The Divided (But Not More Predictable) Electorate: A Machine Learning Analysis of Voting in American Presidential Elections

Seo-young Silvia Kim*[†] and Jan Zilinsky[‡]

[†]American University
[‡]New York University

April 1, 2021

Abstract

Political parties are increasingly homogeneous both ideologically and demographically. With increased party-line voting, a natural corollary of sorting is that membership in demographic groups should be increasingly prognostic of vote choice. We argue that predictability of voting decisions is a useful quantity of interest for testing hypotheses from the literature on partisan and demographic sorting. Contrary to expectations, we find that demographic sorting has not produced a very predictable electorate. Tree-based machine learning models, trained on demographic labels from public opinion surveys between 1952 and 2020, predict only 63.5% of out-of-sample vote choices correctly on average. Moreover, demographics have not grown more predictive over time, while partisanship has. Partisanship's diagnosticity has risen in absolute terms, and its relative dominance over ideology has been stable for the last seven decades. Additional data about voters can still yield superior predictions, but its added value decreases over time as partisanship's predictive power grows.

Keywords: Vote Choice, Elections, Polarization, Sorting, Ideology, Demographics, Machine Learning, Tree-based Models, Random Forests

Word Count: 9,937

*Kim is an Assistant Professor at American University, and is the corresponding author (email: sskim@american.edu). Zilinsky is a Ph.D. candidate at New York University. We thank Jeff Gill, Shiro Kuriwaki, R. Michael Alvarez, Patrick Egan, and the participants at PolMeth Europe 2021, NYU American Politics workshop, and American University works-in-progress workshop for comments.

1 Introduction

Political campaigns segment the electorate into categories based on how voters' observable characteristics are likely to correlate with voting behavior (Fenno, 1978; Hersh, 2015). Scholars also pay close attention to group behavior to explain social and political trends, and there is a deep interest in quantifying and explaining cleavages between groups (Axelrod, 1972). This interest is understandable: whether a readily-perceivable group, typically in terms of demographics, is a reliable base or a swing voting bloc for a particular political party has substantial implications for representation. But how reliable are demographic labels in predicting a presidential vote choice? Moreover, in light of growing polarization, do demographic groups increasingly vote in predictable ways?

Instead of calculating each demographic group's marginal impact, we consider the joint predictive power of observable demographics for inferring voting decisions. If demographic groups have sorted into political camps, their behavior should be increasingly predictable. Using public opinion surveys from 1952 to 2020, we find predictions based on five commonly used demographic variables—age, gender, race, education, and income—have low accuracy and are stable over time. Only 63.5% of vote choices are correctly predicted in hold-out sets of the respondents whose data is deliberately unused in the model-training stage.

Not only is the informativeness of demographic labels surprisingly muted, but their predictive power is virtually flat since the 1970s. On the other hand, when partisan labels are incorporated into prediction algorithms, voting behavior is increasingly predictable over time, which is consistent with the literature on partisan sorting (Levendusky, 2009).

Two main ideas have been referred to as 'sorting.' *Ideological sorting*, which is the more common definition, is that both symbolic ideology (i.e., self-identified liberal vs. conservative placement) and operational ideology (i.e., concrete issue positions) are increasingly correlated with partisanship (Levendusky, 2009; Fiorina et al., 2011; Hetherington, 2009; Weber and Klar, 2019). *Social sorting* is defined as a convergence of social identities and partisan identities, such as race, sexual orientation, religion, occupations, social movements, and so on (Mason, 2016; Mason and Wronski, 2018; Mason, 2018a). Both phenomena are believed to entrench partisanship, decreasing the scope for persuasion.

There is little doubt that the ideological distance between Democrats and Republicans has grown (Webster and Abramowitz, 2017). Accordingly, intra-party heterogeneity has decreased; for example, Norris and Inglehart (2019) observe that "the two major parties gradually shifted to become more homogeneous internally in their cultural positions and more polarized between parties." Consequences include greater partisan animosity (Iyengar and Westwood, 2015; Abramowitz and Webster, 2016; Bougher, 2017; Christenson and Weisberg, 2019; Iyengar et al., 2019), an increase in straight-ticket voting in recent decades (Jacobson, 2017; Burden and Kimball, 2009),[1] and an unprecedented partisan gap in presidential approval rates (Jacobson, 2019).

But what are the consequences of social sorting for voting behavior? A natural corollary of social sorting together with increased party-line voting (Jacobson, 2013) could be that group membership is increasingly prognostic of vote choice.[2] We consider a narrow view of social sorting—namely that *demographic group labels* such as age and race are linked with political behavior—and test whether demographic sorting in vote choice is growing. Specifically, are demographic characteristics and political opinions/partisanship tied in a way that translates into voting behavior, so that belonging to a particular group predicts voting decisions with high accuracy?

Although hinting at it (see a partial list of relevant papers in Section 5), few studies have shown whether the demographic labels' joint predictive power over vote choice has consistently increased over time. In this paper, we first test whether demographic markers

[1] This is despite the gradual rollback of straight-ticket voting options in election administration (NCSL, 2020).

[2] Again, note that when the literature discusses sorting, it usually refers to sorting into "correct parties." See for example Figure 4.1 in Fiorina et al. (2011).

are increasingly more informative of vote choice, based on whether voting decisions can be accurately inferred. For this, we use random forests on data from ANES (1952–2016), CCES (2008–2018), and Nationscape surveys (July 2019–June 2020). The classification accuracy under the training/testing paradigm, we argue, is a quantity of interest that can help validate whether sorting occurs. Contrary to the prevailing narrative, we show that demographics have not become more prognostic of vote choice over time.

Second, we ask whether other data about voters increase the predictive power of voting decisions, as implied by the findings on swing voters' rarity (Panagopoulos, 2020). With the same data and methods, we systematically investigate how well vote choice can be inferred from various combinations of voters' characteristics. We confirm that the (1) predictive power of partisanship is increasing, and (2) while other variables additionally contribute to improving the prediction accuracy, the added value when partisanship is already accounted for is decreasing. These results, in line with the numerous results showing increasing partisan polarization, lend credibility to our initial results.

Finally, we ask whether demographics remain informative predictors when other variables are accounted for, and if so, which ones. We use permutation-based variable importance calculations to determine whether randomly changing the values of a variable of interest significantly reduces the predictive accuracy, thus determining how 'important' a variable is. We find that being a Black voter is a top 10 variable even when party ID and issue positions are accounted for. Other variables such as age and education appear less consistently among the top predictors. However, when all variables are used for prediction, demographics completely disappear from the top 10 variables, giving way to other variables such as beliefs, perceptions, and issue positions.

2 Literature and Hypotheses

Demographic attributes can naturally translate into social identities and, according to the Columbia School of political behavior, they are reinforced by social networks (Lazarsfeld et al., 1944). Membership in social groups is then said to explain politically relevant beliefs and behaviors, and voting can be viewed as akin to cultural activities which "have their origin in ethnic, sectional, class, and family traditions" (Berelson et al., 1954). In this framework, conformity is the norm.

Moreover, political elites have paved the way for a stronger group-party alignment in recent decades, making party positions clearer and more distinguishable to voters (Levendusky, 2009). In response, cross-cutting ties between groups have been decreasing (Mason, 2018b), increasing social polarization. Voters are thus expected to vote in line with their perceived group interests and against members of a disliked out-group.

It then does seem natural to state that group membership should be increasingly informative of vote choice. Essentially, this is the wide-spread assumption that "the link between demographic traits and political orientation is so strong that increases in the share of voters from demographic groups associated with support for the Democrats produce proportionate increases in Democratic support," (a view summarized but not endorsed by Shaw and Petrocik, 2020). Is this assumption true? Drawing on the existing literature, we will derive three hypotheses for the sorting thesis' testable implications.

2.1 Group-based Voting

Although social identities that align with political ideology encompass many categories such as religion and interest groups (Abramowitz and Saunders, 2008; Levendusky, 2009), our focus is on the following five demographic markers: race, education, income, age, and gender. Individuals from opposing parties now differ more on average in political opinions and their observable demographic characteristics. Note that while the selected variables are those considered to be more "objective" labels, how strongly individuals identify with their group may vary.

Race. Although the Democratic Party has had a stable advantage among Black voters for several decades now, there are reasons to expect that the signal from a voter's race is larger than in the past. First, Trump's victory in 2016 was a continuation of "the decades-long expansion of Republican support among white working-class Americans" (Carnes and Lupu, 2020). In addition, the GOP is believed to have activated white identity (Tesler, 2016; Sides et al., 2017, 2019) in response to Barack Obama's electoral wins. In the case of the Latino voters, the Democratic support varies by ethnicity and generation (Abrajano and Alvarez, 2012), while the Asian Americans are, overall, not well courted by either party (Wong et al., 2011). But racial minorities ,on average, may have a weaker incentive to vote Republican. Note also that when the Tea Party emerged in 2010, 80–90 percent of its supporters were white (Williamson et al., 2011), strengthening the hypothesis that prediction accuracy based on race could be increasing.

Education. Three decades' worth of public opinion demonstrates that those with higher educational attainment increasingly associate with the Democratic party (Pew Research Center, 2018). Conversely, Republicans have been gaining support among those citizens who do not have a college degree in the last decade.[3] The partisan education gap reached its peak in 2016, but note that the relationship between education and voting is sensitive to the inclusion of other variables in a model (Schaffner et al., 2018).

Income. Income at the individual level predicts vote choice, but there is some disagreement whether class-based voting has been stable (Gelman et al., 2010) or increasing over time (Stonecash, 2000; Bartels, 2006; McCarty et al., 2008).[4] Preferences and voting are

[3]Throughout the 1990s, the Republican Party did not yet have a lead among white registered voters who were high school graduates. This group of voters was still evenly split between the two major parties.

[4]The best-known argument comes from McCarty et al. (2008, p. 75) who argue that there has been growing "stratification of partisanship by income," with high-income voters increasingly voting Republi-

typically aligned with people's economic self-interest—for example, Ansolabehere et al. (2006) document that "the difference in the rate of Republican voting between an economic Conservative and an economic Liberal is 31 percentage points." However, the extent of the importance of economic issues for voting continues to be debated.[5] In the 2016 presidential election, the income effect is believed to have interacted with education. Carnes and Lupu (2020), for example, show that the diploma divide in 2016 "was driven largely by more affluent Americans."[6]

Age. Young people tend to lean liberal and support Democratic or progressive candidates. In the 2016 presidential popular vote, the vote margin of Democratic minus Republican votes was 24 for Millennials and 28 for Generation Z (Griffin et al., 2020). Higher age, conversely, is correlated with conservatism and voting Republican.[7] Consider also that when respondents were allowed to select up to two groups with which they have most common interests and concerns in a November 2020 YouGov poll, the most frequently mentioned category was "people in the same age group as you," followed by "people in the same political party."[8]

Gender. Voting patterns in exit polls suggest that men are more likely to vote Republican, but in models that control for sexist attitudes, gender does not appear to predict vote choice (Bracic et al., 2019). At the same time, gender interacts with race. For example, Junn (2017) reports that white women voted Republican in 2016 in line with their

can.

