**Chapter 2: Theoretical Utility and the Deductive Approach: Demonstrating BeSiVa with Different Theories of Turnout.**

By Benjamin Rogers

October 2016

**Introduction**

The previous chapter focused on the logistics of how BeSiVa worked and what could be expected from its use, concentrating on what it found when it was applied to synthetically crafted data, comparing algorithm's findings to the true model. A precise discussion of the algorithm's workings was provided, allowing for its recreation as desired. This explanation was necessary to demonstrate the algorithm's utility for making predictive models from relevant variables. While BeSiVa was incapable of finding every relevant variable in a single trial, the models it created made useful predictions, determining the correct value of the dependent variable and variables used in its creation in a certain percentage of cases. This chapter focuses on using BeSiVa deductively, answering a question of interest using real data and a collection of theoretically specified variables. Looking at the choice to vote, the most relevant theories from three different perspectives are tested and compared using the BeSiVa algorithm.

Having demonstrated the algorithm's capabilities at predicting synthetic data, this chapter turns to how BeSiVa can be applied to voter turnout and what a predictive approach may contribute when used deductively. The choice to vote has been considered from many different theoretical angles, with each one posing a differing explanation for why a person decides to turn out to vote. Through the algorithm, the different explanations, and therefore the reasons for including each independent variable, are weighed in comparison to one another, and a set of models using the algorithm's most commonly recommended variables are generated. These models are compared to theoretically specified models, each of which are tested using the same bootstrapping technique to determine how well they make predictions.

The BeSiVa Algorithm's ability to create useful models is based on the principles of predictive analytics, but it is used deductively here as an alternative to null hypothesis significance testing. Too often, the final arbiter for whether a variable matters or not is based entirely on its p-value (Schrodt 2014), which explains whether a variable is statistically significant. Statistical significance, however focuses purely on whether a variable's true value is zero, saying nothing about the substantive significance of that variable (Cohen 1994). Using statistical significance as the deciding factor in whether an independent variable is truly relevant leads to questionable research findings and outcomes that only appear to make advances in our understanding.

The concentration on statistical significance has led to a series of perverse incentives for researchers that lessen the quality of research and the utility of its findings, calling seemingly settled conclusions into question. These questions are raised for a variety of reasons, including when an individual seeking significance inadvertently or deliberately creates models that minimize p-values, even if the independent variables aren't necessarily relevant from a substantive perspective (Gelman and Loken 2013). There is also the problem of overstuffing models with an excess of independent variables, creating 'garbage can' models that show marginal significance for a variable of interest, violating parsimony (Schrodt 2014). Schrodt specifically points out that garbage can models trample on Achen's rule of three, which suggests that any model with more than three independent variables is incorrectly specified (Achen 2002). Researchers' focus on null hypothesis significance testing (hereafter NHST) as the final arbiter of truth has led to incorrectly validated hypotheses and models that are not parsimonious.

The models NHST incentivizes might be reasonable if they spoke to the utility of the theories they pretended to test, but this is highly unlikely given the way that the models are created and implemented. Model utility lies in the ability to not only explain the data that are provided, but to predict new observations as well. Prediction provides a separate test for these models, a test which takes the research further down the trail of inference (Mosteller and Tukey 1977). Prediction informs researchers about the overall utility of their models, how useful they are in a more rigorous test of quality than statistical significance. Imagine a voter turnout model that achieves significance from an explanatory perspective, validating the hypotheses that it sets out to test. But at the same time, the turnout model is incapable of predicting new observations. Even if the hypotheses were demonstrated yielded statistical significance, a model of whether an individual turns out to vote would be useless if it could not reliably determine if a person whose data was not used in the model was likely to vote or not. Unlike statistical significance, prediction explicitly considers how the model performs on new voters. In this situation, prediction serves as a useful test of models and hypotheses, leading to results that are robust to repeated testing, making replication far more likely overall (Hindman 2015) In comparison to NHST, making predictions serves as a useful test of the substantive significance of models, and leads to conclusions that are likely to hold up in the long run.

While making predictions leads to a more rigorous test of hypotheses, failure to make predictions leads to results that are unlikely to replicate, ignoring a useful test of results and creating models that only appear to fit the data well. Attempting to capture all variation in a dependent variable, models tested via NHST may explain provided data well, but are also likely to overfit if new observations are considered. Imagine a model that, when data were provided, yielded the correct value of the dependent variable for every observation used in its creation. In doing so, however, this model failed to predict any observation that wasn’t included in the initial regression. This is the problem of overfitting, when a model is good at giving accurate results for data used to create the model, and incapable of giving an accurate value for data that are not included Using all of the data for estimation then makes it difficult to determine if a given model is useful, as overfitted models may predict provided values of the dependent variable well and fail to predict observations that were not included in estimation (Kuhn and Johnson 2013). The predictive perspective needs to be considered for several reasons including parsimony, theoretical utility, and its general usefulness for testing hypotheses, making an approach like BeSiVa useful for reconsidering hypotheses.

But in light of the utility of the predictive approach, how can a technique like BeSiVa be used in social scientific practice? Using a deductive approach, the BeSiVa algorithm is here used as a means of comparing several different theories of voter turnout. This is in contrast to multiple regression with NHST, which may allow for multiple theories' consideration, but does not choose between them, while NHST has systemic difficulties determining theoretical utility. Multiple theoretical approaches to turnout are developed briefly, before the variables that they specify to predict turnout are provided to the algorithm for testing. Once the algorithm has determined which variables are useful, the selected independent variables are added to models in order of their relevance, and these models tested using a separate bootstrapping approach.

The need for a better way to determining whether a theory is not only true, but useful, is called for. This chapter concentrates on using BeSiVa to make that determination among a selection of theories taken from the literature on voter turnout. It is determined that of these theories, psychological, sociological, and mobilization-based explanations, the BeSiVa algorithm's predictions favor the sociological and psychological explanations. Elements of these theories are found to be most useful for making a prediction, especially education, the strength of party identification from psychology, as well an individual's age, and the time that someone lives in a specific location. The comparison allows for a better understanding not only of what drives an individual to vote, but of prosaic elements that are too often ignored and suggest new directions for research on voter turnout.

**A Short History of Voter Turnout**

**Psychological Origins and The Michigan School**

The choice to vote has a lengthy history within political science, and the multiple causal mechanisms that social scientists suggest drive the decision to turn out makes the area an ideal candidate for using a predictive algorithm deductively. Scholars have debated why people vote for over fifty years, leading to a variety of different theoretical models which could use comparison, beginning with the Michigan School. The founders of the Michigan school determined that individuals are more likely not only to make a choice in voting, but to vote in the first place due to their affinity for a political party. Party identification served as an indicator of interest, knowledge, and concern for the outcome of contests where the party a voter identified with was involved (Campbell et al. 1960). Campbell et al.'s original study of who individuals vote for, focusing on the role of party identification, provided a starting point for decades of research on the question of why people turn out to vote.

**Sociological Theories**

Although they provided a strong explanation of why a person might decide to vote, Campbell et al. admitted to a potential lack of primacy in their causal reasoning, suggesting that no single variable could explain all aspects of the choice to vote. For this reason, voter turnout could be described as governed by other elements of a potential voter's life, a possibility first examined by Berelson, Lazarsfeld and McPhee. In their consideration of Elmira, New York, Berelson et al. concentrated on why individuals choose to vote in a specific way, but they also spoke to the reasons for turning out as well. Noting that “[n]onvoting is related to persistent social conditions having little to do with the candidates or issues of the moment” (32), Berelson et al. discuss the differentiation between social groups and demographics, and were perhaps the first to note the relationship between education, social group involvement, and turnout (1954). Similar to Campbell et al. (1960) in their consideration of vote choice, Berelson et al.’s nascent sociological approach demonstrated a demographic component to explain why individuals turn out to vote, suggesting a new driver for political participation.

