in a past elections effects voting in a future election. To make interpreting the regression analysis easier, I also flip the axis, so that the variable becomes missing the last midterm election rather than voting in the last midterm.

### ***Habit***

Ideally, given the cues that a person is likely to experience during a midterm election and has experienced in past elections, the measure for habit would be able to determine how automatic the act of voting is for that individual. A good habit measure would measure the probability of an individual's voting in the midterm election, but would also take into account the idea that different stimuli trigger different kinds of voting patterns. For instance, some voters may vote only in presidential elections, responding to the hoopla on the news. Others may vote because they received information about the candidates, and the flyer triggers voting in the election. Others may respond to cues we simply don't or can't notice.

Unfortunately, the voter file does not include a scale of automatic voting; it can only be inferred from an individual's voting history. One idea is to use as a measure of habit the

proportion of elections in which an individual has voted in the past. The intuition is that if voting is 75 percent automatic for an individual, then he or she will vote 75 percent of the time. This method has the practical advantage of having a set and constant scale; the measure of habit can only exist between zero and one. Unfortunately, it is also untenable for this study. This research is examining the difference between habitual voters and non-habitual voters.

Individuals who are least likely to have formed a habit of voting are those who have not voted often in the past. However, new voters have not had the opportunity to vote in many elections, so the degree to which the proportion reflects their voting habit is noisy. Consider, for example, a voter who is voting in his or her second election. Since the individual is new to voting and assuming that voting habits do not form quickly, he or she does not have a fully-formed voting habit. However, with only one election under his or her belt, the individual's habit score must be either a one or a zero. Because of the noise of estimating habit in this way, I chose not to use as a measure of habit the proportion of elections in which each person has voted.

Another simple way of measuring habit is to use a count instead of a proportion. This method has disadvantages as well. It does not properly take into account the amount of time that passed between the elections in which the individual voted. A relatively new voter with a habit score of three has voted in almost every election possible. An older voter with a score of three may have voted in a few elections as the mood struck him or her, voted three times in a row and stopped, or started voting after never having voted before (he or she would then effectively be a new voter). Another disadvantage is that the scale changes from one election to the next.

Between 2010 and 1992, when the voter file begins, thirty elections are on the file; so in 2010, a voter could have a score as high as 28 (excluding *this election* and the *last election*). As elections move closer to 1992, the number of possible elections in which a person could vote is



reduced; so the scale changes. However, measuring habit by a count of past elections also has a strong advantage: Someone who has not voted often and so has not had time to form a habit will always have a lower score than an individual who has voted consistently for a long time. For this reason, I used as a measure of habit a count of past elections in which individuals voted.

The next step was to determine which elections should be counted in building a habit score. In this analysis, I chose four different ways of counting habit, attempting to account for the ways in which voting (or not voting) in different kinds of elections might affect the habit of voting in the midterm elections. It could be the case that voting in any election has the same effect on a person's voting habit. Therefore, the Type I habit score was simply a count of all elections, no matter the type, in which an individual voted prior to *this election* (*last election* was also excluded from the count). The second habit score, Type II, is a count of all past elections in which an individual has voted, excluding presidential elections (again *last election* is also excluded). The Type II habit score was based on the argument that individuals who vote in presidential elections have many more reminders to vote than they have in midterm elections, so voting in a presidential election will more than likely have little effect on the habit of voting in a midterm election. Type III, a count of the number of past midterm elections in which an individual voted, is based on the assumption that only elections of the same kind affect the habit of voting in future elections of that kind. The Type IV habit score is a set of six counts, one for past presidential elections; another for past midterm elections (excluding *last election*); and four more counts, one for each kind of primary. Type IV makes the weakest assumption of all the previous types, in that it assumes that all elections effect the formation of a voting habit for midterm elections. However, it also assumes that the effect is different for each kind of past election.

## 5.2 Preliminary Analysis

Prior to working with logistic regression analysis on the full dataset, I created marginal probability tables on a sample of the data to see if the effects of habit would be clear and to verify that the measure of habit followed expectations.

A random sample of 100,000 individuals was taken from the voter file. The sample included all of the midterm elections the 100,000 individuals could have voted. For each measure of habit, individual records were placed in one of two bins. The first bin contained individuals whose voting habit was less than or equal to the mean habit score; the second bin contained individuals whose voting habit was greater than the mean habit score. Then I constructed the marginal probability tables.

The first set of tables are simple two-by-two tables that compare the percentage of people who voted in the last midterm election with those who voted in this midterm election. They also compared high (above the mean) and low (at or below mean) habit scores with voting in this election.

Once habit and past voting were compared to current voting history, I constructed a set of two-by-four margin tables to compare all three variables and to identify possible interactions between habit and previous voting.

## 5.3 Primary Analysis

The hypothesis argues that for an individual with a low habit score, the effect of missing a previous election on the probability that he or she will vote in the next election should be bigger than the effect of missing an election on the probability that an individual with a high

habit score will vote in the next election. Therefore, the core analysis of this paper is based on a difference-in-differences model that examines these probabilities. This section describes the sixteen difference-in-differences logit models that were used. In short, I generated the models by combining the four habit types and four different levels of controls into a logistic regression.

A difference-in-differences model typically consists of an ordinary-least-squares (OLS) regression with two variables of interest and the interaction between the two, which represents the difference in differences. I used a logit model instead of an OLS model because an OLS probability model tends to produce poor results when the probability is close to zero or near one. Since new voters who have not voted in the last election tend to vote in small numbers and individuals with high habit scores tend to vote have high turnout rates, I needed to use a model that could more easily handle probabilities at the low and high end; so I chose to use the logit model. Otherwise, the first model used on each type of habit, Model 1, is a basic difference-in-differences model with no controls. It incorporates the two variables of interest, whether or not the individual voted in the prior midterm election and the individual's habit score. It also includes the interaction between the two, which represents the difference in differences. The formula for the model is below (Equation 3):

$$\begin{aligned} \text{Logit}[\textit{vote midterm}_t] = \\ \beta_0 + \beta_1(-\textit{vote midterm}_{t-1}) + \beta_2(\textit{habit}) + \beta_3(-\textit{vote midterm}_{t-1} * \textit{habit}) \end{aligned}$$

Since the expectation is that missing an election will decrease the probability of voting,  $\beta_1$  in Equation 3, which relates how much missing the last election effects voting in this midterm for an individual with no habit score, should be a negative number.  $\beta_2$ , which describes how increasing habit effects voting in the midterm, should be positive.  $\beta_3$  describes how missing the last election changes from its baseline as an individual's habit score increases; this value should have the opposite sign of  $\beta_1$  and so should be positive.

With no control variables, Model 1 is rather naive. In particular, the weakness of a count as the measure of habit is that a habit score of five means something different for a new voter and a person who has been registered to vote for a long time. In addition, the age of a voter has been shown to effect a person's likelihood of voting. Incorporating age into the model should control both for the weakness of the habit measure and for one variable that could cause bias.

The other weakness of the model is that since the voter file only goes back to 1992, there is a cap on the habit score. The Type I habit score, in which habit is represented by voting in all past elections, the maximum score is 28; but in 2006, the maximum score is only 22. In addition, each election has an idiosyncratic effect on turnout. Incorporating dummy variables to indicate which election *this election* is, should control for both of these issues.

Model 2 incorporates and controls for both the age of the voter and the specific election. Model 2 controls for variables related to time. The formula is described by Equation 4:

$$\begin{aligned} \text{Logit}[\text{vote midterm}_t] = & \\ & \beta_0 + \beta_1(-\text{vote midterm}_{t-1}) + \beta_2(\text{habit}) + \beta_3(-\text{vote midterm}_{t-1} * \text{habit}) + \\ & \beta_4(2010\text{indicator}) + \beta_5(2006\text{indicator}) + \beta_6(2002\text{indicator}) + \beta_7(\text{age}) \end{aligned}$$

Model 3 attempts to take into account, to the extent permitted by the voter file, traits about individuals that may affect how they vote. In particular, Model 3 adds race and gender to Model 2. The formula used for Model 3 is expressed in Equation 5:

$$\begin{aligned} \text{Logit}[\text{vote midterm}_t] = & \\ & \beta_0 + \beta_1(-\text{vote midterm}_{t-1}) + \beta_2(\text{habit}) + \beta_3(-\text{vote midterm}_{t-1} * \text{habit}) + \\ & \beta_4(2010\text{indicator}) + \beta_5(2006\text{indicator}) + \beta_6(2002\text{indicator}) + \beta_7(\text{age}) + \\ & \beta_8(\text{Caucasion}) + \beta_9(\text{Other}) + \beta_{10}(\text{Unknown}) + \beta_{11}(\text{Is Male}) \end{aligned}$$

Model 4 incorporates a dummy variable for the individual's county of origin and helps to account for differences in location that may affect voting behavior. In particular, model 4 takes

into account the ease or difficulty of voting at polling locations and helps to control for areas where efforts are consistently made to increase turnout. Model 4 is expressed in Equation 6:

$$\begin{aligned} \text{Logit}[\text{vote midterm}_t] = & \\ & \beta_0 + \beta_1(-\text{vote midterm}_{t-1}) + \beta_2(\text{habit}) + \beta_3(-\text{vote midterm}_{t-1} * \text{habit}) + \\ & \beta_4(2010\text{indicator}) + \beta_5(2006\text{indicator}) + \beta_6(2002\text{indicator}) + \beta_7(\text{age}) + \\ & \beta_8(\text{Caucasion}) + \beta_9(\text{Other}) + \beta_{10}(\text{Unknown}) + \beta_{11}(\text{Is Male}) + \\ & \beta_{12-106}(\text{county residence indicator}) \end{aligned}$$

#### 5.4 Dealing with Bias

Given the millions of observations in the Tennessee voter file, the results of the model will be precise and will more than likely appear significant. Therefore, the biggest source of error will come from bias. Two broad sources of bias are of special concern. The first involves model misspecification and covers issues such as measuring habit incorrectly, omitting an important variable from the model, or using the wrong functional form of the model. In general, I am dealing with this kind of bias by using and reporting the results of several models. While it doesn't remove the possibility of bias, by using several measurements of habit and a variety of control variables, I can more easily spot bias due to model misspecification. Then I can choose the model or models that are most likely correct.