[5]The relationship between income and Republican partisanship at the individual level, while robust nationally, is moderated by local context (especially ethnic composition) according to detailed analyses of voter files (Hersh and Nall, 2016).

[6]Tree-based methods are ideally suited for identifying interactive relationships between variables and exploiting them to produce accurate predictions.

[7]Williamson et al. (2011) found that at least 75% of Tea Party supporters were over 45 years old.

[8]That is, rather than inferring the importance of group memberships, respondents were asked directly: "Would you say that you share a lot of common interests and concerns with other people of people who are [SAME GROUP], or would you say that age is not really relevant?". In this context, respondents suggested that class ("people who have about the same amount of money as you"), ethnicity, and geographic proximity were less indicative of common interests than age (YouGov, 2020).

behavior in prior elections.

2.2 Hypotheses

Our objective is to test whether demographic sorting is taking place. We argue that the testable implication of demographic sorting is increasing predictive power of demographics on presidential vote choice. Based on the summarized relationships between demographics and political behavior, we derive the following hypothesis:

> *Hypothesis 1 (Increasing Demographic Sorting): Vote choice will become increasingly predictable based on voters' demographic features alone (with other information about voters withheld).*

Beyond demographics, a long scholarly tradition leads us to expect that an explicit self-reported party label should be a powerful signal of vote choice. Campbell et al. (1960) described the psychological, unthinking allegiance to parties with the following consumer analogy: "[l]ike the automobile buyer who knows nothing of cars except that he prefers a given make, the voter who knows simply that he is a Republican or Democrat responds directly to his stable allegiance" (p. 136). On the basis of this argument and the evidence that voter loyalty is increasing Jacobson (2017); Burden and Kimball (2009), we propose the next hypothesis:

> *Hypothesis 2 (Increasing Party ID Sorting): Including explicit party ID will dramatically raise accuracy relative to sparser models using only demographics, and predicting voting decisions with party ID will grow increasingly easy over time.*

Finally, given that parties are seen as ideological brands (Woon and Pope, 2008; Egan, 2013)and voters now generally belong to the "correct" party (Levendusky, 2009) we propose the final hypothesis:

> *Hypothesis 3 (Sufficiency of Party ID): Beyond the initial sets of features (party ID and*

demographics), other voter characteristics, such as issue positions, will contain minimal diagnostic information about vote choice.

2.3 Machine Learning and Political Behavior

There are several reasons for using supervised machine learning methods to test the above hypotheses, especially the tree-based methods that we choose. First of all, the metrics we use to evaluate the results are performance-based on correct out-of-sample predictions. Second, random forests allow flexible interaction structures between variables, uncovering hidden relationships in large datasets (Montgomery and Olivella, 2018). Third, when the set of potential predictors is large, researchers can prune the set of covariates or identify the most important predictors in distinguishing the outcome variable, instead of arbitrarily restricting the set of allowable model specifications (Kim et al., 2020).

We emphasize the first upside of regression trees—and machine learning in general—relative to the family of parametric models usually employed in social science, typically under the maximum likelihood umbrella. A common approach in the existing literature is to estimate a set of logistic regressions and evaluate their performance based on the percent of correctly classified observations (or McKelvey-Zavoina's pseudo R^2) *in-sample*. However, when the out-of-sample fit is not reported, readers cannot evaluate whether the reported models over-fit to the given sample.

Several recent papers, recognizing these advantages, have used these flexible non-parametric methods to explore complex structures in political behavior and perform direct prediction. For example, Bonica (2018) use them to predict legislators' behaviors and issue positions based on campaign contribution records, while Kim et al. (2020) identifies the best predictors of turnout. As the aforementioned authors, we rely on tree-based methods to achieve the best possible prediction given the covariates.

Peterson and Spirling (2018) in particular uses the classification accuracy itself as a substantive quantity of interest: if a trained algorithm easily distinguishes the party label of a political speech, the result is interpreted as a higher degree of polarization.[9] Our logic is quite similar: a high degree of demographic sorting should go hand in hand with a high classification accuracy of the two-party vote choice.

Finally, we wish to derive variable importance measures by permutation (Breiman, 2001) to identify which variables contribute the most to increasing accuracy, instead of relying on statistical significance. The variable importance measures how much the prediction accuracy decreases when a given variable is either removed or reshuffled. This metric is not available under the traditional regression framework where the judgment about variable importance generally involves comparing coefficient magnitudes or significance-based criteria.[10]

Note that our paper is attempting to harness the power of nonparametric modeling of pure prediction algorithms for best classification under the training/testing paradigm. The intent is to assess the accuracy of predictions in a given election cycle and to see if trends are consistent over time. We apply random forests across all data and across model specifications to take advantage of these characteristics consistently, insofar as the number of covariates p in the richest models approaches the number of observations n in one of our datasets.[11] For a comprehensive 'checklist' of differences between regression models and prediction algorithms, see Table 5 in Efron (2020).

[9] In economics, Bertrand and Kamenica (2018) infer income and other consumer traits on the basis of their media consumption, shopping behavior, and other characteristics.

[10] It is important to note that statistical significance does imply that a variable is diagnostic or predictive (Lo et al., 2016).

[11] We compare the performance of these models to logistic regression and classification trees in the Appendix.

3 Data and Methodology

We use three sets of public opinion surveys: American National Election Studies (1952–2016, every four years), Cooperative Congressional Election Study (2008–2018, every two years), and UCLA Nationscape surveys (50 weekly waves in 2019 and 2020). The target variable for prediction is presidential vote choice, which is self-reported,[12] subsetted to respondents who voted for either a Democratic or a Republican candidate.[13] Using all these major surveys in public opinion will help us validate the results and check that our results are not the product of a single survey.

To predict voting decisions, we use random forests, a method for aggregating predictions from regression and classification trees. An individual tree is estimated by sequentially splitting the data on the basis of an optimally chosen cut-off point of the most informative variable.[14] To remove excessive dependence of tree structures on the algorithmic decisions early in the splitting process, a subset of observations and predictors is drawn each time a new classification tree is estimated. An aggregation of tress corrected for inter-tree correlation is the random forests (RF). Random forests put to use for each year/survey separately.

To investigate the extent to which voting behavior is inferrable based on voters' observable characteristics, we use four nested variable specifications for the analysis. Figure 1 shows the labels of and the variables included in each specification. Naturally, the third and the fourth specifications will usually consist of imperfectly overlapping sets of variables for each survey/wave. We include these specifications for benchmark purposes, given that the survey questionnaires reflect the issue cleavages of the day, such as the Iraq war or the Affordable Care Act.

All categorical variables are converted into dummy variables, including a variable to

[12] If there is a post-wave and a pre-wave, we use the post-wave variable. For the CCES mid-term election waves, we use the previous election cycle's presidential vote choice. For Nationscape, we use the

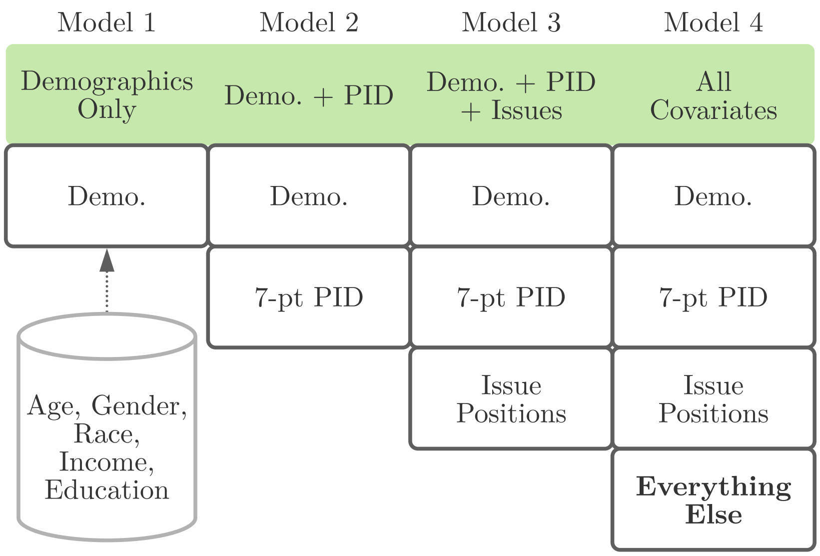


Figure 1: Four Nested Specifications and Corresponding Variables, Visualized

represent nonresponse missing values. Variables with near-zero variance at the 1% level or variables with more than twenty different responses, such as ZIP codes, are dropped to guard against too much sparsity. Only clearly continuous variables—such as age, number of children, or amount donated to political campaigns—are kept as continuous.[15] Nonresponses are treated as a separate category instead of listwise deletion if the variable is categorical.

After cleaning, the data is split into training and testing datasets with an 80:20 ratio. Using the `caret` and the `ranger` package in `R`, we run a class prediction via random forests (Breiman, 2001; Kuhn, 2008; Wright and Ziegler, 2017). All code is publicly available at

respondents' vote intention for the 2020 presidential election.

[13]For ANES, the cumulative dataset was used. For the CCES dataset, seven waves from 2008 and 2018 were used, after extensive wrangling and coding of equivalent variables.

[14]A variable is chosen in a given step if using that variable minimizes deviance.

[15]We chose to encode ordered categorical variables (such as education and party ID) as sets of binary features—so that, for example, a seven-point party ID becomes six binary variables—for several reasons. First, this allows us to be consistent over various surveys and years. Second, we avoid assuming that items measured on an ordered scale have linear or additive effects. Finally, if ordered categorical variables are treated as continuous, item nonresponses would be dropped, but one-hot encoding allows us to keep all survey respondents even when their responses contain missing values, because a separate binary variable is created (e.g., for respondents with a missing income category).

a GitHub repository: https://github.com/sysilviakim/surveyML. For comparison purposes, we also run a logistic regression and a CART model, the results of which are available in the Appendix.

Each model is tuned before a final (best-fitting) model is chosen out of a set of candidate models. The best model is selected on the basis of its performance in the training set while two hyperparameters that may impact model quality are tuned via 10-fold cross-validation.[16]

Note that while we have placed results from different surveys side by side for comparison (24 survey/years $\times$ 4 models, total 96 random forests outputs), we do not claim that they are by default fully comparable. These surveys were designed each for their respective purposes, and the number of respondents and survey modes differ. Even within the ANES survey, the survey modes have undergone some changes (e.g., the addition of the web mode). In particular, questions on cleavage issues are usually different. However, the demographic variables and partisanship variables are relatively similar.

Unlike ANES or CCES, Nationscape is a high frequency online poll where the number of respondents in a typical week is 6,250. We use the data collected prior to the onset of the COVID pandemic because of concerns about changes to the sample composition during an economic crisis. We randomly draw 20% of the pooled dataset, yielding a sample of 25,937 for the training set. We thus deliberately maintain a sample size between that of ANES and the CCES. We then draw 5,187 respondents from the hold-out set and evaluate the models' performance on this set of respondents.