Although Berelson et al. considered the role of demographics in turnout, it was not their primary concentration in *Voting* (1954), and they shied away from explicit considerations of demographics in turnout. Instead, Berelson et al. concentrated on understanding vote choice through qualitative approaches, briefly considering why demographics might affect turnout. The decision to vote or not was examined further in later works, such as *Who Votes* by Wolfinger and Rosenstone. Using the United States Census' current population survey data, Wolfinger and Rosenstone posited that demographics -especially demographics that indicated resources that a voter possessed- could serve as a possible driver of the decision to turn out. Their findings suggested that from the sociological perspective, demographics that indicate resources are key to determining whether an individual is likely to vote.

Wolfinger and Rosenstone demonstrated the sociological approach by comparing a variety of demographic trends. In doing so, they showed that contrary to the class based suggestions of earlier researchers (Schattschneider 1975), the key resources to turning out were based in the abilities those demographics yielded for navigating bureaucracy and understanding voting. The role of these resources was compared along with the instrumental, expressive, and interest based benefits of voting, and the authors found that certain resource-based demographics mattered more than others. Due to its role as a provider of the ability to work with government, education mattered dramatically, imparting skills and experience necessary to maneuver a government bureaucracy with ease. Similarly, age could impart the ability to navigate political situations, providing the experience to act politically even if a person had little formal education. In addition, whether someone was married mattered; marriage provided a strong interpersonal pressure to vote. In a paradoxical finding, an individual's wealth and free time, resources in the more traditional sense of the word, were not particularly indicative of a person's proclivity to vote. Wealth only mattered if the potential voter was unable to attain a level of comfort, and an individual's free time did not matter at all. If a person's occupation affected likelihood of turnout, it was only in terms of their interaction with government due to that job, making farmers more likely voters than other similar people. The key question then, in Wolfinger and Rosenstone's theoretical formation, was whether an individual had necessary resources, especially knowledge and experience, in order to vote (1980). While this element of cognitive resources is demonstrated, and their use of census data is groundbreaking, Wolfinger and Rosenstone were also limited by their reliance on the Census. The use of census data makes it difficult to consider alternative hypotheses of turnout, a limitation which hampers Wolfinger and Rosenstone’s contribution to understanding political participation.

While the authors of the Michigan school were willing to admit their limitations, suggesting roles for life experiences, Wolfinger and Rosenstone's reliance on census data makes it difficult to control for alternative hypotheses of turnout, such as party identification’s role in increasing the likelihood of turning out to vote (Campbell et al. 1960). This inability to test alternatives leads to a failure to consider partisanship, limiting the utility of *Who Votes* due to its inability to test theoretically established alternative hypotheses. Wolfinger and Rosenstone perceive the contribution of the Michigan school, but they are unable to control for its findings in their own research. In addition, Wolfinger and Rosenstone mention a collection of causal mechanisms, including the roles of possible drivers of turnout such as a person's interest as well as instrumental and expressive benefits, but they fail to include any of those potential causal mechanisms beyond acknowledging their existence. For these reasons, Wolfinger and Rosenstone branch out from work that came before, but the limitations of their data makes it impossible to control for non-demographic alternatives, making difficult to describe *Who Votes* as building on prior research.

While they posit an alternative portrait of voting to the Michigan school, Wolfinger and Rosenstone are limited by their data, a problem which required a reconsideration of voter turnout. In his attempt to determine why voting declined in the aggregate, Teixeira focuses on individual turnout, and in doing so manages to overcome some of the limitations that hampered Wolfinger and Rosenstone. Exploring the findings of prior research while also further testing the sociological perspective's implications, Teixeira describes how links to party, the state, and media declined between 1964-1980, leading to turnout's overall decline. Despite this focus on the aggregate, Teixeira created an expansive theoretical picture of individual turnout, one which overcame the difficulties of the Wolfinger and Rosenstone's findings while illustrating a new difficulty of the turnout literature.

To Teixeira, each links between an individual and overarching institutions, the parties, state, and media represented a connection to different aspects of the political process. Links to party led individuals to have an interpretive framework for the issues, and made election outcomes personal. Links to media similarly gave an interpretive framework to the issues and provided independent meaning to both the election and the issues a person cared about. And links to the state gave individuals motivation to vote due to their self-perception as part of an influential group. The decline of each of these potential causes at the individual level led to a collapse in voter turnout in the aggregate, despite the increase in education, which should lead to increased participation (Wolfinger and Rosenstone 1980), which was considered alongside the controls that Wolfinger and Rosenstone and others suggested as relevant. This allowed for the consideration of the hypotheses of several different approaches, but did so at a cost to the overall work's parsimony and interpretability.

Despite Teixeira's consideration of a wealth of causes, it's not clear that each potential theoretically driven causal mechanism is given a fair hearing. Race, region and sex were all dismissed, included as controls despite the author's skepticism about their relevance and the literature's theoretical explanations for their inclusion. This is reasonable, however, given the later findings of Verba et al. who suggested that race did not matter in questions of turnout (1993). In this case the work fell victim to a requirement of the literature, whose discussion of an excess of causes led to a set of models with an excess of variables. But what else could be done? The specification of such models and findings of significance for these variables indicated a need for additional predictors. Even as the models grew to sizes that made their findings difficult to parse, the list of theoretical causes continued to grow, decreasing the likelihood of paring down the theory to a manageable set of explanations.

**Mobilization Theory**

In addition to the challenges of the prior literature, and the failure to trim down the list of previous causes, the addition of further causal mechanisms expanded the literature, branching out without cutting down on the number of causes. Such an addition was made by Rosenstone and Hansen, who suggested a new causal mechanism to explain turnout. While an individual's political efficacy and resources are important to the mobilization model, Rosenstone and Hansen's contribution was the inclusion of political actors who could influence the individual. An individual can participate of their own volition, but if the individual lacks the resources necessary to participate, a political organization may step in to bear individual costs of participation. The act of direct contact by one of these organizations lowers the costs of participation for a citizen to the point where they may be able to participate. Rosenstone and Hansen concentrate on political parties for much of their work, but concede that parties are one of many potential mobilizers. In the mobilization model, social organizations such as civil rights groups and unions are also capable of mobilizing voters (2003), allowing other organizations to drive an individual to participate in politics.

In addition to the direct mobilization that political parties provide, Rosenstone and Hansen suggest that citizens are also mobilized indirectly. Drawing partially on the work of Olsen (1971), Rosenstone and Hansen posit that politically active friends and neighbors may decide to treat voting as a collective action problem. A voter embedded in a network with such individuals receives selective rewards for participation, as well as selective penalties for failure to participate in the electoral process. For this reason, individuals embedded heavily in social networks, indicated by mechanisms such as employment, social group participation, and class, are more likely to be mobilized to participate (Rosenstone and Hansen 2003). For the mobilization model, the act of voting is driven not only by resources and efficacy, but by the mobilization created by political organizations and other actors.

Apart from the concerns of demographically oriented studies, Rosenstone and Hansen contribute a new causal mechanism, avoiding the pitfalls of the literature that concentrated on an increasingly large subset of demographic groups. Despite this necessary contribution, however, mobilization theory leads to a series of other problems. While direct mobilization may be easily captured by questioning people about political contact, indirect mobilization remains challenging to operationalize. The authors argue for social group activity as an indicator, but both mobilization and social group participation may be driven by separate causes. Perhaps the personality trait that leads people to be outgoing may serve as a driver of political activity, or social groups are a natural target for direct mobilization by political groups, making the individual a subject of direct mobilization regardless. While the potential problems of operationalizing these aspects of mobilization cast doubt on the theory, the main concern remains the addition of extra causal mechanisms, as opposed to negating the large body of causes that already exist.