The second form of bias comes from what is missing from the dataset. This was briefly addressed in the "Strengths and Limitations" part of the Data section. Individuals who have not voted in any election over the course of two presidential cycles are removed from the voter file; so any effect these non-voters may have on the model is lessened or removed.

### ***Measuring Bias***

Since voters are removed from the voter file over time, measuring the amount of bias introduced by their removal from the file is possible. If the model is correctly specified, the effect of habit should be about the same for each midterm. However, if bias is introduced into the model due to voters' records being removed, the amount of bias should increase with each cycle. Therefore, a change in the results between election cycles should provide an estimate of bias due to missing voters and also indicate a way of making correction for it.

To estimate the amount of bias generated by individuals' being removed from the voter file, I first created a set of three logit models similar to the last model described in "Primary Analysis" (Equation 6). In each model, I replaced  $midterm_t$  with a particular midterm election and removed additional indicator variables that were not needed to evaluate a single election. Finally, I only used two measures of habit, Type I, which measured habit as the total of all the past elections in which an individual voted, and Type IV, which measured habit related to each type of election. I chose these two measures based on the results of the primary analysis.

$$\begin{aligned} \mathbf{7.1)} \text{ Logit}[\text{voted 2010}] = & \beta_0 + \beta_1(\text{voted 2006}) + \beta_2(\text{habit}) + \beta_3(\text{voted 2006} * \text{habit}) + \\ & \beta_4(\text{age}) + \beta_5(\text{Caucasion}) + \beta_6(\text{Other}) + \beta_7(\text{Unknown Race}) + \beta_8(\text{Is Male}) + \\ & \beta_{9-103}(\text{county residence indicator}) \end{aligned}$$

$$\begin{aligned} \mathbf{7.2)} \text{ Logit}[\text{voted 2006}] = & \beta_0 + \beta_1(\text{voted 2002}) + \beta_2(\text{habit}) + \beta_3(\text{voted 2006} * \text{habit}) + \\ & \beta_4(\text{age}) + \beta_5(\text{Caucasion}) + \beta_6(\text{Other}) + \beta_7(\text{Unknown Race}) + \beta_8(\text{Is Male}) + \\ & \beta_{9-103}(\text{county residence indicator}) \end{aligned}$$

$$\begin{aligned} \mathbf{7.3)} \text{ Logit}[\text{voted 2002}] = & \beta_0 + \beta_1(\text{voted 1998}) + \beta_2(\text{habit}) + \beta_3(\text{voted 2006} * \text{habit}) + \\ & \beta_4(\text{age}) + \beta_5(\text{Caucasion}) + \beta_6(\text{Other}) + \beta_7(\text{Unknown Race}) + \beta_8(\text{Is Male}) + \\ & \beta_{9-103}(\text{county residence indicator}) \end{aligned}$$

If the bias introduced by individuals' being removed from the voter file is more or less constant through time, the relationship among the coefficients in Equations 7.1, 7.2, and 7.3 above can be represented by Equation 8:

$$\beta_{tx} = (2012 - t)\rho + \beta_{truex}$$

The coefficient at time  $t$  is equal to the true coefficient plus bias, which increases linearly. I used OLS regression to estimate the results of the bias introduced into the file each year. The constant in the OLS regression represents a coefficient after correcting for bias due to voters' being removed from the file.

## 6. Results

### 6.1 Marginal Probability Tables

Table 2 presents the results of the preliminary analysis of habit. The table includes several marginal probability tables, grouped together according to different measures of habit. Overall, the table indicates that an individual's voting in the last election, as well as having a high or low habit score, is an important factor in determining the probability of whether or not that individual will vote in the next election. However, the results are dependent on the type of habit score.

The first part of Table 2 shows that for about 71 percent of individuals, voting behavior stays the same from one election to the next. About 40 percent voted in both the last election and in this election. About 31 percent of the population did not vote in either election. This result is simple and expected.

Also expected, individuals who have a high habit score are much more likely to vote in the midterm election than individuals who have a low habit score, no matter how their score is

determined. Of interest is the result indicating that for most ways of calculating voting habit, individuals with low habit scores seem only slightly less likely to vote than not to vote.

The interesting, and slightly off-topic, exception is related to presidential elections. Individuals who have voted in fewer than the mean number (1.7) of past presidential elections have slightly less than one-to-two odds of voting in this midterm election. This might mean that individuals typically begin voting in presidential elections before they move on to vote in other kinds of elections.

The most important results come from combining the two-by-two tables, which group people according to their habit score and also according to whether or not they voted in the last election. For individuals with a low habit score, the difference between the proportion of people who voted last time and those who did not is about 33 percent. This value varies only slightly ( $\pm 4$  percentage points) with the way habit is counted.

The hypothesis states that for individuals with high habit scores, the difference in turnout between those who voted in the last midterm election and those who did not should be much smaller than the roughly 33 percentage points that describe low-habit voters. For most measures of habit, the difference for high-habit voters is less than five percentage points lower than it is for low-habit voters, which is contrary to what a habit would dictate. The two ways of measuring habit that produced substantively lower values were Type II, which counts voting in non-presidential past elections, and oddly enough, an aspect of Type IV, in which habit is calculated by looking only at presidential elections. The difference in differences for these two groups was about thirteen and twelve points respectively, which would support the idea of a voting habit.



## 6.2 Results from the Difference-in-Differences Models

The sixteen difference-in-differences logit models produced fundamentally different results depending on how habit was measured. When using Type I (all past elections) or Type II (all past elections except presidential elections) measures of habit, the results seemed to strongly support the hypothesis. Type III (only past midterm elections) also supports the hypothesis, but its support is much weaker than the results of Type I or II. Type IV, which measures habit with a separate count for each kind of election, showed results that were completely unexpected. For all ways of measuring habit, incorporating more control variables generally muted the apparent effect of both habit and missing the last midterm on voting in this midterm election, but did not substantively change the outcome.

### *Type I and Type II*

Table 3 and Table 4 show the results of the logit regressions for Type I and Type II measures of habit. Measuring habit as a simple count of most or all past elections produces results precisely as the hypothesis would predict. The value associated with how much an effect missing the last midterm election had on voting in this election was both large and negative, hovering around -1.5 and -1.6 for Type I and Type II measures of habit. The value that describes the effect of habit is relatively large with a value around 0.2 and .22 for Type I and Type II habit, respectively. The term that describes how increases in a person's habit score effects the impact of missing the last midterm is around 0.1; this implies that once a person's habit score reaches the mid-to-low teens, forgetting to vote in the last election has no effect on the likelihood of voting in the next election.

While useful, regression tables and particularly logit regression tables, can only do so much to provide a clear understanding of the results. Figures 3-6 present a much clearer picture of the results of the models that used Type I and Type II measures of habit. In the following paragraphs I will focus on Figure 3 and Figure 4, which describe the results of the models that used Type I habit measure. I won't discuss Figure 5 and Figure 6, which deal with Type II habit; but the results are clearly similar.

Figure 3 shows the probability of voting in a given midterm election as predicted by Model I, the naive model without controls. Individuals who have a habit score of zero have never voted in an election, with the possible exception of the last midterm election. Individuals who have a habit score of zero but did vote in the last midterm election have about a 46.7 percent chance of voting in the next election. People with a habit score of zero who did not vote in the last midterm election have only a 17.1 percent chance of voting in the next election. For individuals with a low habit score, there is 29.6 percentage point difference between the likelihood of their voting or not voting in the next election, depending on whether they voted in the last election. A person who voted in each presidential and midterm election (with the possible exception of the last midterm) for the past five cycles and who may have voted once or twice in the presidential primary would have a habit score of ten. In common parlance, this person would be considered to have a voting habit.

A habitual voter, with a habit score of ten, has an 84.3 percent chance of voting in the midterm election, if he or she voted in the last midterm. A habitual voter, with a habit score of ten, who forgot to vote in the previous midterm election still has a 78.1 percent chance of voting in this midterm election. For the habitual voter, failing to vote in the last election generates only a 6.2 percentage point difference in voting behavior. Missing an election has a 29.6 percentage

point effect on new voters, but only has a 6.2 percentage point effect on habitual voters. These results strongly support the hypothesis.