As we have stated in Section 2, we argue our approach has several compelling aspects: better performance, a rich set of interactions between variables that can be flexibly explored, derivation of variable importance measures, and guarding against over-fitting

[16]For a lucid explanation of 'cross-validation,' see Neunhoeffer and Sternberg (2019). In our context, it is a procedure for model tuning that uses 90% of the training data. Each time the model is estimated on a subset of the data, we vary the number of variables available for splitting at each node and the selection of the splitting rule for classification (extremely randomized trees vs. the Gini index).

by evaluating the performance of models out-of-sample. Again, the last aspect yields honest estimates of model performance.[17] Second, regardless of model complexity, by focusing on solving a prediction problem, we can transparently summarize each model with metrics such as out-of-sample accuracy and the AUC.

There are also concerns and criticisms associated with borrowing methods from computer science, including low interpretability and higher computation time and costs. The former concern can be partially mitigated by storing intermediate outputs or visualizing how much predictions can be attributed to individual model inputs. The latter issue is becoming less severe as the availability of high-performance computing (HPC) improves, but we recognize that a lack of access to HPC can be a barrier introducing resource-based inequities across researchers.

4 Results

4.1 How Much Can Demographics Alone Predict Vote Choice?

Before we investigate the evidence for our hypotheses, we first quantify how much, on average, demographics can predict presidential vote choice. Figure 2 shows the time-series plots of out-of-sample accuracy values over time for all three surveys. The top panel shows accuracy rates over time for models estimated using only information on respondents' gender, race, education, income, and age.

We find that when using just demographics, the accuracy for vote choice predictions is generally low, typically less than 65%. More specifically, the average across all surveys and waves is 63.5%.[18] When determined for individual surveys, it is 63.1% for ANES, 64.7% for CCES, and 63.4% for the Nationscape's 2019–20 waves. It is striking to see

[17] Examples of earlier work employing similar methods in political science include Samii et al. (2016), Kim et al. (2020), and Demir et al. (2021).

[18] When the algorithm is changed to logit, it is 64.3%, and when CART, 62.8%.

that these numbers are remarkably similar even across different survey administrations, varying sampling methods, and in different time periods.

Is 63.5% a high number? It is certainly better than random guesses of the two-party vote choice, which on average would be 50%.[19] But is it high enough, considering the emphasis that the literature places on demographic variables? Compared to the spotlight they receive, altogether the accuracy does not seem extraordinarily high. The next question is then whether, per the first hypothesis, the predictive power of demographics has increased over time.

4.2 No Evidence of Demographic Sorting Over Time

Hypothesis 1 (Increasing Demographic Sorting). Can demographic labels alone predict presidential vote choice better in the polarized era compared to the past? The time trends shown in the first panel of Figure 2 do not suggest that this is the case. Test-set accuracy from 1952 to 2020 is remarkably stable.[20]

Two notable exceptions are the elections of 1972 and 2008. For both the Nixon vs. McGovern case and the Obama vs. McCain contest, the ANES-based accuracy is slightly above 70% (71.9% in 1972 and 71.0% in 2008). However, CCES-based accuracy in 2008[21] was only 63.6% and, crucially, not significantly different than the accuracy rates we observe in 2016 (64.6%) or 2020 (63.4% in the Nationscape data). Again, overall, given the area covered by the 95 percent confidence intervals of accuracy, we see that predictions

[19] One may argue that in particular elections when the results were not close, 50% is not a fair comparison. We provide the baseline two-party vote share for each presidential election in Appendix E, which further decreases the value that demographics bring to the prediction (compared to a prediction rule where for each respondent we would have naively predicted a vote for the winning candidate).

[20] In the Appendix, we also provide equivalent plots when (1) geographical information of south vs. non-south states is included, and (2) religious affiliation of whether the respondent is a Protestant, Catholic, Jewish, or none of the above is included. Unfortunately, whether the respondent is specifically an evangelical Christian could not be included for consistency, since the variable is not included in the early years of ANES. The results are similarly stable over time.

[21] Note that the large confidence interval of ANES survey datasets are due to their relatively small size, compared to CCES or Nationscape.

are typically only 10 to 15 percentage points better than random guesses. In fact, in 1960 and 2000, predictions were only marginally better than predictions obtained by chance (respectively 57% and 54% accuracy).

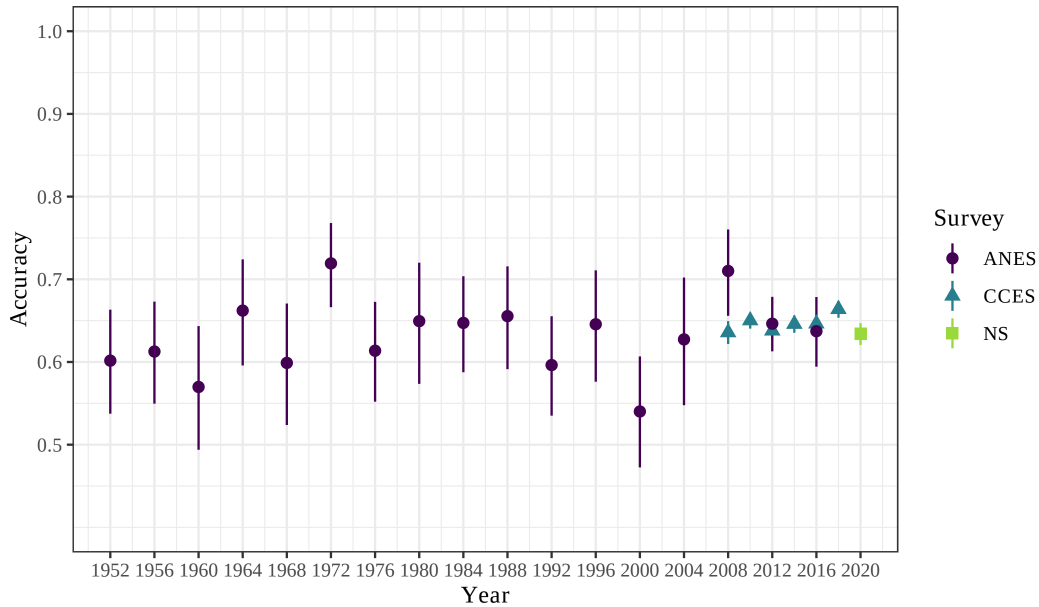
A simple regression slope of accuracy on years with just ANES data is 0.0004 with a standard error of 0.0006 (p-value of 0.46). The slope, while positive, is not statistically significant, thereby providing no evidence to reject the null.[22] If all surveys are pooled—again, with the caveat that accuracy between different surveys may not be directly comparable—the slope is 0.0004 with a standard error of 0.0004 (p-value of 0.24).[23] These null results are robust to the choice of algorithms.

Thus, there is no evidence to suggest that there is demographic sorting—voters' demographic characteristics do not provide more informative signals for vote choice over time. The results suggest that the electorate has not become more polarized along demographic lines in a way that is *informative about voting behavior*.[24] Note that this is not about a specific demographic group (e.g., marginal contribution of white working-class men in their 30s to a party voting bloc) but the overall, joint prediction ability of combinations of demographics, which may explain the disparity between the results of this paper and the literature's recent focus.

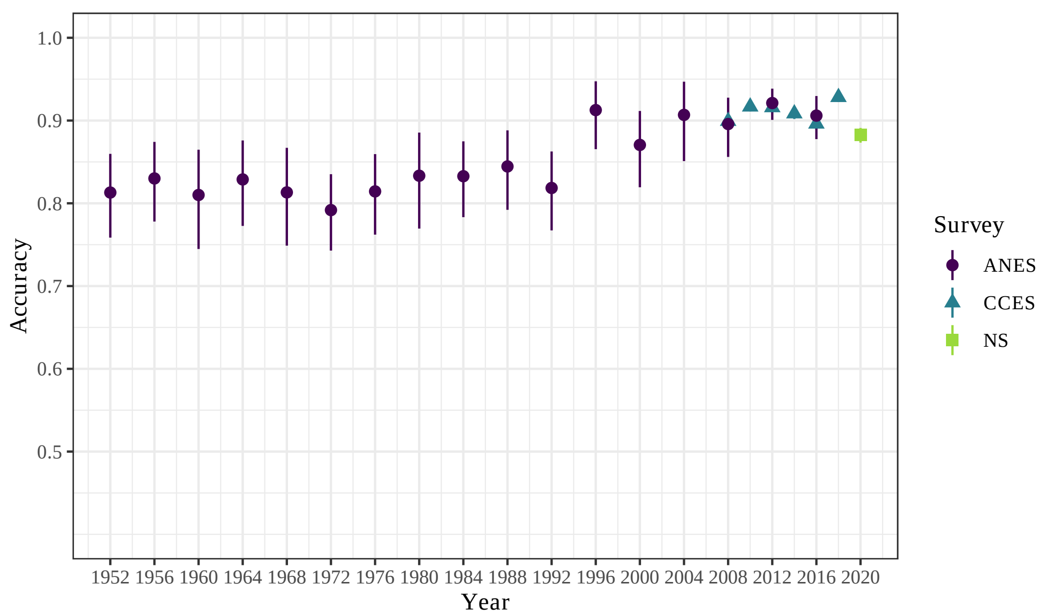
[22] When the algorithm is changed to logit, the corresponding numbers are slope of -0.0012 with a standard error of 0.0006 (p-value of 0.08). When the algorithm is changed to CART, the corresponding numbers are slope of -0.0008875 with a standard error of 0.0007 (p-value of 0.21).

[23] When the algorithm is changed to logit, the corresponding numbers are slope of -0.000683 with a standard error of 0.0004 (p-value of 0.13). When the algorithm is changed to CART, the corresponding numbers are slope of -0.00057 with a standard error of 0.0004 (p-value of 0.22).

[24] Note that AUC (in Appendix), over time, seems to be increasing unlike accuracy, with ANES-only slope of 0.0022 (standard error 0.0005, p-value of < 0.001). The pooled slope of 0.002, standard error 0.0004, p-value of < 0.001, hinting that over all possible threshold values, the ability to separate the Democratic vote from the GOP vote has increased over time (as long we treat the 1950s as a starting point, rather than the 1970s). But when classifying vote choices, researchers rarely use a threshold other than 50%.



(a) Demographics Only



(b) Demo. + PID

Figure 2: Out-of-sample Prediction Accuracy, 95% Confidence Interval, Presidential Vote Prediction Over Time, Demographics Only and Demographics and Party ID, All Surveys, Random Forests.

4.3 Prognostic Power of Party ID and Ideology

Hypothesis 2 (Increasing Party ID Sorting). Next, we turn to tests of our second hypothesis that stated that partisan identification would become more informative over time, consistent with the existing literature. Models that take advantage of data on respondents' explicit partisan identification, measured on a 7-point scale, are summarized in the bottom panel of Figure 2. We also present Figure 3 which shows point estimates of three performance metrics (accuracy, AUC, and the F-1 score), which also displays these performance metrics for the remaining nested specifications. The full set of performance metrics is available in the Appendix.