**Additive Contributions and Habitual Voting**

While there has been some effort to pare down the number of theoretical causes of voting, such as the research of Verba et al. (1993), the habit of the literature appears to be the expansion of theoretical causes. Rather than attempting to pare down theoretically specified causal mechanisms, creating a lean but highly relevant set of predictors in the truest sense of the word, researchers focus on adding causes. The primary goal of the turnout literature is to find alternative explanations, even as this adds to the list of possible reasons a 'true model' would require to explain why someone chooses to vote. This trend can be seen in an offshoot of the psychological literature of the choice to vote, which suggests that the turnout is habitual (Brody and Sniderman 1977, Gerber, Green, and Shachar 2003). In some cases, this leads to sidestepping the theoretical underpinnings of turnout and habit entirely, even mathematically simulating multiple psychological driving mechanisms behind the choice to vote (Fowler 2006). This contribution suggests the true eschewing of theoretical causes, in favor of unnecessary methodological novelties.

While “voting is for many a habit”, as Brody and Sniderman describe it (1977, 349), the literature demonstrates a struggle to explain the underlying reason that habit affects an individual's choice to vote. When it is considered, the theoretical drivers of the choice to vote are explained by multiple causal phenomena, without necessarily suggesting any convincing explanation. In the case of Gerber, Green, and Shachar, individuals have a psychological impetus to vote over time, one which is verified experimentally (2003). While tracing the idea that voting is a habitual behavior to Aristotle, however, Gerber, Green, and Shachar never isolate the reasoning behind habitual voters, instead identifying a wealth of causes with a verifiable effect.

The habitual choice to vote is experimentally verified by Gerber, Green, and Shachar, but their explanation of why voting is habitual is lacking, due to their ambiguous stance on theoretical reasoning behind their findings. Voting may be habitual due to individuals' enduring response tendencies; it may be that the same stimuli cause the same result again and again, due to the fact that voters exist in a persistent electoral environment; or it may be due to a self-reinforcing effect, where the choice to vote in one instance leads to an increased likelihood of voting in the future. Of these three explanations, the first two appear to be akin to one another. Individuals vote due to their enduring response tendencies, something which is consistently stimulated within a persistent electoral environment. Meanwhile, the third explanation is a restatement of the original idea, that individuals vote because voting is habit forming. The idea of voting as habitual may be buoyed by experimental evidence, making it difficult to argue that voting is habit forming, but at the same time, the explanation for why voting is habit forming is lacking, a difficulty that makes considering the habitual nature of voting challenging.

Despite the difficulty in parsing the causal mechanisms behind the idea of voting as a habit, there is precedent in demonstrating a new method on a question related to the choice to vote. Fowler, for instance, demonstrates a series of techniques through simulation designed as a formal model of the choice to vote (2006). Beginning with a critique of the simulation effort by Bendor, Diermeier, and Ting (2003), Fowler suggests a correction to their behavioral model that better captures the habitual nature of voting. By adjusting feedback of voting, Fowler creates a model that captures the possibility that an individual may not vote, while also capturing a majority of simulated voters' behavior.

By correcting the feedback simulation of Bendor, Diermeier, and Ting, Fowler appears to capture the behavior of voters as it appears outside the simulation. It may be, however, that Fowler is talking past Bendor, Diermeier and Ting, due to the possibility of different ways of considering voter turnout, especially if different types of elections are factored in. To use an example, an individual may vote in a specific subset of all elections such as elections for president, satisfying the criterion that Fowler uses as per the theoretical concept of voting as a habit. For this reason, it may be that the two researchers each have much to contribute, but if different conceptions of turnout are differentiated.

When considering how Fowler, in comparison to Bendor, Diermeier and Ting, simulate the choice to vote, it may be that Fowler created the more realistic simulation. Alternatively, Bendor, Diermeier and Ting's simulation may better reflect the large number of elections that voters could participate in, and attempt to simulate a voter's turnout, perhaps in Presidential elections but eschewing smaller electoral contests such as local or special elections. The two differing conceptions of the choice to vote are based purely in simulation, making it difficult to determine whether such different considerations of vote choice are more accurate for each simulation. Although it is possible that only one of these approaches represents a more accurate way of modeling vote choice, multiple types of turnout may mean that for differing choices on voting, the simulations are appropriate at different intervals.

Regardless of whether voting is a habit, the main concern of including prior turnout in models of the choice to vote is its complete overlap with other potential predictors. An individual with increased education is more likely to turn out, as is a person who has a position whose class makes it easier for them to vote (Wolfinger and Rosenstone 1980, Teixeira 1987). The problem lies in determining whether the person's status as a voter last time differs from their status as an educated, wealthy person, affecting their likelihood of turning out in the same manner. The difficulty then, of determining whether voting is a habit from this data lies in the difficulty of disentangling these different ways of predicting whether a person votes or not. In addition, the theoretical underpinnings for habitual voting do not appear to be strongly specified, especially in comparison to the mobilization, psychological, and sociological theories of voting. For this reason, it is a good idea to focus on theoretically specified variables as a means of determining whether their explanations are sufficient to predict, rather than explain the choice to vote, leaving habitual turnout until a better theoretical explanation can be provided.

**A Literature That Adds, But Does Not Weigh Explanations**

In the consideration of each of these differing conceptions of the choice to vote, the problem of the literature is its inability to parse the different theories, to determine which of the conceptions of turnout best reflects what is, rather than what is theorized. This is not reflective of the work of theorists, whose contributions are invaluable, but the means by which the different theories are tested. One of the early critiques made of the social sciences in general, is why multiple phenomena cannot contribute to a topic of interest in equal measure. While the equivocation of different causal mechanisms is not particularly useful due to the fact that it adds little to the understanding of a research topic, the problem of figuring out which theoretical approach provides better predictors of a phenomenon remains unanswered.

While the literature is capable of developing endless theoretical explanations, the chance to force them to compete with one another remains untaken. Too often, the literature focuses on adding, rather than subtracting elements of the models, until giant lists of variables are necessary to suitably control for the collection of theoretically suggested causes. Such an approach to determining veracity falls decisively in opposition to parsimony exemplified in Achen's rule of three. To Achen, any model with more than three independent variables is meaningless due to poor specification, arguing that treating the groups within enormous models as something that can be controlled for through dummy variables is incorrect (2002). But the main thrust of the literature does not appear to be to answer Achen, either by arguing against his rule, or by trying to work within it, but to ignore his critiques of research methodology entirely (Schrodt 2014). For this reason, a different approach is needed to

While a few efforts, such as Verba et al. (1993) have been made to cut down on the number of variables that explain political participation, the main thrust of the literature is to continue expanding lists of relevant variables. The literature focuses on building, rather than sculpting the collection of variables that should be included in understanding the turnout. While multiple causes may drive the choice to vote, and may even simultaneously explain parts of the variation in turnout, a predictive approach such as BeSiVa, similar to other variable selection approaches, allows for a comparison of variables equivalent to sculpting, rather than accumulating. BeSiVa compares the variables which are provided to it, using the information to create a model that best predicts subsets of data held out for the purpose of prediction.