Figure 4, like Figure 3, shows the probability of voting in a given midterm election; however, it depicts the results of Model 4, which is the model that incorporated all of the control variables. Figure 4, therefore, depicts a prototypical Tennessee voter who is white (modal), female (modal), forty-seven years old (mean age), and living in Washington county (median of voting effect). In addition, Figure 4 holds the election constant at the 2010 midterm election. The model suggests that if the prototypical voter were a new voter, with a habit score of zero, and had not voted in the last midterm election, the probability that he or she would vote in 2010 would be about 5.7 percent. If the same individual had voted in the last midterm, 2006, the likelihood that he or she would vote in 2010 would jump to 22.7 percent, a difference of 17.0 percentage points. However, if the prototypical voter was a habitual voter, with a habit score of ten, and had voted in the last midterm election, in 2006, the probability that he or she would vote in the 2010 election would be 69.4 percent. The prototypical voter with a habit score of ten, who forgot to vote in the last midterm election, would have a 57.9 percent chance of voting in 2010. Therefore, for the habitual prototypical voter, missing an election creates about an 11.5 percentage point difference in the probability of voting in the next election. There was a 23.4 percentage point difference in differences in Model I, but once all the control variables are incorporated into the model, the difference in differences drops to 5.5 percentage points. This still is supportive of the hypothesis, but the overall effect is greatly muted.

### *Type III*

Type III, which measures habit as a count of all past midterm elections, produces results similar to those of Type I and Type II. However, this measure implies that voting habits are formed over a longer period of time than Type I and Type II measures of habit imply. As with Type I and Type II, the Type III measure of habit indicates (see Table 5) that missing the last midterm election is detrimental to a voter's likelihood of voting in this election (-1.05 in Model 4). A higher degree of habit translates into a higher likelihood of voting (0.62). Most important, the more frequently an individual has voted in past midterm elections, the less of difference voting in the last midterm election makes on whether he or she votes in this election (interaction term is about 0.08).

Figure 8 begins to show more clearly the difference between Type III and Types I and II measures of habit. While the red and blue lines in Figure 8 get closer to each other—at a habit score of zero, the difference between the two is 18.5 percentage points; at a habit score of four, the difference is 11.4 percentage points, which closes the gap by 7.1 percentage points—the lines seem flatter and more parallel than they are in Figure 4 or Figure 6. One reason the two lines appear more parallel in Figure 8 is that they do not meet in the frame of graph. Model 4 projects that the intersection between the two lines would occur at a Type III habit score of 12.9. While this number is similar to the intersection point of 14.6 (Type I) and 13.1 (Type II), its implications are different.

The intersection between the red and blue lines is the point at which failing to vote in the last midterm election no longer has an effect on voting in this election. It is a reasonable point at which to state that a voting habit is fully formed. This means that the minimum amount of time it would take for someone to achieve a fully-formed voting habit is fifteen cycles or by the time

an individual is about seventy-eight years old, according to the Type III measure of habit.

According to the Type I and Type II measures, a habit could be formed in as few as 2.5 cycles or by the time a person is about twenty-three years old, though the average age of individuals with such high habit scores is forty-four.

#### *Type IV*

Type IV, which measures habit as six different counts of all past elections, may yield the strangest results. The reason for breaking the measures of habit into six measures instead of one was to respond to the idea that different elections might affect a person's voting habit differently. Table 6, the results of four models related to the Type IV measure of habit, certainly confirms this notion. In Model 1, which incorporates no controls, the first few values are expected: Missing the last midterm election has a fairly large and negative (-0.95) on voting in this election. Individuals who have voted in past presidential elections, past midterms, and past August primaries during the midterm are more likely to vote in the next election as would be expected of individuals with a voting habit. (The coefficients related to those three elections are 0.82, 0.37, 0.18.) Then comes the surprise: The coefficients related to the remaining three elections are negative (-0.05, -0.17, -0.17). Other factors being equal, an individual who has voted in any previous presidential or local elections or in the August primary in presidential years is less likely to vote in the midterm election compared with an individual who has not voted in these elections.

Similar results occur with the interaction term. Increasing the number of times an individual has voted in previous presidential or midterm election reduces the effect that missing an election has on voting in the next election, as a voting habit would predict. However,

surprisingly, having a history of voting in any primary elections seems to increase the effect of missing the last midterm on voting in this election.

These unexpected results persist, even after introducing a variety of control variables. They also persist when building three models, one for each midterm election (see Table 8). The persistence of the negative values implies that these coefficients, which indicate that voting in certain kinds elections decreases the probability of voting in the next midterm election, is not purely a statistical fluke. One possible explanation is that there are different kinds of voting habits. In particular, the results would be consistent with the notion that a sufficiently large segment of the population votes only in local or primary elections. Perhaps, individuals are intentionally forming a voting habit to only vote in elections in which their vote would make a difference. If enough people developed a voting habit only in local and primary elections, then they would have a high count of voting in the variety of primaries and not voting in the midterm elections, which would create negative coefficients for the relationship between voting in the primary and voting in the midterm. Since not everyone who votes in primaries votes only in local and primary elections, knowing whether or not an individual missed the midterm election ceases to be an indicator of how formed a voting habit is but instead becomes an indicator of the kind of voting habit an individual has. This would explain why the interaction terms are negative.

### **6.3 Measuring Bias results and another Look at Study Limitations**

With only three midterm elections included in the research, the attempt to measure bias was difficult and produced largely insignificant results. However, the point estimates of the bias tend to indicate that the primary analysis more than likely exaggerates the effect of all three variables of interest (failing to vote in the last election, habit, and interaction). Finally, a careful

look at the results may indicate the presence of model selection bias because I did not fully control for the effect a particular election has on voting.

Models based on three observations are not likely to be accurate. For the model to gauge the effect of bias, the effect of bias would need to be both relatively large and strongly correlated with time; and it is not. The process required estimating bias on twenty-six different variables (see Table 7 and Table 8). Only two, failing to vote in the last election (in Table 7) and voting in previous presidential elections (Table 8), were significant at the .05 confidence level<sup>2</sup>. But having measured twenty-six variable, the two variables that were significant at the .05 level may have simply happened by chance.

With so few significant results, making adjustments to the estimates presented in the primary analysis is not a good idea. However, a close look at Table 7 and Table 8 shows that the magnitude of the relationship between failing to vote in the last election and voting in this election consistently increases both between 2010 and 2006, 2006 and 2002. This would imply that any bias that is on the file is probably inflating the effect failing to vote has on voting in the next election, though by how much is not clear. The two tables show similar results for habit and the interaction term, indicating that the results from the primary analysis are probably inflated to some degree, though how much is unclear.

In 2006, there was a close race for senate. The Democratic candidate appealed strongly to classical Democratic voters. Table 7 and Table 8 clearly an increase in these demographic groups: youth, women, and Blacks. Controlling for voter history, a young, Black, or female voter was more likely to vote in 2006 than in 2010 or 2002. The increase in voting means that different demographic groups change behavior in different elections. In the primary analysis, an

---

<sup>2</sup> With one degree of freedom, .05 significance is 12.7 standard errors from the mean, rather than the typical 1.96.

indicator variable was incorporated to control for particular elections. The effect of the 2006 election on the specific demographic groups implies that more may have needed to be done to fully control for the effect of particular elections.

Overall, the attempt to measure the amount of bias introduced by voters' being removed from the voter file has placed strong limitations on the conclusions that can be made about habit. The process has provided some evidence that any bias more than likely does effect the results of how habit informs voting in a midterm election. It also has shown that there is likely to be bias due to omitted variables from the various models. However, attempts to measure bias so that corrections could be made for the bias, led to results so imprecise as to be unhelpful.

## **7. Conclusions**

For voting to be a habit, the act of voting needs to have a significant effect on voting again in the future. The habitual act of voting forms the voting habit. However, once the habit is formed, the inclination to vote should persist; so the high likelihood of an individual's voting should be about the same whether or not he or she actually voted in the last election. In this thesis, I used data from the May 2012 Tennessee voter file, a list of all the registered voters in the state and their voting history, to see if voting had the shape of a habit. To find the form of voting habits, I compared several different logistic regression models. I began with a naive difference-in-differences model that included only whether a person voted in the last election, one of several measures of habit, and the interaction between the two. Then I made the model progressively more complex by incorporating control variables that the data set would allow. Finally, due to concerns about possible selection bias introduced into the dataset from individuals' records that were removed from the voter file, I made a measure of bias by



comparing individual election years in a way that should progressively introduce more bias in a measurable way.

The attempts to measure bias were largely unsuccessful, but they did seem to imply the existence of bias. This limits the strength of the conclusions I am able to draw.

That being said, the results of the difference-in-differences analysis seem to suggest that voting does indeed take on the form of a habit. Voting in the last election seems to matter more for the individual who has not voted a lot in past elections, compared to someone who has voted in many of the last several elections. Though the way habit is measured makes a significant difference, our results seem to suggest that an individual's habit of voting in midterm elections becomes fully formed after he or she has voted in about fourteen past elections. After fifteen elections, a person's likelihood of voting in the next midterm election remains about the same, whether or not the individual remembered to vote in the last midterm.