We find that partisanship, jointly with basic demographics, has indeed become a significantly more prognostic variable over time. Before the 1992 election, PID-based accuracy (together with accuracy) never exceeded 85%. However, starting with Bill Clinton's re-election, the same specification generally classifies at least 90% of voting decisions correctly. The linear regression slope testing for a temporal trend in the rates displayed in the bottom panel of Figure 2 is 0.0018 with a standard error of 0.0003 (p-value < 0.001), which is in terms of effect size more than four times the coefficient size from the first specification reported above.[25]

In addition to accuracy, Figure 3 also displays the evolution of the AUC (a metric that guards against class imbalance) and the F-score (which balances precision and recall) over time. We see that—for party-inclusive specifications—both of these metrics are increasing, confirming the patterns mentioned above. We thus find empirical support for our second hypothesis, which aligns with the well-established results of an increasingly polarized U.S. electorate.

To be sure, the predictions based on the second specification leave on the table other knowable attributes of voters. We explore the implications of their inclusion, which

[25]Pooled results yield a slope of 0.0018 with a standard error of 0.0002 (p-value < 0.001).

brings us to our final hypothesis.

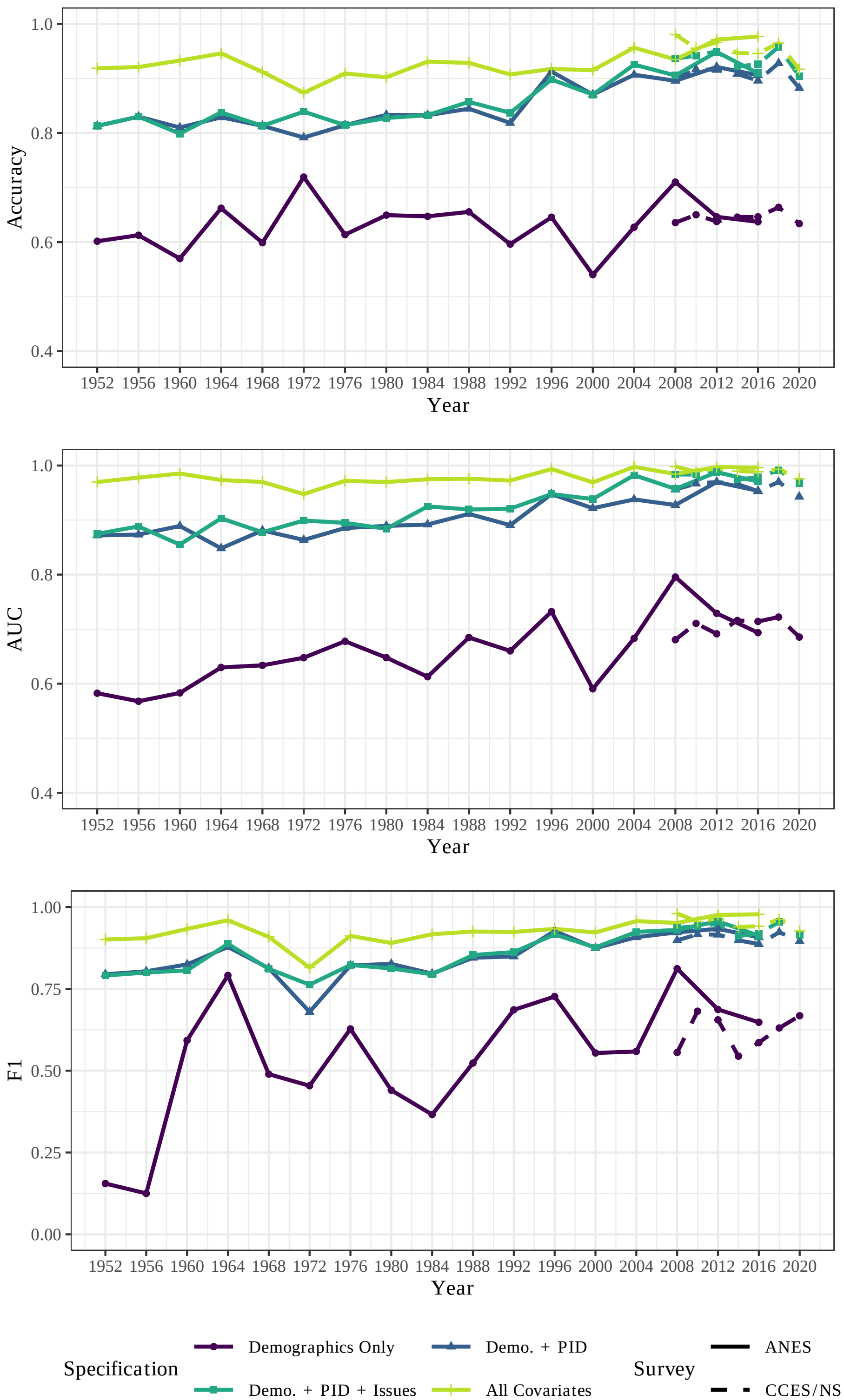


Figure 3: Performance of Presidential Vote Prediction Over Time,
All Models and All Surveys, Random Forests, Accuracy/AUC/F1 Scores

Hypothesis 3 (Sufficiency of Party ID). While partisanship is now more prognostic of vote choice, other factors continue to be important in the sense of providing additional useful information for inferring respondents' vote choice, as we can see in Figure 3. More specifically, on average, accuracy of ANES predictions improves by 0.7 percentage points when issue variables are added on top of demographics and party ID. In addition, accuracy improves by 8.2 percentage points once all other variables have been added.[26] Some examples of variables included in the fourth and final specification are non-policy opinions. For example, the top variable in terms of permutation importance for the 2018 CCES was a belief that Trump colluded with Russia to influence the 2016 election, and for ANES 2016, it was the perception about whether honesty well describes the Democratic presidential candidate.

Hence the evidence generally does not favor the third hypothesis. The patterns uncovered by these models suggest that it is possible to glean significant information about voters' behavior even after accounting for their partisanship. Views on policy issues consistently reveal more information about behavior, above and beyond partisanship. Moreover, other questions asked on public opinion surveys (occupation, subjective class identification, group attitudes, political knowledge, media consumption, beliefs, perceptions, and so on) still contain a significant amount of additional information that can be used to improve predictions about voters' behavior.

However, it is clear that in recent years, the value-added from the set of all variables in terms of prediction is, on average, decreasing. Once party ID is accounted for along with demographics, the ability of other variables to be put to good use in better predicting vote choice is more limited compared to the era with lower mass-level polarization.

[26] When we pool accuracy rates from all surveys, they improve by 1.3 percentage points when issue variables are added on top of demographics and party ID, and by 7.2 percentage points when all other variables have been added.

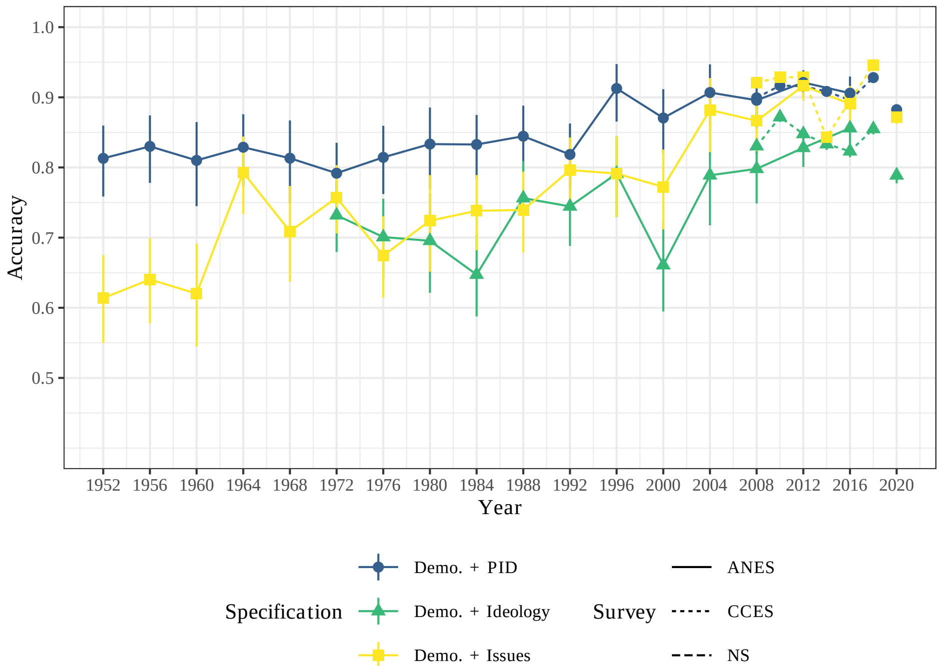


Figure 4: Performance of Presidential Vote Prediction Over Time.
Comparison of Party ID, Symbolic Ideology, and Operational Ideology, All Surveys.
The prediction algorithm employed is the Random Forests.

Party ID vs. Symbolic Ideology vs. Operational Ideology. Given the results above, it is also worth digressing to investigate the predictive power of three distinct concepts which are sometimes confused: explicit party ID, self-identified symbolic ideology along the liberal-conservative scale, and operational ideology, or specific positions on issues (Ellis and Stimson, 2012). Figure 4 shows the accuracy of predictions based on the following models: demographics and 7-point party ID (original model 2, displayed in Figure 3), demographics and 3-point symbolic ideology (liberal vs. moderate vs. conservative[27]), and demographics and operational ideology, or all issue questions. Note that some caution in interpretation must be forewarned due to the number of varying covariates between this model.

Three patterns are worth noting. First, symbolic ideology is nearly as informative as

[27]Unfortunately, the question is not available during 1952–1966 because ANES only provides the question starting from 1972.

operational ideology. This is quite striking, given that random forests extract vote-related signal from dozens of issue-related questions on the surveys. Second, in spite of a large amount of information available in the data in the operational ideology model, issues still underperform compared to party ID. Finally, the convergence of the predictive power based on demographics and operational ideology vis-à-vis the demographics and party ID model highlights the growing alignment between partisanship and policy views.

This exercise makes it clear that party labels that voters give themselves trump both types of ideologies. In the next section, turn back to demographics and describe how their relative importance has changed over time.

4.4 Variable Importance of Demographics

Given Section 4.2, we explore a related question: do demographics remain among the important variables when other variables are accounted for? If so, which ones? We calculate this with permutation-based variable importance measures. Table 1 shows which demographic characteristics remain as the top 10 variables when either (1) only party IDs are included (Specification 2), or (2) on top of that, issue variables are also included (Specification 3). The variables that appear in the same row are aligned by importance from left to right.

The demographics quickly give way to party ID in variable importance and further disappear once issues are accounted for. The only variable that consistently stays as informative about vote choice is identifying as Black. Age and education are somewhat important but are less significant and less consistent over different surveys and waves. Identifying as Black consistently retains strong prediction power on vote choice even after accounting for party IDs.