**The Data**

In order to understand what drove voter turnout, the American National Election Study, or ANES, was selected as a source for data. The ANES serves as a logical choice to study turnout, especially at the level of individual voters. The study's time series cumulative data file allows for consideration of differing drivers behind the choice to vote ranging back to 1948. In addition, the data allows for a consideration of a large collection of potential drivers of turnout in a manner that lets BeSiVa make a comparison of each predictor's relevance. Once any category deemed missing was recoded appropriately, it was necessary to consider what data should be chosen to use with the algorithm. While BeSiVa could work on any data set, a selection was needed, and the 2000 election was selected at random from a set of possible years. Like the other possible datasets, the ANES 2000 survey contained the variables suggested by theory, had a sizable number of observations, and was readily available.

**The Dependent Variable**

Given the focus on the choice to vote, the operationalization of the dependent variable was relatively straightforward. In one of their survey responses, the ANES asked whether an individual voted in a given election, which was used as a proxy for their behavior. There is an understanding, however, that an individual might decide not to vote, but when queried, would specify that they did due to response bias, (Belli, Traugott, and Beckmann 2001, Tourangeau and Yan 2007). Despite this potential drawback of using survey responses rather than voting records, the risk of response bias is accepted as a limitation of the design, and is irrelevant to the algorithm's ability to predict. Given that the main point of using turnout is to test the algorithm, the risk of response bias in the dependent variable is secondary to the utility of the predictions created. Missing data were properly recoded, and the survey responses were changed into a numeric dichotomous variable with 1 for yes and 0 for no, but no other changes were made to the dependent variable. Having selected turnout as a dependent variable, the algorithm only needed a series of independent variables to create a predictive model of the choice to vote.

**Independent Variables**

Having determined that self-reported choice to vote would serve as the dependent variable, multiple independent variables were provided to BeSiVa for its consideration. The first variable that was included for consideration was an individual's party identification. Dating back to the Michigan School, party identification may affect an individual's likelihood to turn out due to interest and attachment to the outcome (Campbell et al. 1960, Dalton and Wattenberg 1993). In the ANES data, party identification was measured as a 7 item categorical variable, ranging from strong partisan, weak partisan, independent but leaning towards a party, and truly independent. For the algorithm's consideration, and due to the fact that party strength, rather than party, is theorized to lead a person to turn out to vote, however, party identification was recoded as a numeric variable capturing the strength of an individual's identification. This operationalization of the strength of party ID ranged from 0 for independents to 3 for strong partisans of either party. Thus, party identification, long associated with turnout, was included in a manner consistent with the component expected to correspond to turnout behavior, along with other variables suggested by the literature.

In addition to party identification, a collection of variables based off the opinions of sociological theorists were included, starting with education. Operationalized based on if someone had achieved certain levels of schooling, Wolfinger and Rosenstone theorize that education increases how much individuals pay attention to politics. Education makes politics more enjoyable to follow, as further additional education makes it easier to understand the rules and impact of politics (1980). It has also been suggested that education aids individuals in maneuvering the mechanics of voting (Teixeira 1987), or enables the individual to better parse political information (Rosenstone and Hansen [1993] 2003). Education was included in the ANES as a 6 category factor variable, ranging from finishing grade school to advanced degrees, with values in between to capture whether someone had finished high school or a bachelor's degree. The variable, however, was taken and transformed into a numeric variable, one which captured a person's level of education as a continuous predictor. This operationalization of education was relatively straightforward, and its multiple backing theories and relative agreement made it a necessary addition to the list of variables to consider.

Education was one example of a variable with multiple proposed causal mechanisms related to turnout behavior, as was age. Age was hypothesized to affect voter turnout as a proxy for political experience, which could substitute for education (Wolfinger and Rosenstone 1980). It was also a potential driver due to the existence of shared generational experience or differing drives at stages of the life cycle (Rosenstone and Hansen [1993] 2003). Age was asked about on the ANES as a numeric variable, and was provided to the algorithm in its unrecoded form, as well as an age squared term, to capture possible non-linearities. This meant that age was included based on how old the individual was at the time of the survey, a continuous predictor of an individual's experience.

In addition to variables that indicated a person's cognitive resources, such as age and education, the literature described individuals' connections to media as a possible determinant of turnout behavior. Teixeira suggested that links to the media, operationalized based on how often someone read the newspaper, allowed voters to be more engaged, giving each election a sense of meaning and making an individual more likely to turn out (1987). The ANES inquired about newspaper reading directly, and the number of days per week an individual read the paper was provided to the algorithm as a possible predictor of turnout. The number of days reading the paper was included as a numeric predictor, and ranged from none to seven. With this way of determining how politically connected through the media an individual was, the algorithm could capture these links' overall relevance.

Due to the fact that it has been defined in multiple ways, political efficacy was challenging to include for the algorithm's consideration. Political efficacy, based in the notion that an individual believes they influence the outcome, has been theorized to increase the likelihood of voting due to the increased sense of accomplishment that such a feeling provides (Teixeira 1987). It can also make things easier for a potential voter, with a sense of personal competence makes someone feel more comfortable participating in politics overall (Rosenstone and Hansen [1993] 2003). Political efficacy is operationalized using this sense of influence over the outcome, helping to determine its role in the prediction of whether a person votes by including it for the algorithm's consideration.

Despite the fact that it has been suggested as a key driver in some corners of the literature, the role of income in predicting turnout has been disputed. While such a resource may make individuals more capable of participating in politics (Schattschneider 1975), Wolfinger and Rosenstone suggested that income only mattered to the point of comfort (1980). Teixeira, by comparison, suggested that greater incomes were capable of easing the challenge of voting as a component of socioeconomic status (1987), while Rosenstone and Hansen suggested that it would make participation more likely, due to an increased likelihood of sharing social circles with the political class ([1993] 2003). To try and capture income's potential role, the income categories of the ANES were recoded as a numeric predictor, treating the different selections of the respondent's income quantile as a driver of political participation. Once it had been recoded, its ability to capture whether income could potentially predict turnout made it an invaluable addition to the list of variables for consideration.

The use of race as a predictor of whether someone turns out to vote has been disputed (Verba et al. 1993) as are individuals' sex and the region in which someone lives (Teixeira 1987). Their roles as operationalizations of historical privilege and political culture (Wolfinger and Rosenstone 1980, Teixeira 1987), however, and their disputed status meant that these demographic predictors should be included in the list of variables to consider. Race was recoded from a 7 point scale based on a self-identification of ethnicity, which was changed to a dichotomous variable based on whether an individual was in the minority. Region was a 4-fold classification based on whether someone lived in the northeast, north central, south, or west of the United States. Similarly, sex was included for consideration as a binary dichotomous categorical variable. Given their disputed status, these variables are ideal for a predictive algorithm, allowing for their consideration in a more systemic fashion.

Unlike the disputed status of race and other demographic predictors, marital status is a relatively uncontroversial addition to the list of independent variables for the algorithm to consider. Being married lessens the costs on an individual to go out and vote while providing a separate incentive to do so due to an additional stake in the election (Teixeira 1987). While not directly mentioned, marital status makes intuitive sense from a mobilization perspective as well. A second person increases the likelihood of being embedded in social networks, increasing the likelihood of direct and indirect mobilization (Rosenstone and Hansen [1993] 2003). To capture marital status, its effect on networks, and where it put someone, an individual's marital history was considered. Although it was asked about in the ANES from the perspective of whether someone had been married, the variable was dichotomized to focus on whether or not someone was divorced, as a way of separating out those who had been married and separated from those who had not. With this operationalization of an individual's overall marital status, BeSiVa could determine its relationship to turnout due to the inclusion of a variable operationalizing it.