For an individual who usually votes in presidential and midterm elections, fifteen elections is about seven full election cycles, or about thirty years, before the voting habit is fully formed. This result conflicts with Franklin's work, in which it was suggested that a voting habit would be formed in three election cycles. The conflict is more than likely explained by two factors. Franklin's work was examining presidential elections, while this research examined midterms. This thesis attempted to find and examine the strength of an individual's habit, whereas Franklin work was looking to see how many cycles it took before a person consistently voted or did not vote. Further research would be useful to pinpoint the discrepancy.

In both politics and political science, there is not only an implicit assumption about voting being a habit, which research by now has probably established, but also assumptions about different kinds of voter habits. Phrases such as "presidential voter," "super voter,"

"presidential and midterm voter" all suggest similar but distinctive voting habits. Most of the work done on voting habits to date has skirted the issue of the variety of voting habits by either examining the same type, for example, examining the pattern of voting in presidential elections or by examining two adjacent elections such as a primary and the following general. In this thesis, when habit was measured not as a single value, but as a count of all previous elections, I stumbled upon evidence of distinct voting habits. In particular, I found that individuals who voted in local elections were less likely to vote in midterm elections, providing evidence that there are "local voters," or perhaps "primary voters." These are individuals who vote only in local elections or perhaps only in primary elections.

Discovering "local voters" was a surprise; but if true, it raises several new questions for research into voting habits. What are the different types of voting habits? Can a voter's habit be shifted simply by inducing them to vote in different kinds of elections, or does there need to be a shift in the voter's thinking about voting before the habit can be changed? For example, would a local voter change his or her voting habits if he or she were led to believe that voting matters in the general election as well as in the local election?

In addition to pointing toward ideas for future research, this thesis has important implications for activists who are attempting to increase voter turnout. Efforts to increase turnout, or to "get out the vote," aren't easy. Possibly the best way to effect a long-term increase in voter turnout is to induce a habit of voting in individuals. Toward this notion of inducing people to become habitual voters, this research seems to recommend focusing "get out the vote" efforts on two groups of voters: Voters, who voted in the last election, but have not otherwise voted much at all; and individuals who have some history of voting, but forgot to vote in the last election. Focusing on these voters would seem to have the greatest effect, though field tests

would be recommended to confirm this finding. In addition, this research would suggest that individuals who have voted several times in the past need not be contacted. They are likely to vote anyway; and as their habit of voting is mostly formed, their failing to vote in this election does not decrease the likelihood of their voting in the next election.

The possible finding of individuals who only vote in local elections is somewhat more disconcerting for those who are trying to increase turnout. Depending on the mechanism of habit formation, there may be an upper limit of turnout across all elections, as individuals, dictated by habit are more likely to vote in certain types of elections and not in others. It could also mean that organizations working to get out the vote in several elections are working at cross purposes with themselves, encouraging different voting habits that are more or less exclusive. However, since this research did not examine the effect of “get out the vote” efforts on voting habits and because more work needs to be done on different kinds of voting habits, conclusions about campaign strategies remain, at best, speculation.

## References

- Aldrich, J. H., Montgomery, J. M., & Wood, W. (2011). Turnout as a Habit. *Political Behavior*, 33(4), 535–563.
- Armitage, C. J. (2005). Can the Theory of Planned Behavior Predict the Maintenance of Physical Activity? *Health Psychology*, 24(3), 235–245.
- Ashenfelter, O., & Kelley, S. (1975). Determinants of Participation in Presidential Elections. *Journal of Law and Economics*, 18(3), 695–733.
- Brody, R. A., & Sniderman, P. M. (1977). From Life Space to Polling Place: The Relevance of Personal Concerns for Voting Behavior. *British Journal of Political Science*, 7(3), 337–360.
- Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). *The American voter*. New York: Wiley.
- Cutts, D., Fieldhouse, E., & John, P. (2009). Is Voting Habit Forming? The Longitudinal Impact of a GOTV Campaign in the UK. *Journal of Elections, Public Opinion & Parties*, 19(3), 251–263.
- Denny, K., & Doyle, O. (2009). Does Voting History Matter? Analysing Persistence in Turnout. *American Journal of Political Science*, 53(1), 17–35.
- Erikson, R. S. (1981). Why Do People Vote? Because They Are Registered. *American Politics Research*, 9(3), 259–276.
- Finkel, S. E. (1985). Reciprocal Effects of Participation and Political Efficacy: A Panel Analysis. *American Journal of Political Science*, 29(4), 891–913. doi:10.2307/2111186
- Fowler, J. H. (2006). Habitual Voting and Behavioral Turnout. *The Journal of Politics*, 68(2), 335–344.
- Franklin, M. N. (2004). *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge ; New York: Cambridge University Press.
- Gerber, A. S., Green, D. P., & Shachar, R. (2003). Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment. *American Journal of Political Science*, 47(3), 540–550.
- Grofman, B., Collet, C., & Griffin, R. (1998). Analyzing the turnout-competition link with aggregate cross-sectional data. *Public Choice (1986-1998)*, 95(3-4).
- Gosnell, H. F. (1927). *Getting out the vote; an experiment in the stimulation of voting*. Chicago, Ill: The University of Chicago press.
- Green, D. P., & Shachar, R. (2000). Habit Formation and Political Behaviour: Evidence of Consuetude in Voter Turnout. *British Journal of Political Science*, 30(4), 561–573.

- Highton, B. (1997). Easy Registration and Voter Turnout. *The Journal of Politics*, 59(2), 565–575.
- Huckfeldt, R., & Sprague, J. (1992). Political Parties and Electoral Mobilization: Political Structure, Social Structure, and the Party Canvass. *The American Political Science Review*, 86(1), 70.
- Lake, R. L. D., & Huckfeldt, R. (1998). Social Capital, Social Networks, and Political Participation. *Political Psychology*, 19(3), 567–584.
- Lally, P., Van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6), 998–1009.
- Lally, P., Wardle, J., & Gardner, B. (2011). Experiences of habit formation: A qualitative study. *Psychology, Health & Medicine*, 16(4), 484–489.
- Miller, W. E. (1996). *The new American voter*. Cambridge, MA: Harvard University Press.
- Haberman, M., & Shultheis, E. (2012). Mitt Romney's pollster, Neil Newhouse's Moment. *POLITICO*. Retrieved from <http://www.politico.com/news/stories/1112/83288.html>
- Plutzer, E. (2002). Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood. *The American Political Science Review*, 96(1), 41–56.
- Verba, S. (1972). *Participation in America: political democracy and social equality*. New York: Harper & Row.
- Wolfinger, R. E. (1980). *Who votes?* New Haven: Yale University Press.
- Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114(4), 843–863.

## Tables

Table 1	Turnout	
	Presidential	Midterm
<b>Race</b>		
White	73.8%	53.7%
Black	68.2%	43.2%
Other	69.0%	42.5%
Unknown	76.1%	48.9%
<b>Gender</b>		
Female	73.3%	49.0%
Male	72.7%	52.1%
<b>Age</b>		
18-34	61.2%	29.1%
35-49	73.4%	49.0%
50-64	80.0%	62.1%
65+	82.6%	69.2%
<b>Partisanship<sup>1</sup></b>		
Democrat	80.6%	63.9%
Republican	83.7%	68.8%
Non-Partisan	67.4%	40.5%
<sup>1</sup> partisanship is defined as having voted in at least 1 primary for the Party		

Last Midterm	Midterm		Total
	N	Y	
Y	13%	40%	54%
N	31%	15%	46%
Total	45%	55%	100%

**Table: 2**

N = 275,330 person-elections

Type I habit	N		Y	Total
	Low	High		
	37%	8%	24%	61%
			32%	39%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	40%	15%	60%	100%
			85%	100%
N	72%	40%	28%	100%
			60%	100%
Total	45%		55%	100%

Midterms/Type III habit	N		Y	Total
	Low	High		
	39%	6%	31%	69%
			25%	31%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	34%	17%	66%	100%
			83%	100%
N	68%	50%	32%	100%
			50%	100%
Total	45%		55%	100%

Type II Habit	N		Y	Total
	Low	High		
	38%	7%	26%	64%
			30%	36%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	39%	14%	61%	100%
			86%	100%
N	71%	40%	28%	100%
			60%	100%
Total	45%		55%	100%

August Primary in Midterm	N		Y	Total
	Low	High		
	38%	7%	26%	64%
			30%	36%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	39%	14%	61%	100%
			86%	100%
N	72%	41%	28%	100%
			58%	100%
Total	45%		55%	100%

Past Presidential	N		Y	Total
	Low	High		
	34%	10%	15%	49%
			41%	51%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	48%	16%	52%	100%
			84%	100%
N	80%	35%	20%	100%
			65%	100%
Total	45%		55%	100%

August Primary in Presidential	N		Y	Total
	Low	High		
	36%	8%	27%	63%
			29%	37%
Total	45%		55%	100%

	N		Y	Total
	Low	High		
Y	36%	15%	64%	100%
			85%	100%
N	73%	44%	27%	100%
			56%	100%
Total	45%		55%	100%