However, note that in Specification 4 (all covariates), none of the variables remain in the top 10 variables. With ANES data, you can expand the threshold up to the top 15

Year	Rank #1	Rank #2	Rank #3	Rank #4	Rank #5
1952	Black	Some college	2-year college		
1956	Income: 68-95 %tile	Income: refused	High school graduate	Age	
1960	Income: 96-100 %tile	Age	Income: 68-95 %tile	Income: 34-67 %tile	
1964	Black	2-year college	Age	Income: 34-67 %tile	Income: 68-95 %tile
1968	Black	Some college	High school graduate	Age	Income: 34-67 %tile
1972	Black	Age	Income: 34-67 %tile	2-year college	
1976	Black	Age	Income: 68-95 %tile	2-year college	
1980	Black	Income: 96-100 %tile	Hispanic	Age	2-year college
1984	Black	High school graduate	2-year college	Age	
1988	Black	Hispanic	Income: 68-95 %tile	Gender	
1992	Black	2-year college	High school graduate	Age	
1996	Black	2-year college	Hispanic	Gender	
2000	Black	Income: 96-100 %tile	2-year college	High school graduate	
2004	Black	Age	2-year college	Some college	
2008	Black	Hispanic	2-year college	High school graduate	
2012	Black	Hispanic	2-year college	Age	
2016	Black	2-year college	Some college	Hispanic	

(a) PID Included (ANES)

Year	Rank #1	Rank #2
1952	Black	
1956	Income: 68-95 %tile	
1960	Age	
1964		
1968	Black	Age
1972	Black	
1976	Black	
1980	Black	
1984	Black	
1988	Black	
1992	Black	
1996		
2000		
2004		
2008	Black	
2012	Black	
2016	Black	

(b) PID/Issues Included (ANES)

Year	Rank #1	Rank #2	Rank #V3	Rank #4
2018	Black	Age	Post-grad	High school graduate
2016	Black	Post-grad	High school graduate	4-year college
2014	Black	Age	High school graduate	4-year college
2012	Black	Age		
2010	Black	High school graduate	4-year college	Some college
2008	Black	Age	High school graduate	4-year college

(c) PID Included (CCES)

Year	Rank #1	Rank #2
2018	Black	
2016	Black	
2014	Black	Age
2012	Black	
2010	Black	
2008		

(d) PID/Issues Included (CCES)

Table 1: Demographics Remaining in the Top 10 Variables By Variable Importance, Presidential Vote Choice, Random Forests, ANES (1952–2016) and CCES (2008–2018)

variables, whereby for 1972 and 2008, identifying as Black will still be counted as one of the more important variables. With CCES data, the threshold needs to be expanded to the top 30 variables, whereby for 2014 and 2018 being Black emerges in the top important variables set. Given the results in Section 4.2, we could say that identifying as Black is a strong predictor in the sparse covariate set, but not *growing* stronger over the years.

5 Discussion: A Contested Role of Demographics

The degree of voter loyalty has substantial implications for representation. Per the typical rational choice model, campaigns will focus on catering to persuadable voters rather than pandering to a part of the electorate that will, conditional on turning out, behave predictably. If a voting bloc is "too reliable," a normative concern is that the group's interests will not be represented adequately relative to its size and importance. For example, the Democratic party has been criticized for not giving priority to issues that matter to Black voters, a key voting bloc that delivered the Biden victory (Scott, 2020). Predictability is directly linked to the leverage that a voting bloc has over parties and candidates (Axelrod, 1972).

Turning the clock back a few years, Donald Trump's unexpected victory in 2016 left the media and researchers searching for explanations of the winner's appeal, and they often zoomed in on the voting blocs that supported Trump, such as white voters without a college degree (Schaffner et al., 2018; Abramowitz and McCoy, 2019). A common concern has been that existing political cleavages between demographic groups were widening, and many of the post-election explanations focused on the deepening partisan divide by demographics such as race or education (Porter, 2016; Morgan and Lee, 2018).

Along these lines, Sides (2017) writes that "[t]he Democratic Party has an increasing advantage among nonwhite people. Among Hispanics, Democrats outnumbered Republicans by 23 points in 2002 but 36 points in 2016." The reduction of the white population and increasing racial diversity of the electorate (Teixeira et al., 2015) have been used to project the advent of the Democratic party.[28]

But other scholars have cautioned that belonging to a social group "does not necessarily prescribe a specific political outlook" (Huddy, 2018). Clearly, unexpected behaviors from some groups can lead to wrong inferences about election competitiveness; for instance, McCall and Orloff (2017) have observed that

> *... some Democratic commentators bemoaned the fact that a majority of white women had voted for Trump, and called it a kind of betrayal, underlining their expectation that women would naturally, on the basis of their gender interests and identity, support a woman with politics and policies understood to be women-friendly.*

This is particularly important because in this paper, we are using demographic *labels*, which are self-reported categorizations and not the *degrees* to which individuals identify with the group labels. To put it another way, objective inclusion in groups is not equal to the internalized sense of membership (Huddy, 2013). In this sense, the demographic labels are basis for but not equal to the social identity and affinity discussed in Green et al. (2002), Achen and Bartels (2016), or Mason (2015, 2016). Therefore, theoretically, our results show that the demographic sorting based on observable 'labels' does not extend to a better prediction of vote choices. Moreover, we do not use variables such as evangelicalism which are known to be strong identity groups outside demographics.[29]

Our results also have substantial implications for practitioners and the public. The results could help dispel the myth that demographics are fully or strongly deterministic of vote choices. Such myths are perpetuated due to the horse-race coverage of pre-election

[28] To be sure, identity politics is deployed by both sides: Abramowitz and McCoy (2019) conclude that Trump's 2016 campaign slogan of "Make America Great Again" has successfully pulled the white working-class, especially those without a college degree, away from the Democratic party. See also Lamont et al. (2017) and McQuarrie (2017).

[29] For variations of Specification 1 with geography and religion considered, see Appendix D. Note that due to data limitations—to make things consistent over 1952–2020—we could not use many important variables that have emerged as important social groups.

and post-election polls based on group membership. They have the potential to wreak real-world damage by widening the emotional gap between groups or wrong ecological inferences.

To be sure, all this is not to say that demographics should be disregarded altogether when campaign strategies are formulated. Because we investigate vote choice, the data sample itself is conditional on turnout, one of the key factors that determine a group's contribution to the voting bloc (Axelrod, 1972). Therefore, demographics can still play an important force in shaping voting behavior by playing a decisive role in turnout. For example, Krupnikov and Piston (2015) show that when there are Black candidates on the ballot, racial prejudice may prompt even co-partisan voters to stay home instead of turning out to vote. This indicates that if demographics are strong determinants of turnout in certain races, the attempt to see the relationship between demographics and vote choice may suffer from bias induced by conditioning on a collider. We leave further probes to future research.

6 Conclusion

Demographic attributes can function as markers of social identities, and membership in these social groups does, to an extent, carry political meaning. Voters' demographic characteristics help improve vote choice predictions compared to random guesses—given the five demographic characteristics, the probability of an accurate prediction with a random forests model is on average 63.5%. This, although higher than 50%, suggests that for most people, memberships in their income group, age group, gender, education group, or even ethnic group is not politically 'sorted' strongly enough to translate to particularly accurate signals about their voting decisions.

Moreover, the accuracy of predicted vote choice inferred based on voters' demographic attributes has not grown over the years. Our findings based on three sets of national surveys from 1952 to 2020 show that out-of-sample prediction accuracy has not significantly increased over the last seven decades. While the main results use random forests, the results hold when the prediction algorithm is changed to others such as a simple logistic regression or a CART. Our results are therefore not consistent with the first hypothesis: demographic sorting is not increasing.

We validate our first result by showing results from four nested model specifications, increasingly incorporating more information about respondents. In line with existing findings showing higher polarization, we show that once partisanship is no longer withheld from the set of predictors, we do observe, as expected, a massive increase in accuracy. Partisanship's diagnosticity in absolute terms is unmatched by either symbolic or operational ideology, although issue positions and partisanship's predictive power converge in recent years. Furthermore, inferring vote choice with just the combination of demographics and party ID grows easier over time.

We also note that predictions based on partisanship and demographics can further be improved by including issue positions from each survey. Adding extra variables (including non-policy features such as voters' political knowledge, media consumption, attitudes tapping into identity considerations, and other survey instruments) to the set of model features also generally yields higher accuracy. While not fully comparable due to questionnaires changing over time, the full models generally perform well, lending credibility to the method of choice. However, once partisanship has been accounted for, the added value from additional variables is decreasing over time. In recent survey waves, there is only minimal gain from richer specifications besides partisanship and demographics.

To check that demographics are not vital to the prediction, we also check permutation-based variable importance measures to see whether demographics remain among the top 10 important variables richer models. Only the strongest demographic signal of identifying as Black consistently remains among the top 10 important variables in specifications that include partisanship and/or issues. However, in the full model, where all possible covariates are used for prediction, no demographic features remain, suggesting that signals from other combinations of variables are sufficiently strong to push these out of the top 10.

Considering all this, we conclude that demographic sorting has not translated into voting behavior. Without information about respondents' partisanship, even sophisticated random forests models typically only achieve out-of-sample accuracy up to 65%, and this performance metric has not increased over the last seventy years. As the predictive power of party ID grows stronger, it dominates the signal from other covariates, diminishing their marginal predictive power. Therefore, while our results validate scholarly findings on ideological sorting and polarization, we find no support for vote-based demographic sorting.

References

Abrajano, Marisa and R. Michael Alvarez (2012). *New Faces, New Voices: The Hispanic Electorate in America*. Princeton University Press.

Abramowitz, Alan and Jennifer McCoy (2019). United States: Racial Resentment, Negative Partisanship, and Polarization in Trump's America. *The ANNALS of the American Academy of Political and Social Science 681*(1), 137–156.

Abramowitz, Alan I. and Kyle L. Saunders (2008). Is polarization a myth? *The Journal of Politics 70*(2), 542–555.

Abramowitz, Alan I. and Steven Webster (2016). The Rise of Negative Partisanship and the Nationalization of U.S. Elections in the 21st Century. *Electoral Studies 41*, 12–22.

Achen, Christopher and Larry Bartels (2016). *Democracy for Realists*. Princeton University Press.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder (2006). Purple America. *Journal of Economic Perspectives 20*(2), 97–118.

Axelrod, Robert (1972). Where the Votes Come From: An Analysis of Electoral Coalitions, 1952-1968. *The American Political Science Review 66*(1), 11–20.

Bartels, Larry M. (2006). What's the Matter with What's the Matter with Kansas? *Quarterly Journal of Political Science 1*(2), 201–226.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. University of Chicago Press.

Bertrand, Marianne and Emir Kamenica (2018). Coming Apart? Cultural Distances in the United States over Time. National Bureau of Economic Research Working Paper Number 24771. DOI: 10.3386/w24771.

Bonica, Adam (2018). Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning. *American Journal of Political Science 62*(4), 830–848.

Bougher, Lori D. (2017). The Correlates of Discord: Identity, Issue Alignment, and Political Hostility in Polarized America. *Political Behavior 39*(3), 731–762.