Due to their consideration as potential operationalizations of mobilization, a collection of variables was included to determine whether individuals voted due to political contact. Multiple variables to determine whether an individual had been contacted by a party or some other organization were included for the algorithm's consideration, as well as their employment status, a key predictor of being at the center of a network of potential mobilizing agents (Rosenstone and Hansen [1993] 2003). The ANES captured these responses through a series of dichotomous questions, asking whether an individual had been contacted by the Democratic or Republican parties, any party, or by another political organization. These variables were included for the algorithm's consideration as a means of capturing whether someone had been directly mobilized.

Direct mobilization is a key means of determining participation in the mobilization model of political participation, but it remains half of the main drivers of mobilization theory, and merited inclusion. This was done in the form of an individual's connection to networks through religious participation, asking whether they attended church or not. Church attendance was used to further capture the potential role of such a network, which was asked as a 5-item question based on how often someone went to church over the course of a year. This was recoded, however, into a dichotomous response to determine whether someone attended church or not. With this means of capturing whether an individual was connected to a greater network, a selection of variables most relevant to understanding whether an individual chose to vote or not could be considered by the algorithm.

**Methods and Results**

Once the variables were recoded into more theoretically appropriate forms, they were provided to the algorithm, which went to work determining which independent variables best predicted turnout behavior. Instead of running the algorithm once, however, the algorithm was run on the same data 100 times, randomly separating the data into different training and test sets each time. The training and test sets were varied by the use of different random seeds, guaranteeing a different division of data each time while also making it possible to recreate results as needed. The data, independent variables, and dependent variables were always the same, as were the arguments provided to the algorithm. These arguments specified five iterations, meaning that the maximum number of variables in any single model the algorithm created was 5. It also specified a threshold of 0.1%, meaning that any variable that could be added to the model would need to make the PCP improve by at least one tenth of one percent, or the algorithm would stop and return its results. The algorithm was run on the data, and the results of the PCP as well as the independent variables were saved, analyzed, and compared against theoretically specified models.

[Figure 1 about here]

From the 100 runs of the algorithm, two of the major outputs related to the final model that BeSiVa recommended, its PCP and the independent variables that it specified, were saved. The PCPs for each of the runs may be seen in figure 1, displayed in the histogram and kernel density plot. This shows the distribution of the PCPs, and notably the mean PCP. On average, the percent correctly predicted of these models is 70.5%, and the median is 70.1%, showing that the models created by BeSiVa are better than a random guess (which would lead to a PCP of 50%), and better than predicting the modal category of the dependent variable for all voters (which would have a PCP of 65.4%). The predictions are concentrated tightly around the mean, with a standard deviation of 2.8%, suggesting that the predictions the algorithm makes are consistent. In addition, a normal distribution created using the mean and standard deviation of the PCPs has been superimposed over the plot, displaying a marked similarity to the PCPs' kernel density[1](" \l "sdfootnote1sym). This plot provides a demonstration of the PCPs from the models that the algorithm generates, which may be compared to the predictions made by theoretically specified models.

[Figure 2 about here]

To compare BeSiVa's models to the theoretically specified models from the literature, it is necessary to compare predictions. To make this comparison, the models using variables specified by Campbell et al., Teixeira, and Rosenstone and Hansen were created with 20% of observations held out at random, placing them in a test set. Campbell et al.’s model only included party identification, while Rosenstone and Hansen included efficacy, education, income, age, and the network placement indicators, including church attendance and contact from parties and other organizations. Teixeira, meanwhile, included a vast number of variables. These variables included the demographic data seen in other models, links to the media, operationalized by the days someone read the paper, political efficacy, and the strength of partisan identification. Just like the BeSiVa algorithm, PCPs were generated on the held out data for the theoretical models. This process was repeated 100 times, and the PCPs for these models and one created from the four variables suggested by BeSiVa were saved and displayed in figure 2. Figure 2 shows a set of box plots which are compared to choosing the mode for all observations. As signified by the blue line, choosing the mode of the dependent variable for all observations would lead to a prediction that is 65.4% accurate, the proportion of people who said they voted. This shows that where accuracy is concerned, only BeSiVa's and Campbell et al.'s variable choices do better than picking the modal category for all observations. The black bars in the boxes represent the median, the central point of the data, allowing the theoretical models to be compared to one created by the algorithm, whose models had a median PCP of 70.1%. The box plots show that the median PCPs generated by the theoretically specified models are smaller than the median PCP of BeSiVa. This suggests that despite their theoretical specification, these models are not as capable of making predictions as one specified by the algorithm.

[Figure 3 about here]

In addition to the PCPs, the algorithm also provided a selection of independent variables in its final model, those that maximized the percent correctly predicted. These variables were saved and plotted in figure 3, with the x axis representing the number of times each variable was selected, and the y axis naming each independent variable selected by the algorithm. Unsurprisingly, the most commonly selected variable for predicting whether an individual turned out to vote is their level of education, chosen in approximately four fifths of the algorithm's selections of independent variables. Following closely is the strength of an individual's party identification, chosen over half of the time. These two variables are chosen often enough that based on the strength of the theories underlying their inclusion, and their performance in the algorithm, any model created from the results of BeSiVa ought to include education and party identification.

While education and party identification are the two most commonly included predictors in the algorithm's results, it is hardly surprising that these two well explored, theoretically validated variables were found to be useful in predicting vote choice. Surprisingly, however, the operationalization of mobility, the time spent in one's house, is the third most selected predictor, suggesting that residential mobility is important in determining an individual's likelihood to vote2[2](" \l "sdfootnote2sym). The time spent in one's house is highly favored by the algorithm as a predictor, selected third most often of all included independent variables. Perhaps this is due the entrenchment of individuals in a network, making them more susceptible to pressures to vote (Rosenstone and Hansen [1993] 2003), or due to more prosaic concerns. After all, moving requires that a person change their voter registration, a barrier to voting that may lead individuals to stay home on election day. Regardless of the theoretical underpinnings, BeSiVa selected the time an individual has lived in a house as the third most important predictor in vote choice, necessitating further investigation.

After time spent living in the same house, the rest of the independent variables selected by BeSiVa include few surprises, except perhaps for the fact that after party identification, a predictor favored by the psychological approach (Dalton and Wattenberg 1993), BeSiVa has a definite preference for the sociological approach. The age and age squared terms are selected, demonstrating the predictor's strength in predicting likelihood to vote. The algorithm prefers other demographic predictors, such as an individual's sex, status as a minority, and region, suggesting roles for these sociological predictors in determining vote choice. After the main demographic variables, two of the Mobilization theory's predictors are considered, but these variables were selected in less than 5 of the 100 times the algorithm was run, suggesting that their selection is unlikely to predict turnout in the majority of cases. This test demonstrated the relative quality of the variables selected, suggesting that among those tested, the psychological and sociological theories of vote choice were most relevant for predicting turnout.

**Model Validation through Bootstrapping**

Having obtained a sense of where theories stood for predicting whether someone chose to turn out or not, the question of how to determine the quality of the variables and models selected by BeSiVa, and compare them in a systemic way to the theoretically specified models became imperative. After all, BeSiVa had provided 100 separate recommendations for independent variables, each of which included a slightly different selection of variables provided to it. In an attempt to establish the quality of models according to the predictive criterion, a process similar to BeSiVa was attempted, comparing models with the variables it suggested against models containing theoretically relevant predictors, which can be seen in figures 4-5. The results suggested the remarkable power of education in predicting whether or not someone chose to vote, and the difficulty in determining the necessity for predictors beyond 4.