TABLE: 3	Difference in Differences Logit Results using Habit Type I							
	Model 1		Model 2		Model 3		Model 4	
	b	se	b	se	b	se	b	se
<b>Failed to Vote Habit</b>	-1.4473	0.0030	-1.5551	0.0031	-1.5684	0.0031	-1.5874	0.0032
	0.1816	0.0003	0.1903	0.0004	0.1914	0.0004	0.2043	0.0004
<b>Interaction</b>	0.1035	0.0007	0.0906	0.0007	0.0937	0.0007	0.1085	0.0007
<b>Age</b>			0.0065	0.0001	0.0069	0.0001	0.0060	0.0001
<b>1998 Election</b>			0.7411	0.0028	0.7377	0.0028	0.7139	0.0028
<b>2006 Election</b>			0.3591	0.0026	0.3517	0.0026	0.3285	0.0026
<b>2010 Election</b>			-0.7425	0.0026	-0.7589	0.0026	-0.8234	0.0026
<b>Race: Caucasian</b>					-0.1952	0.0028	0.0490	0.0031
<b>Race: Other</b>					-0.0620	0.0038	0.0340	0.0040
<b>Race: Unknown</b>					-0.0882	0.0034	0.1784	0.0041
<b>Gender: Male</b>					0.1165	0.0019	0.1233	0.0019
<b>County of Origin</b>							<OMITTED>	
<b>_cons</b>	-0.1312	0.0023	-0.4193	0.0042	-0.3467	0.0048	-0.8022	0.0100
<b>N_clust</b>	2,875,402		2,875,402		2,875,402		2,875,402	
<b>AIC</b>	8,458,877		8,007,234		7,995,989		7,853,246	

TABLE: 4	Difference in Differences Logit Results using Habit Type II							
	Model 1		Model 2		Model 3		Model 4	
	b	se	b	se	b	se	b	se
<b>Failed to Vote Habit</b>	-1.3887	0.0024	-1.5160	0.0025	-1.5264	0.0025	-1.5132	0.0026
	0.2225	0.0004	0.2174	0.0004	0.2187	0.0004	0.2387	0.0005
<b>Interaction</b>	0.1049	0.0009	0.0882	0.0009	0.0921	0.0009	0.1153	0.0009
<b>Age</b>			0.0098	0.0001	0.0103	0.0001	0.0094	0.0001
<b>1998 Election</b>			0.8586	0.0028	0.8565	0.0028	0.8400	0.0028
<b>2006 Election</b>			0.5265	0.0026	0.5215	0.0026	0.5153	0.0026
<b>2010 Election</b>			-0.5216	0.0025	-0.5341	0.0025	-0.5787	0.0025
<b>Race: Caucasian</b>					-0.1950	0.0028	0.0691	0.0031
<b>Race: Other</b>					-0.0838	0.0038	0.0178	0.0040
<b>Race: Unknown</b>					-0.0968	0.0033	0.1473	0.0040
<b>Gender: Male</b>					0.0992	0.0018	0.1068	0.0018
<b>County of Origin</b>							<OMITTED>	
<b>_cons</b>	0.1612	0.0019	-0.3741	0.0041	-0.2897	0.0047	-0.7361	0.0097
<b>N_clust</b>	2,875,402		2,875,402		2,875,402		2,875,402	
<b>AIC</b>	8,582,360		8,153,824		8,144,252		7,994,614	



TABLE: 5	Difference in Differences Logit Results using Habit Type III							
	Model 1		Model 2		Model 3		Model 4	
	b	se	b	se	b	se	b	se
<b>Failed to Vote Habit</b>	-1.0414	0.0029	-1.0448	0.0030	-1.0460	0.0030	-1.0564	0.0030
	0.5435	0.0014	0.6096	0.0015	0.6101	0.0015	0.6208	0.0015
<b>Interaction</b>	0.1039	0.0026	0.0272	0.0026	0.0308	0.0026	0.0818	0.0026
<b>Age</b>			0.0164	0.0001	0.0166	0.0001	0.0165	0.0001
<b>1998 Election</b>			0.8349	0.0028	0.8354	0.0028	0.8318	0.0028
<b>2006 Election</b>			0.4111	0.0026	0.4102	0.0026	0.4079	0.0026
<b>2010 Election</b>			-0.6725	0.0026	-0.6765	0.0027	-0.7017	0.0027
<b>Race: Caucasian</b>					-0.0915	0.0027	0.0875	0.0030
<b>Race: Other</b>					-0.0852	0.0038	-0.0326	0.0039
<b>Race: Unknown</b>					-0.0436	0.0033	0.1329	0.0039
<b>Gender: Male</b>					0.1132	0.0017	0.1171	0.0017
<b>County of Origin</b>							<OMITTED>	
<b>_cons</b>	0.1604	0.0026	-0.7507	0.0041	-0.7395	0.0047	-1.0613	0.0091
<b>N_clust</b>	2,875,402		2,875,402		2,875,402		2,875,402	
<b>AIC</b>	9,118,332		8,554,686		8,548,732		8,474,529	

TABLE: 6	Difference in Differences Logit Results using Habit Type IV							
	Model 1		Model 2		Model 3		Model 4	
	b	se	b	se	b	se	b	se
<b>Failed to Vote</b>	-0.9538	0.0044	-1.0360	0.0046	-1.0363	0.0046	-1.0249	0.0047
<b>Habit:</b>								
<b>Presidential</b>	0.8246	0.0017	0.8977	0.0017	0.8981	0.0017	0.9069	0.0017
<b>Midterm</b>	0.3716	0.0019	0.4038	0.0020	0.4071	0.0020	0.3350	0.0021
<b>August - Midterm</b>	0.1854	0.0015	0.1199	0.0015	0.1222	0.0015	0.1981	0.0016
<b>August - Presidential</b>	-0.0564	0.0019	-0.1439	0.0019	-0.1469	0.0019	-0.1393	0.0019
<b>Presidential Primary</b>	-0.1768	0.0019	-0.0970	0.0019	-0.0980	0.0019	-0.1035	0.0020
<b>Local Primary</b>	-0.1705	0.0016	-0.2113	0.0016	-0.2092	0.0016	-0.1963	0.0018
<b>Interaction:</b>								
<b>Presidential</b>	0.4154	0.0025	0.4257	0.0026	0.4251	0.0026	0.3950	0.0026
<b>Midterm</b>	0.2402	0.0034	0.2046	0.0035	0.2065	0.0035	0.1822	0.0035
<b>August - Midterm</b>	-0.0980	0.0024	-0.1260	0.0025	-0.1220	0.0025	-0.0786	0.0025
<b>August - Presidential</b>	-0.1832	0.0034	-0.0952	0.0036	-0.0988	0.0036	-0.0889	0.0036
<b>Presidential Primary</b>	-0.0898	0.0037	-0.1301	0.0039	-0.1271	0.0039	-0.1121	0.0039
<b>Local Primary</b>	-0.0873	0.0030	-0.1179	0.0031	-0.1156	0.0031	-0.0858	0.0032
<b>Age</b>			0.0110	0.0001	0.0114	0.0001	0.0116	0.0001
<b>1998 Election</b>			1.2983	0.0033	1.2962	0.0033	1.3053	0.0034
<b>2006 Election</b>			1.1575	0.0031	1.1528	0.0031	1.1779	0.0032
<b>2010 Election</b>			-0.0134	0.0031	-0.0257	0.0031	-0.0049	0.0032
<b>Race: Caucasian</b>					-0.1352	0.0028	0.0456	0.0032
<b>Race: Other</b>					-0.0575	0.0039	0.0345	0.0041
<b>Race: Unknown</b>					-0.0436	0.0034	0.0990	0.0041
<b>Gender: Male</b>					0.1276	0.0019	0.1363	0.0019
<b>County of Origin</b>							<OMITTED>	
<b>_cons</b>	-1.2738	0.0037	-2.4190	0.0053	-2.3970	0.0058	-2.7223	0.0104
<b>N_clust</b>	2,875,402		2,875,402		2,875,402		2,875,402	
<b>AIC</b>	7,584,284		7,137,441		7,129,948		7,017,866	

TABLE: 7	Measuring Bias in of Type I habit									
	2010		2006		2002		Bias/year		Adjusted Value	
	b	se	b	se	b	se	p	se	B <sub>true</sub>	se
<b>Failed to Vote</b>	-1.6867	0.0049	-1.7664	0.0062	-1.8624	0.0096	-0.022*	0.0012	-1.640*	0.0081
<b>Habit</b>	0.1783	0.0004	0.2174	0.0008	0.2405	0.0014	0.0078	0.0012	0.165*	0.0079
<b>Interaction</b>	0.0450	0.0008	0.1081	0.0013	0.2206	0.0019	0.0219	0.0036	-0.0071	0.0244
<b>Age</b>	0.0004	0.0001	-0.0014	0.0001	0.0109	0.0002	0.0013	0.0010	-0.0046	0.0069
<b>Race:</b>										
<b>Caucasion</b>	0.1449	0.0054	-0.2481	0.0061	0.4056	0.0065	0.0326	0.0755	-0.0948	0.5160
<b>Race: Other</b>	0.2752	0.0066	-0.1278	0.0078	0.1769	0.0085	-0.0123	0.0511	0.1819	0.3489
<b>Race:</b>										
<b>Unknown</b>	0.2964	0.0064	-0.2028	0.0079	0.2989	0.0096	0.0003	0.0722	0.1290	0.4935
<b>Gender: Male</b>	-1.2579	0.0158	0.0142	0.0036	0.1229	0.0041	0.1726	0.0840	-1.4092	0.5736
<b>County of Origin</b>	<OMITTED>		<OMITTED>		<OMITTED>		<OMITTED>		<OMITTED>	
<b>_cons</b>	-1.2579	0.0158	0.1344	0.0188	-0.6787	0.0222				
<b>N_clust</b>	2,875,402		2,114,654		1,653,674					
<b>aic</b>	2,779,839		1,974,862		1,510,981					