Bracic, Ana, Mackenzie Israel-Trummel, and Allyson F. Shortle (2019). Is Sexism for White People? Gender Stereotypes, Race, and the 2016 Presidential Election. *Political Behavior 41*(2), 281–307.

Breiman, Leo (2001). Random Forests. *Machine Learning 45*(1), 5–32.

Burden, Barry C. and David C. Kimball (2009). *Why Americans Split Their Tickets: Campaigns, Competition, and Divided Government*. University of Michigan Press.

Campbell, Angus, Philip E Converse, Warren E Miller, and Donald E Stokes (1960). *The American Voter*. John Wiley.

Carnes, Nicholas and Noam Lupu (2020). The White Working Class and the 2016 Election. *Perspectives on Politics*.

Christenson, Dino P. and Herbert F. Weisberg (2019). Bad Characters or Just More Polarization? The Rise of Extremely Negative Feelings for Presidential Candidates. *Electoral Studies 61*, 102032.

Demir, Mehmet Özer, Biagio Simonetti, Murat Alper Başaran, and Sezgin Irmak (2021). Voter Classification Based on Susceptibility to Persuasive Strategies: A Machine Learning Approach. *Social Indicators Research*, 1–16.

Efron, Bradley (2020). Prediction, Estimation, and Attribution. *Journal of the American Statistical Association 115*(530), 636–655.

Egan, Patrick J. (2013). *Partisan priorities: How issue ownership drives and distorts American politics*. Cambridge: Cambridge University Press.

Ellis, Christopher and James A. Stimson (2012). *Ideology in America*. Cambridge University Press.

Fenno, Richard F. (1978). *Home Style: House Members in Their Districts*. Longman.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy Pope (2011). *Culture War?: The Myth of a Polarized America*. Longman.

Gelman, Andrew, Lane Kenworthy, and Yu-Sung Su (2010). Income inequality and partisan voting in the United States. *Social Science Quarterly 91*(5), 1203–1219.

Green, Donald, Bradley Palmquist, and Eric Schickler (2002). *Partisan Hearts and Minds*. New Haven, CT: Yale University Press.

Griffin, Rob, William H. Frey, and Ruy Teixeira (2020). America's electoral future: The coming generational transformation. Center for American Progress. URL: https://www.brookings.edu/research/americas-electoral-future-the-coming-generational-transformation/.

Hersh, Eitan D. (2015). *Hacking the Electorate: How Campaigns Perceive Voters*. Cambridge University Press.

Hersh, Eitan D. and Clayton Nall (2016). The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records. *American Journal of Political Science 60*(2), 289–303.

Hetherington, Marc J. (2009). Putting Polarization in Perspective. *British Journal of Political Science 39*(2), 413–448.

Huddy, Leonie (2013). From Group Identity to Political Cohesion and Commitment. In *The Oxford Handbook of Political Psychology* (Second ed.). Oxford University Press.

Huddy, Leonie (2018). The group foundations of democratic political behavior. *Critical Review 2*(30), 71–86.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science 22*(1), 129–146.

Iyengar, Shanto and Sean J. Westwood (2015). Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science 59*(3), 690–707.

Jacobson, Gary C. (2013). Partisan Polarization in American Politics: A Background Paper. *Presidential Studies Quarterly 43*(4), 688–708.

Jacobson, Gary C. (2017). The Triumph of Polarized Partisanship in 2016: Donald Trump's Improbable Victory. *Political Science Quarterly 132*(1), 9–41.

Jacobson, Gary C. (2019). Extreme Referendum: Donald Trump and the 2018 Midterm Elections. *Political Science Quarterly 134*(1), 9–38.

Junn, Jane (2017). The Trump majority: white womanhood and the making of female voters in the U.S. *Politics, Groups, and Identities 5*(2), 343–352.

Kim, Seo-young Silvia, R. Michael Alvarez, and Christina M Ramirez (2020). Who voted in 2016? Using fuzzy forests to understand voter turnout. *Social Science Quarterly 45*(1), 5–11.

Krupnikov, Yanna and Spencer Piston (2015). Racial Prejudice, Partisanship, and White Turnout in Elections with Black Candidates. *Political Behavior 37*(2), 397–418.

Kuhn, Max (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software 28*(1), 1–26.

Lamont, Michèle, Bo Yun Park, and Elena Ayala-Hurtado (2017). Trump's electoral speeches and his appeal to the American white working class. *The British Journal of Sociology 68*(S1), S154–S180.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet (1944). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Duell, Sloan and Pearce.

Levendusky, Matthew (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: University of Chicago Press.

Lo, Adeline, Herman Chernoff, Tian Zheng, and Shaw-Hwa Lo (2016). Framework for making better predictions by directly estimating variables' predictivity. *Proceedings of the National Academy of Sciences 113*(50), 14277–14282.

Mason, Lilliana (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science 59*(1), 128–145.

Mason, Lilliana (2016). A Cross-Cutting Calm: How Social Sorting Drives Affective Polarization. *Public Opinion Quarterly 80*(S1), 351–377.

Mason, Lilliana (2018a). Losing Common Ground: Social Sorting and Polarization. *The Forum 16*(1), 47–66.

Mason, Lilliana (2018b). *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press.

Mason, Lilliana and Julie Wronski (2018). One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship. *Political Psychology 39*(S1), 257–277.

McCall, Leslie and Ann Shola Orloff (2017). The multidimensional politics of inequality: Taking stock of identity politics in the U.S. Presidential election of 2016. *The British Journal of Sociology 68*(S1), S34–S56.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal (2008). *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press.

McQuarrie, Michael (2017). The revolt of the Rust Belt: place and politics in the age of anger. *The British Journal of Sociology 68*(S1), S120–S152.

Montgomery, Jacob M. and Santiago Olivella (2018). Tree-based models for political science data. *American Journal of Political Science 62*(3), 729–744.

Morgan, Stephen and Jiwon Lee (2018). Trump Voters and the White Working Class. *Sociological Science 5*, 234–245.

NCSL (2020). Straight Ticket Voting States. https://www.ncsl.org/research/elections-and-campaigns/straight-ticket-voting.aspx.

Neunhoeffer, Marcel and Sebastian Sternberg (2019). How Cross-Validation Can Go Wrong and What to Do About It. *Political Analysis 27*(1), 101–106.

Norris, Pippa and Ronald Inglehart (2019). *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press.

Panagopoulos, Costas (2020). *Bases Loaded: How US Presidential Campaigns Are Changing and Why It Matters*. Oxford University Press.

Peterson, Andrew and Arthur Spirling (2018). Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization in Westminster Systems. *Political Analysis 26*(1), 120–128.

Pew Research Center (2018). Wide Gender Gap, Growing Educational Divide in Voters' Party Identification.

Porter, Eduardo (2016). Where were Trump's votes? Where the jobs weren't. New York Times. December 13.

Samii, Cyrus, Laura Paler, and Sarah Zukerman Daly (2016). Retrospective Causal Inference with Machine Learning Ensembles: An Application to Anti-recidivism Policies in Colombia. *Political Analysis 24*(4), 434–456.

Schaffner, Brian F., Matthew MacWilliams, and Tatishe Nteta (2018). Understanding white polarization in the 2016 vote for president: The sobering role of racism and sexism. *Political Science Quarterly 133*(1), 9–34.

Scott, Eugene (2020). Black voters delivered Democrats the presidency. Now they are caught in the middle of its internal battle. *The Washington Post*. URL: https://www.washingtonpost.com/politics/2020/11/14/black-voters-delivered-democrats-presidency-now-they-are-caught-middle-its-internal-battle/.

Shaw, Daron and John Petrocik (2020). *The Turnout Myth*. Oxford University Press.

Sides, John (2017). Race, Religion, and Immigration in 2016: How the Debate over American Identity Shaped the Election and What It Means for a Trump Presidency. *Democracy Fund Voter Study Group*. URL: https://www.voterstudygroup.org/publication/race-religion-immigration-2016.

Sides, John, Michael Tesler, and Lynn Vavreck (2017). How Trump Lost and Won. *Journal of Democracy 28*(2), 34–44.

Sides, John, Michael Tesler, and Lynn Vavreck (2019). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton University Press.

Stonecash, Jeff (2000). *Class And Party In American Politics*. Routledge.

Teixeira, Ruy, William H. Frey, and Rob Griffin (2015). States of Change: The Demographic Evolution of the American Electorate, 1974–2060. Technical report, Center for American Progress. URL: https://www.americanprogress.org/issues/democracy/reports/2015/02/24/107261/states-of-change/.

Tesler, Michael (2016). *Post-Racial or Most-Racial?: Race and Politics in the Obama Era*. University of Chicago Press.

Weber, Christopher and Samara Klar (2019). Exploring the Psychological Foundations of Ideological and Social Sorting. *Political Psychology 40*(S1), 215–243.

Webster, Steven W and Alan I. Abramowitz (2017). The ideological foundations of affective polarization in the u.s. electorate. *American Politics Research 45*(4), 621–647.

Williamson, Vanessa, Theda Skocpol, and John Coggin (2011). The Tea Party and the Remaking of Republican Conservatism. *Perspectives on Politics 9*(1), 25–43.

Wong, Janelle S., S. Karthick Ramakrishnan, Taeku Lee, and Jane Junn (2011). *Asian American Political Participation: Emerging Constituents and Their Political Identities*. Russell Sage Foundation.

Woon, Jonathan and Jeremy C. Pope (2008). Made in congress? testing the electoral implications of party ideological brand names. *Journal of Politics 70*(3), 823–836.

Wright, Marvin N. and Andreas Ziegler (2017). Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software 77*(1), 1–17.

YouGov (2020). HuffPost: Common Interests. URL: https://big.assets.huffingtonpost.com/athena/files/2020/12/03/5fc94b67c5b6810771273f2c.pdf.

Online Appendices (Supporting Information)

Appendix A Model Outputs and Performance: ANES **2**
A.1 Accuracy Range Comparison Between Methods 2
A.2 Full Performance Tables 3
A.3 ROC Curves 9

Appendix B Model Outputs and Performance: CCES **11**
B.1 Accuracy Range Comparison Between Methods 11
B.2 Full Performance Tables 12
B.3 ROC Curves 14

Appendix C Nationscape **16**
C.1 Full Performance Tables 16
C.2 ROC Curves 16

Appendix D Alternative Specifications for Demographics Only **18**

Appendix E Benchmarking Accuracy Rates **19**

Appendix A Model Outputs and Performance: ANES

Ranges of accuracy rates for all 4 specifications, broken down by algorithm, are are displayed in Table A.1. Table A.2 shows the AUC, accuracy, recall, precision, and the F-score for RF-based models from 1952 onwards, based on the ANES data. These performance metrics are included in Table A.3 for logit models, and A.4 for classification trees.

Performance of logistic regressions, classification trees, and Random Forests which include respondents' demographic characteristics as features is compared for each election between 1952 and 2016 in Figure A.1 via ROC curves. The set of features is then extended to also include partisanship in Figure A.2, issue positions (Figure A.3), and all available features (Figure A.4).