[Figures 4-5 about here]

To determine the suitability of the independent variables selected by the BeSiVa algorithm, the predictors were listed in the order of the number of times they were selected, and then added to models. The first model contained education, while the second contained education and party identification, following the pattern of variables seen in figure 3. Each of these models was subject to a cross-validation strategy, similar to the process of BeSiVa. A random subset of 20% of observations were kept from each model, and the model was reestimated without that data a specific number of times. Then, the held out data was predicted and compared against its measured value. This was done for models featuring each of the independent variables or variables selected by BeSiVa, and for three models based off of theory drawn from the mobilization, sociological, and psychological perspectives of turnout.

Once the models were tested and compared, their average PCP were stored, and plotted as the points in figures 4-53[3](" \l "sdfootnote3sym). There are a few things to note about these figures. First, there is the relative similarity of all of the models predicted by BeSiVa. Despite adding upwards of 13 variables, the models all have a relatively similar range of PCPs. The model with the largest PCP, signified by the X, appears to decrease from 7 to 4, as seen in the figures. While this might indicate that as the number of bootstrapped runs increases, the large confidence bands suggest that the maximum PCP is barely decreasing. Despite this, a confident statement may be made about the number of variables necessary to make a good prediction based on these confidence bands, due to their increase above the prediction that would be made if the mode were chosen for all observations.

If the mode were chosen for all observations, as displayed by the blue line, the percent of correctly predicted observations would not vary. Specifically, it would fall at 0.654, the proportion of individuals who said they voted in the ANES' 2000 Survey. By picking this value for all voters, however, someone attempting to make a prediction would outperform theoretically specified approaches, such as those suggested by Teixeira and Rosenstone and Hansen, as seen in the lines near the bottom of the chart. The addition of many variables, as Teixeira (1987) felt required to do, made for poor predictions, and the model suggested by Rosenstone and Hansen's mobilization approach made a similar prediction. The Michigan school's contribution, that the strength of party identification makes an individual more likely to turn out, nearly outperforms the initial model, but is quickly outperformed by BeSiVa. With the exception of the Michigan model, the highly theoretically specified models are incapable of making better predictions than choosing the modal category for all voters.

Although picking the modal category for all voters outperforms two of the theoretically specified approaches, it also may outperform some of the models created by BeSiVa. Given that in figures 4-5, the mode occasionally falls within the 95% confidence intervals, the models created with variables suggested by BeSiVa may not always outperform choosing the mode in all cases. Despite this possibility, however, some of the models do outperform choosing the mode. Figures 4-5 all show that after 4 variables are added, the confidence band falls above choosing the mode, which continues until ten to twelve variables are added to the model, depending on the number of times cross-validation was run. In the interest of parsimony, it may then be suggested that four variables, education, party identification, time spent in one's house, and age, theoretically specified and chosen by the algorithm, maximize prediction on the choice to vote using BeSiVa.

**A Comparison to Statistical Significance**

Having demonstrated the results suggested by the algorithm, it makes sense to compare them to the results generated by null hypothesis significance testing, examining the statistical significance of the variables and how the theoretically specified models differ from those the algorithm proposes. Instead of running bootstrapped regressions to compare predictions, each of the models suggested in the last step were run as separate logistic regressions, as were regressions featuring the variables recommended by Campbell et al. (1960), Teixeira (1987), and Rosenstone and Hansen (1993). All available data were used for these regressions, with no separation of data for testing purposes, and the models resulting from each of these regressions may be seen in tables 1-3. Comparing the models of these tables, it is clear that statistical significance, while capable of explaining variation in the dependent variable, makes it impossible to determine a variable’s ability to make predictions.

The models seen in tables 1-3 have many variables with statistical significance, but the variables that significance suggests are relevant frequently make the predictions worse. Although many of the models that seem most appropriate for predictive purposes have significant variables, not all of the variables recommended by the algorithm for making predictions are statistically significant. The first four most commonly selected variables all are statistically significant, suggesting an accord between the algorithm, bootstrapping, and significance testing. The remainder of the models, however, display scattered statistical significance, some with p-values well under the widely accepted threshold of 0.05, despite the fact that no model performs better than the one created by the first four recommended variables according to the 1000 run bootstrap or seven variables according to the 100 run bootstrap. In addition to the quality of fit statistics, Tables 1-3 feature the average PCPs from the 1000 run bootstrap, which are also included separately in table 4. Regardless of which model is capable of making the best predictions, the scattershot nature of statistical significance makes it difficult to determine which variables are truly useful in predicting turnout. With some variables, their inclusion appears to make sense from an explanatory perspective, but when their effect on the average PCP is considered, it is clear that these variables’ inclusion make a model’s ability to predict worse.

Depending on the number of bootstrapping runs to generate the confidence intervals on the PCPs, an argument may be made about whether the seventh or the fourth model is more capable of making predictions. Looking at statistical significance, however, is counterproductive if the goal is to make a prediction from the model. As an example, the role of the squared term in age is debatable; in the 100 run bootstrap, it appears to lead to a maximizing of prediction, but the small coefficient size and varying significance make it difficult to determine whether it should be included. An individual's minority status appears to affect turnout, and it is significant until party contact is added, but this suggests that contact supersedes an individual's racial identity in a model that predicts whether a person turns out. This is especially troubling, due to the fact that based on the bootstrapping tests and the mean PCPs generated from those tests, party contact's inclusion leads to empirically worse predictions. Similarly, gender is never found to be statistically significant, despite improving models' predictive power according to the 100 run bootstrap. It may be that the 100 run bootstrap is doing a poorer job of determining which model is most capable of predicting turnout, but statistical significance decisively misleads the researcher, overstating the importance of predictors like the days an individual read the paper, church attendance, and especially party contact. The use of statistical significance may inform the researcher if a non-zero relationship exists, but the idea that p-values inform the utility, substantive significance, or even relevance of a particular independent variable on predicting the dependent variable is hopefully abated.

**Quality of Fit Statistics**

Although significance demonstrably misleads if a prediction is to be made, this is not necessarily a problem with significance itself; just because a variable is considered significant doesn’t necessarily mean it is supposed to be predictive. This is contrary, however, to the way that statistical significance is frequently considered. Too often, NHST and quality of-fit statistics and predictive capability are conflated (Shmueli 2010), and the quality-of-fit statistics confuse the issue further. Looking at models' quality-of-fit statistics, it is clear that despite the algorithm's prescription for parsimonious models, the estimates of the models’ deviance and Akaike Information Criterion (AIC) do not change in a manner that reflect the predictive capability of each model. If used to estimate predictive capability, the quality-of-fit statistics suggest that the models are improving at making predictions, decreasing consistently (with a slight exception) as variables are added to the models. Given that these statistics are meant to suggest the difference between models’ performances, serving as measures of goodness of fit (Long 1997), this is especially problematic given that they are thought to signify a model’s ability to make good predictions (Shmueli 2010), which is explicitly stated for the Akaike Information Criterion (Forster and Sober 1994, Forster 2002). Given the utility of the models suggested by the algorithm, and the predictions made by the models in the bootstrapping, a model’s ability to predict and its ability to explain, as signified by the quality-of-fit statistics, must be explicitly differentiated.

As previously described, tables 1 and 2 feature a set of regressions based on variables suggested by the algorithm, and they include the quality of fit statistics for these models, as well as their mean PCP. Now, if the 100 run bootstrap is considered, then the model with seven variables is the first to outperform choosing the mode for all voters. If the 1000 iteration bootstrapping process is considered, then the four variable model features the only four variables necessary to create the best prediction of the choice to vote. This would be difficult to determine from looking at the quality-of-fit statistics, however. These estimates of model quality are supposed to measure how well the model explains the provided data, or in the case of the AIC, how well the model separates the predictive trend from the noise in the data (Forster and Sober 1994, Forster 2002), an explicit measure designed to predict new data comparable to the PCP (Clark 2004). Their performance, however, indicates the need for a prediction as a separate criterion based on how they improve, nearly consistently, given additional independent variables.