\*value significant at the 95% confidence interval given the model only has 1 degree of freedom

TABLE: XXXX	Measuring Bias in of Type IV habit									
	2010		2006		2002		Bias/year		Adjusted Value	
	b	se	b	se	b	se	p	se	b <sub>true</sub>	se
<b>Failed to Vote Habit:</b>	-0.9217	0.0078	-1.0968	0.0094	-1.2147	0.0162	-0.0366	0.0041	-0.8580*	0.0282
<b>Presidential</b>	1.0717	0.0026	1.0270	0.0041	0.7921	0.0058	-0.0350	0.0137	1.1733	0.0938
<b>Midterm</b>	0.3483	0.0027	0.4512	0.0045	0.3527	0.0096	0.0006	0.0145	0.3808	0.0993
<b>August - Midterm</b>	0.1809	0.0022	0.1529	0.0032	0.2989	0.0054	0.0148	0.0126	0.1224	0.0858
<b>August - Presidential</b>	-0.1636	0.0027	-0.2000	0.0042	-0.0404	0.0061	0.0154	0.0141	-0.2271	0.0966
<b>Presidential</b>	-0.2048	0.0026	-0.0828	0.0045	-0.2032	0.0064	0.0002	0.0175	-0.1648	0.1195
<b>Primary</b>	-0.1576	0.0026	-0.2433	0.0038	-0.2965	0.0055	-0.0174	0.0023	-0.1283	0.0160
<b>Local Primary</b>										
<b>Interaction:</b>										
<b>Presidential</b>	0.3289	0.0048	0.4317	0.0062	0.5108	0.0071	0.0227*	0.0017	0.2874*	0.0117
<b>Midterm</b>	0.1108	0.0052	0.1139	0.0075	0.3777	0.0114	0.0334	0.0188	0.0006	0.1285
<b>August - Midterm</b>	-0.0659	0.0041	-0.1029	0.0050	-0.1225	0.0066	-0.0071	0.0013	-0.0547	0.0086
<b>August - Presidential</b>	-0.1040	0.0066	-0.0627	0.0079	-0.1323	0.0084	-0.0035	0.0080	-0.0784	0.0547
<b>Presidential</b>	-0.1734	0.0066	0.0418	0.0090	0.0395	0.0097	0.0266	0.0157	-0.1904	0.1072
<b>Primary</b>	-0.0526	0.0056	-0.0683	0.0066	-0.0365	0.0072	0.0020	0.0034	-0.0645	0.0234
<b>Local Primary</b>										
<b>Age</b>	0.0092	0.0001	0.0063	0.0001	0.0145	0.0002	0.0007	0.0008	0.0060	0.0055
<b>Race: Caucasian</b>	0.1274	0.0058	-0.2652	0.0065	0.4271	0.0069	0.0375	0.0783	-0.1283	0.5349
<b>Race: Other</b>	0.2130	0.0073	-0.1853	0.0084	0.2106	0.0091	-0.0003	0.0573	0.0812	0.3915
<b>Race: Unknown</b>	0.2093	0.0069	-0.2287	0.0083	0.3529	0.0101	0.0180	0.0736	0.0035	0.5027
<b>Gender: Male</b>	0.2316	0.0033	0.0255	0.0038	0.1443	0.0043	-0.0109	0.0234	0.1993	0.1602
<b>County of Origin</b>	<OMITTED>		<OMITTED>		<OMITTED>		<OMITTED>		<OMITTED>	
<b>_cons</b>	-2.8675	0.0167	-1.3388	0.0207	-1.7615	0.0262				
<b>N_clust</b>	2,875,402		2,114,654		1,653,674					
<b>AIC</b>	2,401,402		1,796,929		1,401,171					

\*value significant at the 95% confidence interval given the model only has 1 degree of freedom