A.1 Accuracy Range Comparison Between Methods

Variable Specification	Year	Logit	CART	RF
Demographics Only	2016	[0.6139, 0.6972]	[0.5806, 0.6656]	[0.5943, 0.6786]
Demo. + PID	2016	[0.8733, 0.9262]	[0.7634, 0.8339]	[0.8776, 0.9296]
Demo. + PID + Issues	2016	[0.8948, 0.9429]	[0.7655, 0.8357]	[0.8818, 0.9330]
All Covariates	2016	[0.8733, 0.9262]	[0.9013, 0.9479]	[0.9601, 0.9880]
Demographics Only	2012	[0.6360, 0.7009]	[0.6385, 0.7032]	[0.6129, 0.6788]
Demo. + PID	2012	[0.9047, 0.9417]	[0.8133, 0.8641]	[0.9008, 0.9385]
Demo. + PID + Issues	2012	[0.9127, 0.9480]	[0.8298, 0.8786]	[0.9314, 0.9626]
All Covariates	2012	[0.8929, 0.9321]	[0.9207, 0.9543]	[0.9576, 0.9815]
Demographics Only	2008	[0.7106, 0.8087]	[0.6831, 0.7846]	[0.6559, 0.7602]
Demo. + PID	2008	[0.8560, 0.9276]	[0.7981, 0.8824]	[0.8560, 0.9276]
Demo. + PID + Issues	2008	[0.8232, 0.9024]	[0.7487, 0.8415]	[0.8672, 0.9358]
All Covariates	2008	[0.4475, 0.5622]	[0.8414, 0.9165]	[0.9012, 0.9598]
Demographics Only	2004	[0.6248, 0.7713]	[0.5668, 0.7196]	[0.5477, 0.7021]
Demo. + PID	2004	[0.8810, 0.9654]	[0.8147, 0.9224]	[0.8510, 0.9469]
Demo. + PID + Issues	2004	[0.7794, 0.8969]	[0.8363, 0.9373]	[0.8734, 0.9609]
All Covariates	2004	[0.4294, 0.5888]	[0.9044, 0.9783]	[0.9125, 0.9823]
Demographics Only	2000	[0.4681, 0.6024]	[0.4948, 0.6285]	[0.4725, 0.6068]
Demo. + PID	2000	[0.8346, 0.9228]	[0.7358, 0.8456]	[0.8194, 0.9116]
Demo. + PID + Issues	2000	[0.6927, 0.8094]	[0.7358, 0.8456]	[0.8194, 0.9116]
All Covariates	2000	[0.4459, 0.5805]	[0.7895, 0.8887]	[0.8707, 0.9482]
Demographics Only	1996	[0.6112, 0.7428]	[0.5563, 0.6925]	[0.5762, 0.7108]
Demo. + PID	1996	[0.8712, 0.9512]	[0.7932, 0.8954]	[0.8654, 0.9474]
Demo. + PID + Issues	1996	[0.7504, 0.8618]	[0.7932, 0.8954]	[0.8484, 0.9358]
All Covariates	1996	[0.4537, 0.5941]	[0.7717, 0.8787]	[0.8712, 0.9512]
Demographics Only	1992	[0.5314, 0.6517]	[0.5351, 0.6553]	[0.5351, 0.6553]
Demo. + PID	1992	[0.7834, 0.8758]	[0.6998, 0.8056]	[0.7673, 0.8626]
Demo. + PID + Issues	1992	[0.6998, 0.8056]	[0.6998, 0.8056]	[0.7875, 0.8790]
All Covariates	1992	[0.4756, 0.5977]	[0.7834, 0.8758]	[0.8664, 0.9392]
Demographics Only	1988	[0.5741, 0.6997]	[0.5870, 0.7117]	[0.5913, 0.7157]
Demo. + PID	1988	[0.8061, 0.8990]	[0.6789, 0.7941]	[0.7921, 0.8881]

Demo. + PID + Issues	1988	[0.7599, 0.8623]	[0.6789, 0.7941]	[0.8061, 0.8990]
All Covariates	1988	[0.4099, 0.5403]	[0.8392, 0.9239]	[0.8881, 0.9578]
Demographics Only	1984	[0.5876, 0.7037]	[0.5876, 0.7037]	[0.5876, 0.7037]
Demo. + PID	1984	[0.8032, 0.8908]	[0.8113, 0.8972]	[0.7833, 0.8749]
Demo. + PID + Issues	1984	[0.7595, 0.8554]	[0.6974, 0.8026]	[0.7833, 0.8749]
All Covariates	1984	[0.5028, 0.6231]	[0.8113, 0.8972]	[0.8942, 0.9579]
Demographics Only	1980	[0.6033, 0.7469]	[0.5559, 0.7039]	[0.5736, 0.7201]
Demo. + PID	1980	[0.7568, 0.8756]	[0.7759, 0.8903]	[0.7695, 0.8854]
Demo. + PID + Issues	1980	[0.7504, 0.8707]	[0.7759, 0.8903]	[0.7631, 0.8805]
All Covariates	1980	[0.4121, 0.5653]	[0.7504, 0.8707]	[0.8482, 0.9420]
Demographics Only	1976	[0.5597, 0.6800]	[0.5290, 0.6508]	[0.5520, 0.6727]
Demo. + PID	1976	[0.7786, 0.8729]	[0.6574, 0.7695]	[0.7622, 0.8594]
Demo. + PID + Issues	1976	[0.7622, 0.8594]	[0.6574, 0.7695]	[0.7622, 0.8594]
All Covariates	1976	[0.4606, 0.5843]	[0.7663, 0.8628]	[0.8678, 0.9409]
Demographics Only	1972	[0.6564, 0.7592]	[0.6663, 0.7680]	[0.6663, 0.7680]
Demo. + PID	1972	[0.7564, 0.8466]	[0.6630, 0.7651]	[0.7429, 0.8352]
Demo. + PID + Issues	1972	[0.7666, 0.8552]	[0.7127, 0.8091]	[0.7940, 0.8778]
All Covariates	1972	[0.4671, 0.5798]	[0.7362, 0.8294]	[0.8322, 0.9083]
Demographics Only	1968	[0.5517, 0.6968]	[0.5574, 0.7020]	[0.5238, 0.6707]
Demo. + PID	1968	[0.7731, 0.8859]	[0.7731, 0.8859]	[0.7489, 0.8670]
Demo. + PID + Issues	1968	[0.7428, 0.8622]	[0.7731, 0.8859]	[0.7489, 0.8670]
All Covariates	1968	[0.5461, 0.6916]	[0.8227, 0.9227]	[0.8612, 0.9489]
Demographics Only	1964	[0.5773, 0.7071]	[0.6098, 0.7368]	[0.5958, 0.7241]
Demo. + PID	1964	[0.7826, 0.8838]	[0.7141, 0.8280]	[0.7727, 0.8759]
Demo. + PID + Issues	1964	[0.7189, 0.8320]	[0.7286, 0.8401]	[0.7826, 0.8838]
All Covariates	1964	[0.4324, 0.5676]	[0.8127, 0.9069]	[0.9075, 0.9718]
Demographics Only	1960	[0.5107, 0.6595]	[0.4218, 0.5728]	[0.4939, 0.6435]
Demo. + PID	1960	[0.7387, 0.8599]	[0.7084, 0.8353]	[0.7448, 0.8647]
Demo. + PID + Issues	1960	[0.7448, 0.8647]	[0.7084, 0.8353]	[0.7326, 0.8550]
All Covariates	1960	[0.5561, 0.7020]	[0.8723, 0.9566]	[0.8858, 0.9649]
Demographics Only	1956	[0.5376, 0.6616]	[0.5256, 0.6502]	[0.5496, 0.6730]
Demo. + PID	1956	[0.8041, 0.8949]	[0.7737, 0.8707]	[0.7780, 0.8742]
Demo. + PID + Issues	1956	[0.8085, 0.8983]	[0.7737, 0.8707]	[0.7780, 0.8742]
All Covariates	1956	[0.7266, 0.8318]	[0.7910, 0.8846]	[0.8806, 0.9510]
Demographics Only	1952	[0.5047, 0.6319]	[0.5457, 0.6711]	[0.5375, 0.6633]
Demo. + PID	1952	[0.7630, 0.8633]	[0.6975, 0.8084]	[0.7586, 0.8597]
Demo. + PID + Issues	1952	[0.7674, 0.8669]	[0.6975, 0.8084]	[0.7586, 0.8597]
All Covariates	1952	[0.6589, 0.7747]	[0.7763, 0.8741]	[0.8772, 0.9496]

Table A.1: Accuracy Range Comparison, Presidential Vote Choice, ANES 1952–2016

A.2 Full Performance Tables

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.6935	0.6372	[0.5943, 0.6786]	0.6566	0.6397	0.6480
Demo. + PID	2016	0.9537	0.9060	[0.8776, 0.9296]	0.9339	0.8824	0.9074
Demo. + PID + Issues	2016	0.9707	0.9098	[0.8818, 0.9330]	0.9245	0.9007	0.9125