In considering the quality-of-fit statistics for each model, it would be impossible to determine that the best prediction, on average, is made by a model with four to seven independent variables when looking at the deviance. Despite the fourth model's status as the most predictive model in the 1000 iteration bootstrap, the deviance of each model continues to decrease even after variables are added that do not improve a model’s predictive capability. Deviance, a comparison between a perfect prediction of the results and the model created and displayed, is meant to compare between models, determining which ones are best at fitting the data (Long 1997). In this case, however, models with variables that lead to demonstrably poorer predictions still have lower deviance. Instead of one of the algorithm’s suggestions, relying the deviance would suggest that the most predictive model is either one proposed by Teixeira or Rosenstone and Hansen, if it is interpreted as a measure of prediction. The deviance almost continuously drops as variables are added, sometimes dramatically even if the prediction is worse, suggesting that deviance cannot be used to determine which model makes the best predictions. The deviance may be useful from an explanatory standpoint, but it misleads a researcher who seeks to use a model for making predictions.

In considering the models' summary statistics, deviance is not useful for determining whether a model’s predictions are better or worse than other models. The deviance is likely overfitting due to its purpose, which is to compare the relationship between the model at hand to a model that is capable of perfectly predicting the data. The Akaike Information Criterion (AIC), however, is meant to be a measure of model performance for predicting data that was not included (Forster and Sober 1994, Forster 2002), and its use as an information criterion has been suggested in contrast to using a test set for making predictions (Clark 2004). Despite this formulation, however, the AIC experiences a problem similar to the deviance, suggesting that it too is vulnerable to overfitting. In tables 1 and 2, the AIC continues to decrease, suggesting that the predictions that each model makes are improving, and the AIC decreases dramatically when contact by a party is added to the model. By comparison, the average PCP is direct a measure of the predictions that each model makes, and it suggests that once 4 variables are added, no improvement on prediction is possible. Just like the Deviance, the AIC suggests that there is a nearly consistent increase in the quality of predictions as variables are added, with only a slight hiccup when the variable accounting for the south is added. By comparison, the PCP suggests that the reverse is true; adding many of these variables leads to poorer predictions, suggesting that similar to the deviance, the AIC cannot determine a model's predictive capability despite its specification as being able to do so.

When comparing the models suggested by the algorithm, the AIC's value drops when almost every variable is added, despite the fact that adding more variables to models made their predictions worse, as demonstrated by the bootstrapping. The AIC continues to decrease in value among the models built from theoretically specified variables, with the exception of the Michigan school's recommendation to consider strength of party identification. Notably, the lowest value for the AIC among all of the explanatory statistics is held by the model with variables suggested by Teixeira, which had a lower predictive accuracy than any model suggested by the algorithm, and is comparable to a model containing the variables recommended by Rosenstone and Hansen. The PCPs, however, actually decreased, showing that the AIC was underperforming as a way of determining how well a model could make predictions. The AIC also suggested that of the three theoretically specified models, the best model for making predictions was Teixeira 1987, with Rosenstone and Hansen 1993 in second place, and Campbell et al. 1960’s predictions performing the worst. Empirically, the reverse is true, with the Michigan school doing a better job of making predictions than the other two models in table 3, as seen by comparing their average PCPs. When any measure, statistical significance or quality-of-fit, is compared to actual predictions, it becomes clear that no tool currently used by political science is capable of making accurate, relevant predictions. This calls for an addition to political science’s toolbox, tools that focus directly on making predictions. This chapter demonstrated two such tools, and how they can be used in a deductive mode of inference: BeSiVa and the bootstrapping approach. With these techniques, the question of where relationships exist may be considered, but also what is predictive, and therefore what is useful in determining the value of a given dependent variable.

**Conclusion:**

This chapter represents a demonstration of the utility of the predictive approach when employed with a deductive research design. This demonstration involved collecting a series of theoretically relevant variables and providing them to the BeSiVa algorithm. Once the variables were provided, the algorithm used its predictive criterion to determine which variables were necessary to create the most predictive model. From there, the variables it recommended were added to a model, one at a time, testing their utility separately using a bootstrapping approach. In addition to the algorithm's variables, models were created from three different theoretical perspectives of the choice to vote, and they were compared to the algorithm's recommended variables, via bootstrapping and the more conventional regression tables. These tables only served to demonstrate how a model may possess statistically significant variables, with a veritable constellation of stars suggesting significance, while making predictions that barely differ from flipping a coin. Based on the algorithm and the bootstrap, four variables appear to create the most predictive model: Education, the strength of party ID, how long someone lives in a single house, and age.

In addition to the comparison of different theories behind voter turnout, using the algorithm enables a consideration of several methodological critiques of the discipline. The idea that a well specified model contains only three independent variables (Achen 2002) may be accurate, but additional independent variables may add to the model's utility from a predictive standpoint. Whether such an addition is warranted requires testing the models using the algorithm and bootstrapping techniques, but through these techniques, it is possible to determine how preferable parsimony really is. If it is possible to include fewer variables while maximizing prediction accuracy, then parsimony is a useful goal. If the inclusion of a smaller number of variables fails to create a more predictive model, then parsimony is less desirable.

The BeSiVa algorithm and the bootstrapping used to consider the variables it selected systemically demonstrated that in the case of the choice to vote, a model with more than three independent variables may provide additional benefits from a predictive standpoint. But the bootstrapping, which demonstrated how BeSiVa's selections could be used to create predictive models, issues a potential response to a second charge Achen levies. Achen suggests that methods such as logistic regression are overused, and not the most appropriate way to model data. Through predictive criteria such as the PCP, the determination not only of the most appropriate variables, but also the method's appropriateness used may be considered from a more objective standpoint.

While it may be beneficial to consider the use of alternatives to logistic regression, prediction lets researchers determine the necessity of finding an alternative. If logistic regression leads to good predictions on data that has been kept away from the regression for testing purposes, then a hunt for a better estimator than the logistic estimator is not essential. If logistic regression does not make good predictions, however, then the methods used to create the model must be reconsidered. Through the BeSiVa algorithm and bootstrapping, however, such a necessity may be determined. This can be done by comparing the predictions against 1/C, where C is the number of categories that can be predicted, as described by Kuhn and Johnson (2013), or by comparing the predictions against the prediction from choosing the mode for all observations. The fact that by a predictive criterion, such a comparison is possible with logistic regression shows that the estimator has utility beyond what its critics suggest.

The predictive criterion may be used to make the most capable model, using a prespecified technique and variables that are chosen for their theoretical capability. Using such a criterion may even concur with ironclad theoretical findings. But the power of an approach that explicitly considers the ability of a variable, model, or theory to make predictions arises from the possibility that a well specified theoretical model, or a variable whose relevance seemed unimpeachable may not be capable of aiding in actual prediction. This was demonstrated by comparing theoretically specified models and their ability to predict whether an individual turned out to vote, to the models created by the BeSiVa algorithm, and finding the theoretical models wanting. This criterion allows for the consideration of how well models make predictions, making it possible to determine whether theories, either through groupings of variables or the overall selection of what to give to the algorithm, leads to a good prediction. It also demonstrates that statistical significance is incapable of determining what makes a good prediction. Despite the apparent relevance of a given variable according to statistical significance, only an actual predictive test may determine a variable’s utility and justify its inclusion from a substantive perspective.