Figure 1: Age When Registered

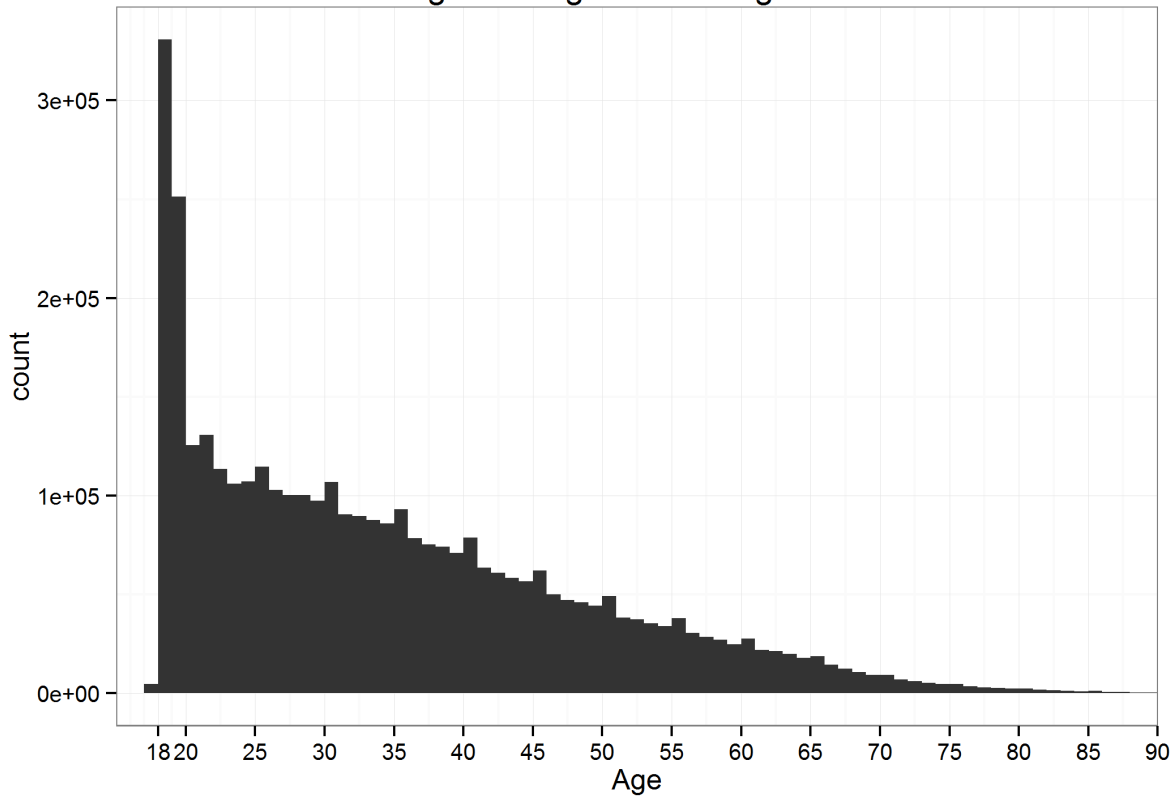


Figure 2: Comparing True Turnout from Voterfile

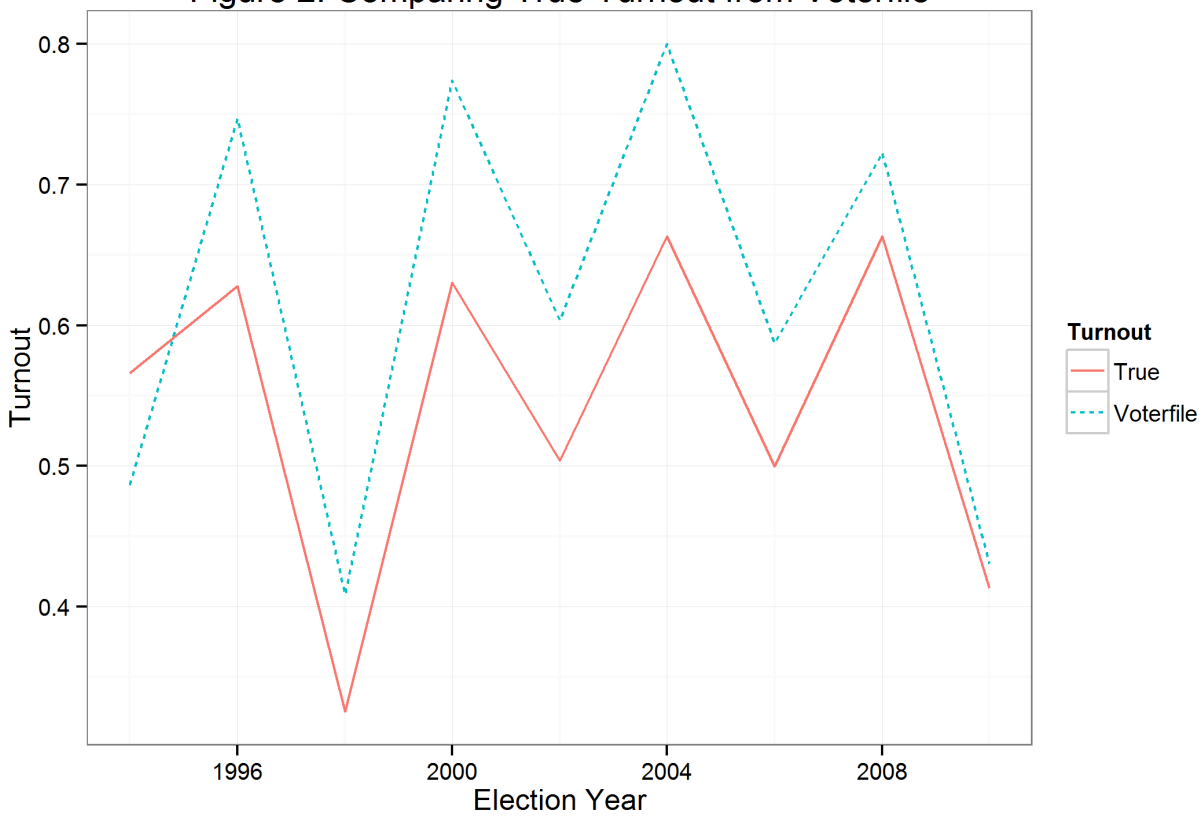


Figure 3: Probability of Voting - Model 1, Type I Habit

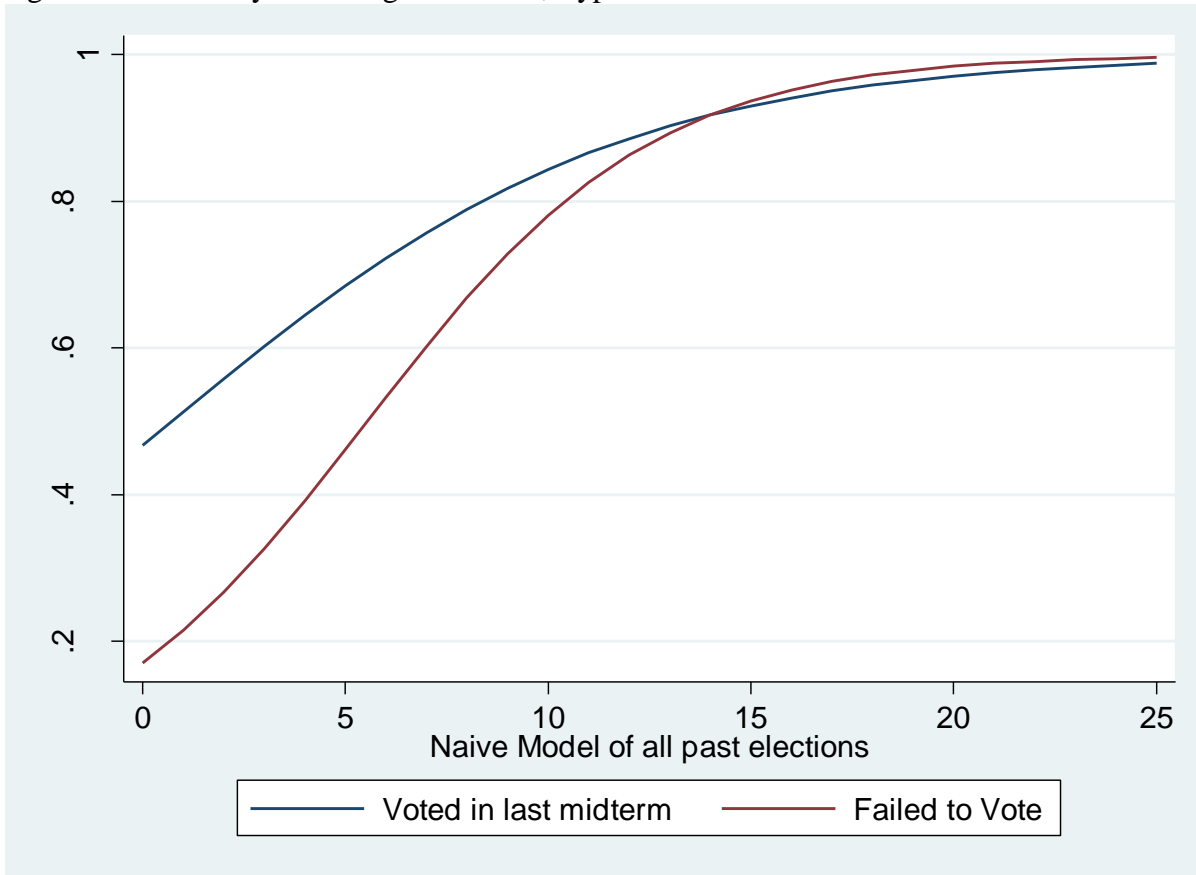


Figure 4: Probability of Voting For a Typical Voter- Model 4, Type I Habit

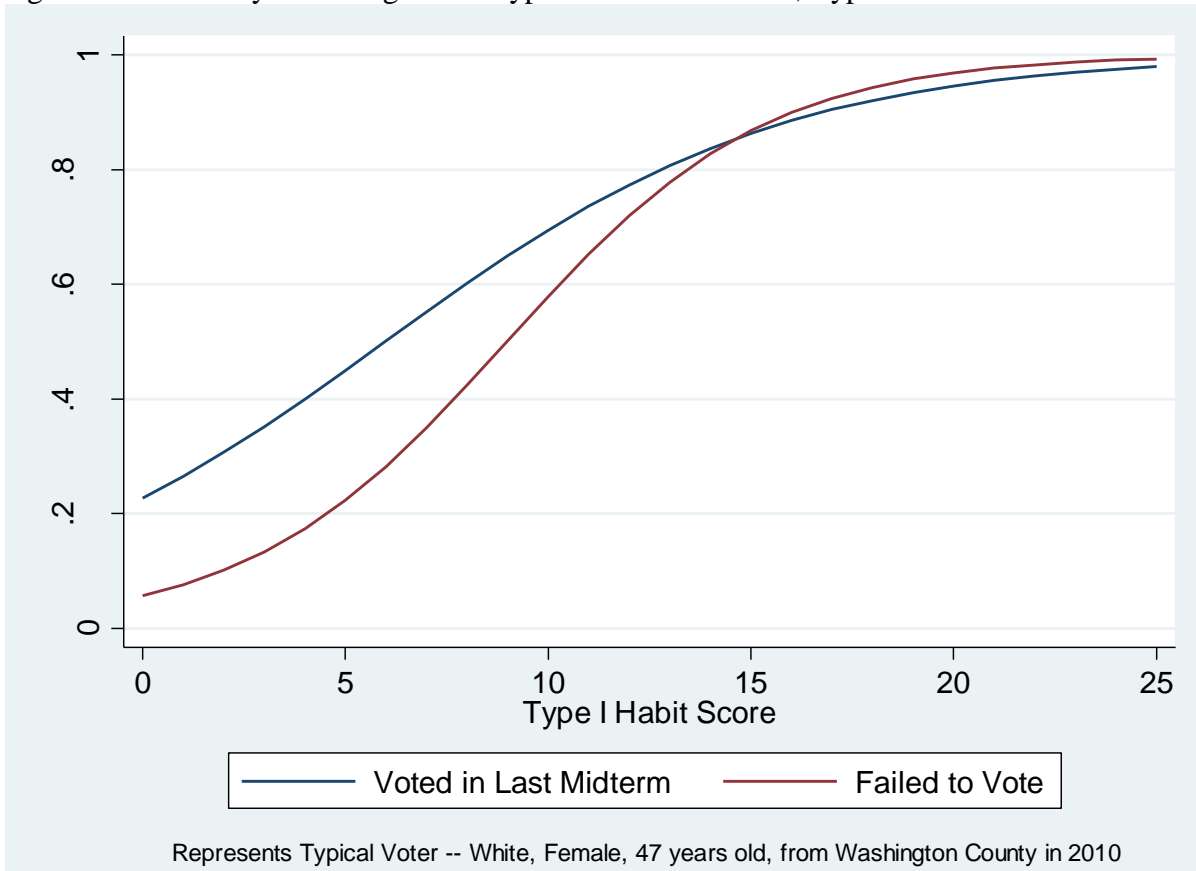


Figure 5: Probability of Voting - Model 1, Type II Habit

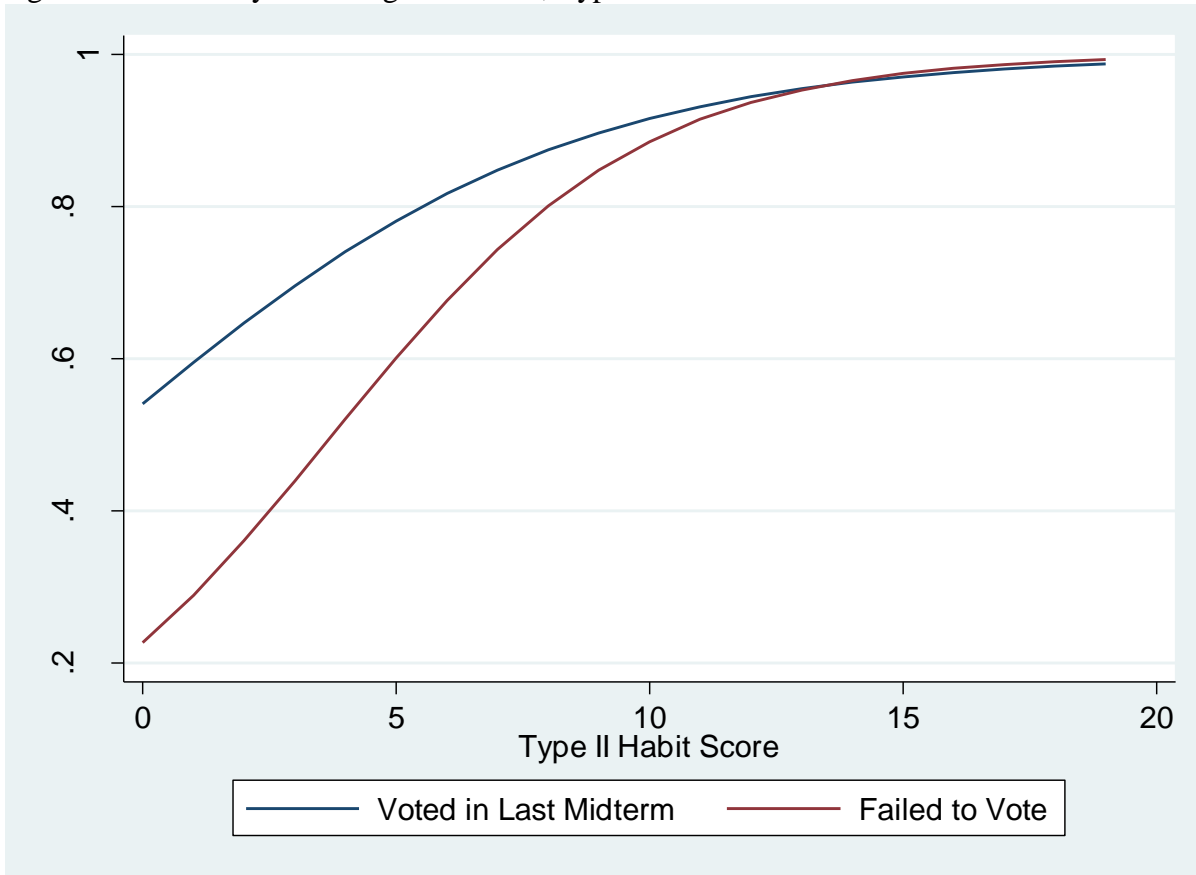


Figure 6: Probability of Voting For a Typical Voter- Model 4, Type II Habit

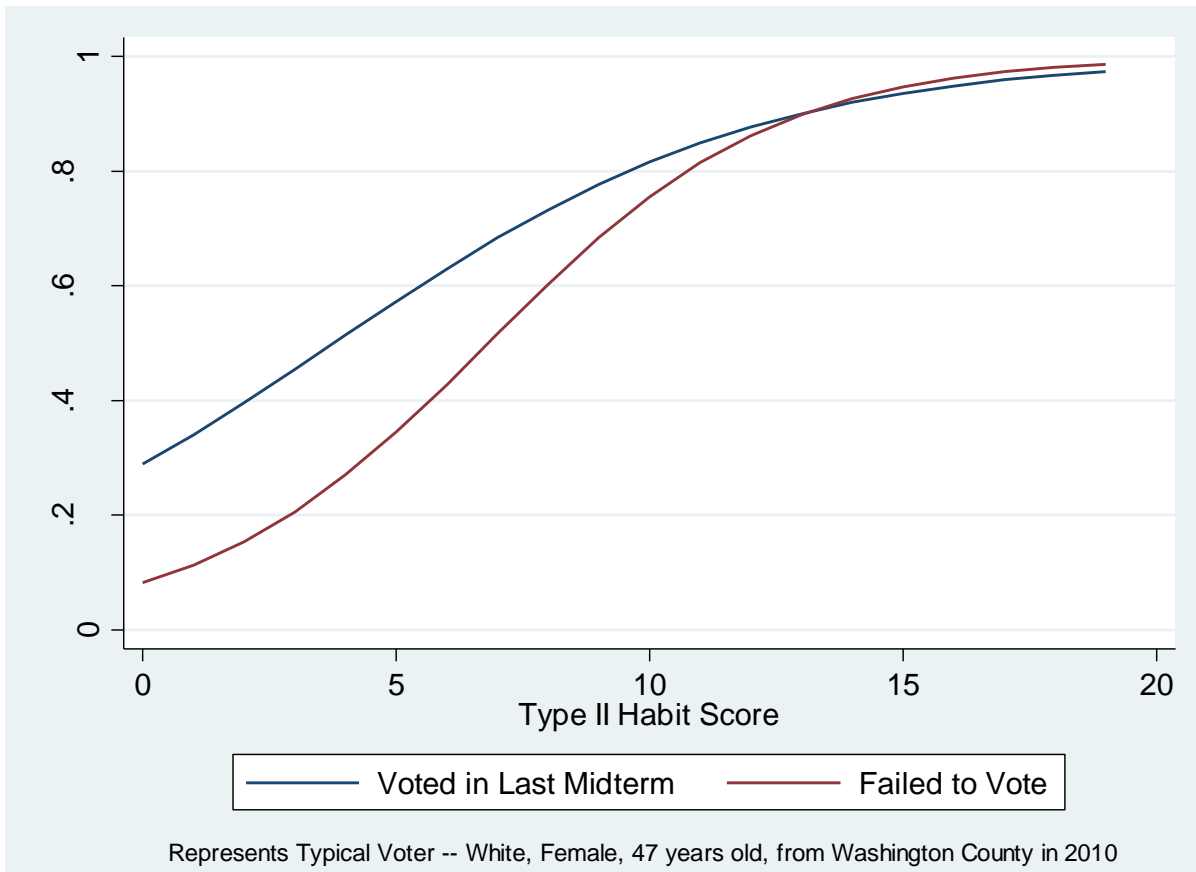


Figure 7: Probability of Voting - Model 1, Type III Habit

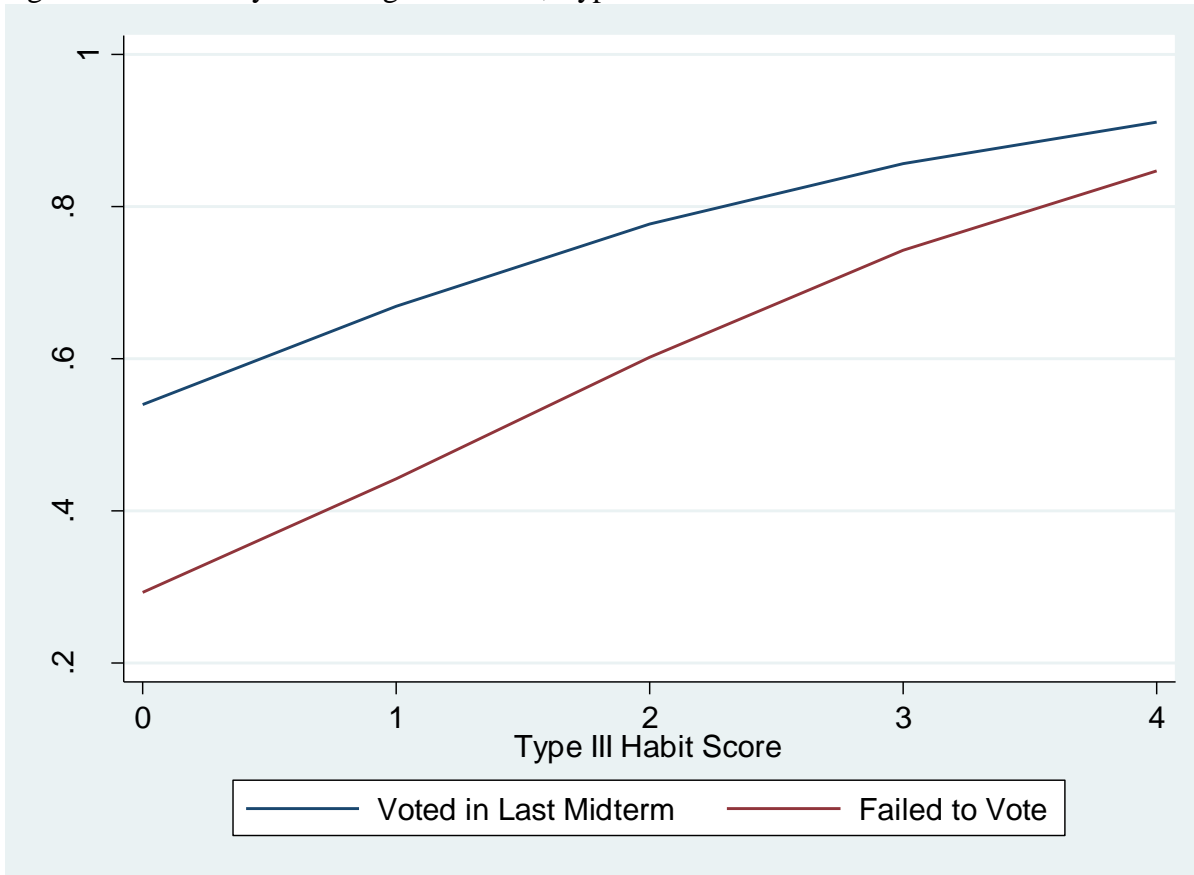


Figure 8: Probability of Voting For a Typical Voter- Model 4, Type III Habit