[1](" \l "sdfootnote1anc). While the histogram bore marked similarities to a normal distribution, the bounded nature of the data makes the conventional means of plotting confidence intervals problematic, due to the central limit theorem's lack of applicability to bounded distributions.

[2](" \l "sdfootnote2anc). While such a conclusion validates a hypothesis mentioned by Teixeira (1987), the variable meant to operationalize mobility in this case may also captures more about the individual than anticipated. Given the lack of class-based variables in the algorithm's selections, despite the inclusion of income in the list of potential independent variables, the length of time a person spends in a single house may be capturing elements of economic class and age. Given its number of selections, time spent in the same house may also be capturing elements of these two independent variables, among others, better than the variables meant to operationalize themit may be that hypotheses about time spent in a location are underrated, it may also be that this variable is highly correlated with other theoretically relevant predictors. The correlations between the amount of time living in one's house and age is slight, it is significant, as signified by their 0.18 Pearson's R, which was also the case between time spent in a house and income for a value of 0.14. The correlations between these predictors make sense; a stable place to live, while not necessary, assists with the stability needed to attain a substantial income, and time spent in house necessarily correlates with age. The older an individual is, the longer they can spend living in one place, making age necessary, if not sufficient, to stay in one place for some time. The correlation of these two predictors with the amount of time spent in one's house suggests that it is picking up more information than the question's formulation might expect.

[3](" \l "sdfootnote3anc). In addition to determining the average, however, the confidence bands were desired, despite the difficulty of getting a confidence band for a bounded value like the PCP, which ranges between zero and one. For this reason, the confidence bands were determined through bootstrapping; the 2.5th and 97.5th percentiles of the PCPs captured by the cross-validation were measured, creating a bootstrapped 95% confidence band akin to those created for continuous variables

Figure 1: The algorithm's percent correctly predicted. The final models for BeSiVa had a mean of slightly below .7, and bear superficial similarities to a normal distribution.

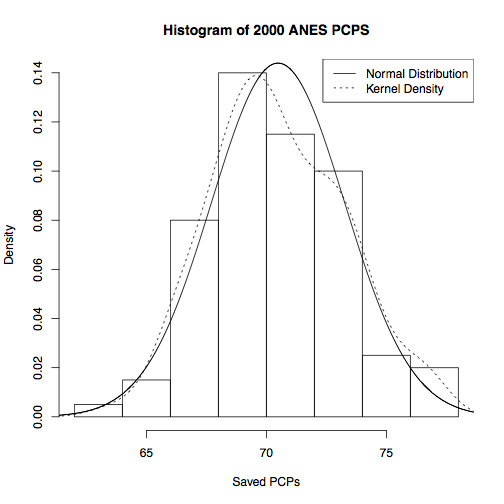


Figure 2: A comparison of the PCPs between theoretically specified models. By comparison to the PCP in the histogram, the medians here fall well below 70%, suggesting that these models are less capable of predicting whether a person votes, Even Campbell et al. 1960, whose model is still well below the mean percentage of the percent correctly predicted.

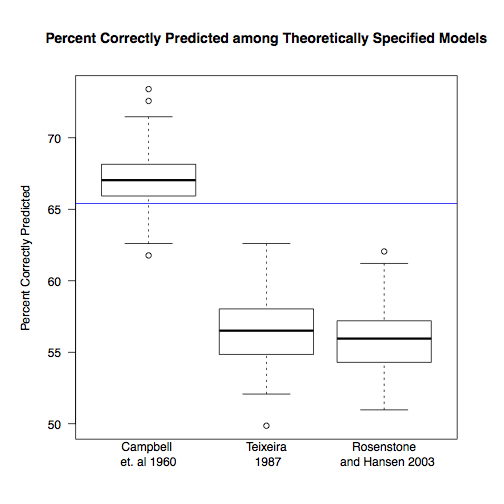


Figure 3: BeSiVa's selections. Out of 100 runs, BeSiVa selected education most often, followed by party ID and time spent in someone's house.

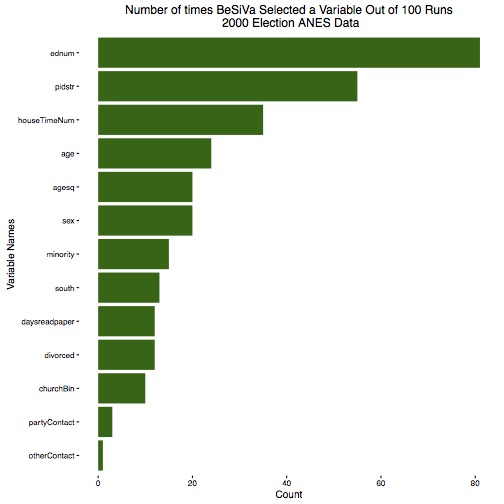


Figure 4: A test of the independent variables. The independent variables selected by BeSiVa in figure 3 were added to models one at a time, and cross-validated 100 times. In this case, the best model is the one that has 7 variables included, but the variation is large.

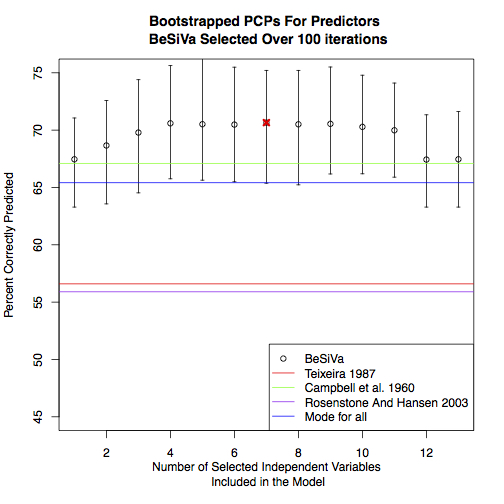


Figure 5: the same test, repeated 1000 times. Note that in this case, the largest PCP is also the first one whose 95% confidence band fall outside the choosing the mode for category, with the first 4 independent variables.

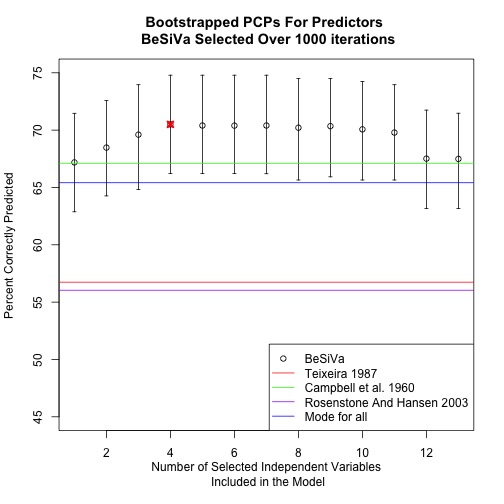


Table 1: The models with variables suggested by the algorithm and used in the first six iterations of the bootstrapping process. The bootstrapping tests suggest that the fourth or seventh iterations are most useful for making predictions. In this case, no data are held out, to show the coefficients and significance that would arise from an explanatory standpoint.



Table 2: The seventh through thirteenth iterations of the bootstrap. Similar to table 1, no data were held out, and despite the limited predictive capability of any model after 7, and especially 12 and 13, significance continues to arise for some of the variables.



Table 3: Theoretically specified models. The summary statistics on both Teixeira, and Rosenstone and Hansen are both expansive, suggesting their utility from the explanatory standpoint.



Table 4: The average percentages correctly predicted. Variables listed refer to a model with that variable, and all previous variables included as an independent variable. The models at the bottom are theoretically specified.