All Observables	2016	0.9962	0.9770	[0.9601, 0.9880]	0.9815	0.9743	0.9779
Demographics	2012	0.7290	0.6464	[0.6129, 0.6788]	0.7271	0.6513	0.6871
Demo. + PID	2012	0.9697	0.9211	[0.9008, 0.9385]	0.9374	0.9299	0.9336
Demo. + PID + Issues	2012	0.9877	0.9486	[0.9314, 0.9626]	0.9597	0.9539	0.9568
All Observables	2012	0.9971	0.9713	[0.9576, 0.9815]	0.9684	0.9840	0.9761
Demographics	2008	0.7956	0.7101	[0.6559, 0.7602]	0.7164	0.9366	0.8118
Demo. + PID	2008	0.9275	0.8958	[0.8560, 0.9276]	0.9139	0.9317	0.9227
Demo. + PID + Issues	2008	0.9571	0.9055	[0.8672, 0.9358]	0.9190	0.9415	0.9301
All Observables	2008	0.9845	0.9349	[0.9012, 0.9598]	0.9469	0.9561	0.9515
Demographics	2004	0.6830	0.6273	[0.5477, 0.7021]	0.6667	0.4810	0.5588
Demo. + PID	2004	0.9378	0.9068	[0.8510, 0.9469]	0.8810	0.9367	0.9080
Demo. + PID + Issues	2004	0.9823	0.9255	[0.8734, 0.9609]	0.9241	0.9241	0.9241
All Observables	2004	0.9975	0.9565	[0.9125, 0.9823]	0.9286	0.9873	0.9571
Demographics	2000	0.5905	0.5402	[0.4725, 0.6068]	0.5664	0.5424	0.5541
Demo. + PID	2000	0.9217	0.8705	[0.8194, 0.9116]	0.8938	0.8559	0.8745
Demo. + PID + Issues	2000	0.9384	0.8705	[0.8194, 0.9116]	0.8803	0.8729	0.8766
All Observables	2000	0.9691	0.9152	[0.8707, 0.9482]	0.8898	0.9576	0.9224
Demographics	1996	0.7321	0.6456	[0.5762, 0.7108]	0.6599	0.8083	0.7266
Demo. + PID	1996	0.9476	0.9126	[0.8654, 0.9474]	0.9048	0.9500	0.9268
Demo. + PID + Issues	1996	0.9477	0.8981	[0.8484, 0.9358]	0.8779	0.9583	0.9163
All Observables	1996	0.9935	0.9175	[0.8712, 0.9512]	0.8815	0.9917	0.9333
Demographics	1992	0.6601	0.5963	[0.5351, 0.6553]	0.6296	0.7532	0.6859
Demo. + PID	1992	0.8907	0.8185	[0.7673, 0.8626]	0.8263	0.8734	0.8492
Demo. + PID + Issues	1992	0.9207	0.8370	[0.7875, 0.8790]	0.8519	0.8734	0.8625
All Observables	1992	0.9724	0.9074	[0.8664, 0.9392]	0.8889	0.9620	0.9240
Demographics	1988	0.6847	0.6555	[0.5913, 0.7157]	0.7500	0.4018	0.5233
Demo. + PID	1988	0.9113	0.8445	[0.7921, 0.8881]	0.7953	0.9018	0.8452
Demo. + PID + Issues	1988	0.9194	0.8571	[0.8061, 0.8990]	0.8250	0.8839	0.8534
All Observables	1988	0.9758	0.9286	[0.8881, 0.9578]	0.9130	0.9375	0.9251
Demographics	1984	0.6127	0.6473	[0.5876, 0.7037]	0.7368	0.2435	0.3660
Demo. + PID	1984	0.8918	0.8327	[0.7833, 0.8749]	0.8108	0.7826	0.7965
Demo. + PID + Issues	1984	0.9251	0.8327	[0.7833, 0.8749]	0.8165	0.7739	0.7946
All Observables	1984	0.9748	0.9309	[0.8942, 0.9579]	0.9211	0.9130	0.9170
Demographics	1980	0.6478	0.6494	[0.5736, 0.7201]	0.7273	0.3158	0.4404
Demo. + PID	1980	0.8892	0.8333	[0.7695, 0.8854]	0.7582	0.9079	0.8263
Demo. + PID + Issues	1980	0.8841	0.8276	[0.7631, 0.8805]	0.7738	0.8553	0.8125
All Observables	1980	0.9697	0.9023	[0.8482, 0.9420]	0.8734	0.9079	0.8903
Demographics	1976	0.6778	0.6136	[0.5520, 0.6727]	0.6143	0.6418	0.6277
Demo. + PID	1976	0.8854	0.8144	[0.7622, 0.8594]	0.8014	0.8433	0.8218
Demo. + PID + Issues	1976	0.8949	0.8144	[0.7622, 0.8594]	0.7972	0.8507	0.8231
All Observables	1976	0.9718	0.9091	[0.8678, 0.9409]	0.8986	0.9254	0.9118
Demographics	1972	0.6476	0.7192	[0.6663, 0.7680]	0.7400	0.3274	0.4540
Demo. + PID	1972	0.8635	0.7918	[0.7429, 0.8352]	0.7527	0.6195	0.6796
Demo. + PID + Issues	1972	0.8991	0.8391	[0.7940, 0.8778]	0.8039	0.7257	0.7628
All Observables	1972	0.9476	0.8738	[0.8322, 0.9083]	0.8544	0.7788	0.8148
Demographics	1968	0.6336	0.5989	[0.5238, 0.6707]	0.5932	0.4167	0.4895
Demo. + PID	1968	0.8810	0.8132	[0.7489, 0.8670]	0.7551	0.8810	0.8132

Demo. + PID + Issues	1968	0.8776	0.8132	[0.7489, 0.8670]	0.7604	0.8690	0.8111
All Observables	1968	0.9698	0.9121	[0.8612, 0.9489]	0.8696	0.9524	0.9091
Demographics	1964	0.6298	0.6622	[0.5958, 0.7241]	0.6794	0.9467	0.7911
Demo. + PID	1964	0.8482	0.8288	[0.7727, 0.8759]	0.8457	0.9133	0.8782
Demo. + PID + Issues	1964	0.9031	0.8378	[0.7826, 0.8838]	0.8353	0.9467	0.8875
All Observables	1964	0.9733	0.9459	[0.9075, 0.9718]	0.9662	0.9533	0.9597
Demographics	1960	0.5830	0.5698	[0.4939, 0.6435]	0.5600	0.6292	0.5926
Demo. + PID	1960	0.8891	0.8101	[0.7448, 0.8647]	0.7619	0.8989	0.8247
Demo. + PID + Issues	1960	0.8551	0.7989	[0.7326, 0.8550]	0.7732	0.8427	0.8065
All Observables	1960	0.9852	0.9330	[0.8858, 0.9649]	0.9231	0.9438	0.9333
Demographics	1956	0.5678	0.6126	[0.5496, 0.6730]	0.7000	0.0686	0.1250
Demo. + PID	1956	0.8734	0.8300	[0.7780, 0.8742]	0.7521	0.8627	0.8037
Demo. + PID + Issues	1956	0.8883	0.8300	[0.7780, 0.8742]	0.7611	0.8431	0.8000
All Observables	1956	0.9779	0.9209	[0.8806, 0.9510]	0.8796	0.9314	0.9048
Demographics	1952	0.5825	0.6016	[0.5375, 0.6633]	0.6923	0.0874	0.1552
Demo. + PID	1952	0.8719	0.8130	[0.7586, 0.8597]	0.7355	0.8641	0.7946
Demo. + PID + Issues	1952	0.8748	0.8130	[0.7586, 0.8597]	0.7436	0.8447	0.7909
All Observables	1952	0.9698	0.9187	[0.8772, 0.9496]	0.9192	0.8835	0.9010

Table A.2: Performance Metrics, Presidential Vote Choice, Random Forests, ANES 1952–2016

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.7034	0.6564	[0.6139, 0.6972]	0.6768	0.6544	0.6654
Demo. + PID	2016	0.9522	0.9021	[0.8733, 0.9262]	0.9202	0.8897	0.9047
Demo. + PID + Issues	2016	0.9581	0.9213	[0.8948, 0.9429]	0.9459	0.9007	0.9228
All Observables	2016	0.9576	0.9021	[0.8733, 0.9262]	0.9108	0.9007	0.9057
Demographics	2012	0.7342	0.6691	[0.6360, 0.7009]	0.7832	0.6152	0.6891
Demo. + PID	2012	0.9744	0.9247	[0.9047, 0.9417]	0.9523	0.9198	0.9358
Demo. + PID + Issues	2012	0.9719	0.9319	[0.9127, 0.9480]	0.9547	0.9299	0.9421
All Observables	2012	0.9677	0.9140	[0.8929, 0.9321]	0.9313	0.9238	0.9276
Demographics	2008	0.8354	0.7622	[0.7106, 0.8087]	0.8300	0.8098	0.8198
Demo. + PID	2008	0.9522	0.8958	[0.8560, 0.9276]	0.9139	0.9317	0.9227
Demo. + PID + Issues	2008	0.9194	0.8664	[0.8232, 0.9024]	0.8981	0.9024	0.9002
All Observables	2008	0.5162	0.5049	[0.4475, 0.5622]	0.6803	0.4878	0.5682
Demographics	2004	0.7228	0.7019	[0.6248, 0.7713]	0.8163	0.5063	0.6250
Demo. + PID	2004	0.9437	0.9317	[0.8810, 0.9654]	0.9048	0.9620	0.9325
Demo. + PID + Issues	2004	0.8452	0.8447	[0.7794, 0.8969]	0.8214	0.8734	0.8466
All Observables	2004	0.5037	0.5093	[0.4294, 0.5888]	0.5000	0.4557	0.4768
Demographics	2000	0.5817	0.5357	[0.4681, 0.6024]	0.5714	0.4746	0.5185
Demo. + PID	2000	0.9159	0.8839	[0.8346, 0.9228]	0.9107	0.8644	0.8870
Demo. + PID + Issues	2000	0.7545	0.7545	[0.6927, 0.8094]	0.7739	0.7542	0.7639
All Observables	2000	0.5062	0.5134	[0.4459, 0.5805]	0.5385	0.5339	0.5362
Demographics	1996	0.7351	0.6796	[0.6112, 0.7428]	0.7109	0.7583	0.7339
Demo. + PID	1996	0.9451	0.9175	[0.8712, 0.9512]	0.9256	0.9333	0.9295
Demo. + PID + Issues	1996	0.8893	0.8107	[0.7504, 0.8618]	0.8092	0.8833	0.8446

All Observables	1996	0.5307	0.5243	[0.4537, 0.5941]	0.6058	0.5250	0.5625
Demographics	1992	0.6548	0.5926	[0.5314, 0.6517]	0.6319	0.7278	0.6765
Demo. + PID	1992	0.9179	0.8333	[0.7834, 0.8758]	0.8383	0.8861	0.8615
Demo. + PID + Issues	1992	0.7532	0.7556	[0.6998, 0.8056]	0.8026	0.7722	0.7871
All Observables	1992	0.5403	0.5370	[0.4756, 0.5977]	0.6170	0.5506	0.5819
Demographics	1988	0.6548	0.6387	[0.5741, 0.6997]	0.6970	0.4107	0.5169
Demo. + PID	1988	0.9191	0.8571	[0.8061, 0.8990]	0.8250	0.8839	0.8534
Demo. + PID + Issues	1988	0.8800	0.8151	[0.7599, 0.8623]	0.8148	0.7857	0.8000
All Observables	1988	0.4727	0.4748	[0.4099, 0.5403]	0.4414	0.4375	0.4395
Demographics	1984	0.6295	0.6473	[0.5876, 0.7037]	0.6607	0.3217	0.4327
Demo. + PID	1984	0.8929	0.8509	[0.8032, 0.8908]	0.8246	0.8174	0.8210
Demo. + PID + Issues	1984	0.8745	0.8109	[0.7595, 0.8554]	0.7788	0.7652	0.7719
All Observables	1984	0.5649	0.5636	[0.5028, 0.6231]	0.4823	0.5913	0.5313
Demographics	1980	0.6637	0.6782	[0.6033, 0.7469]	0.8571	0.3158	0.4615
Demo. + PID	1980	0.9013	0.8218	[0.7568, 0.8756]	0.7586	0.8684	0.8098
Demo. + PID + Issues	1980	0.9070	0.8161	[0.7504, 0.8707]	0.7821	0.8026	0.7922
All Observables	1980	0.4902	0.4885	[0.4121, 0.5653]	0.4235	0.4737	0.4472
Demographics	1976	0.7005	0.6212	[0.5597, 0.6800]	0.6214	0.6493	0.6350
Demo. + PID	1976	0.8896	0.8295	[0.7786, 0.8729]	0.8201	0.8507	0.8352
Demo. + PID + Issues	1976	0.9107	0.8144	[0.7622, 0.8594]	0.8058	0.8358	0.8205
All Observables	1976	0.5280	0.5227	[0.4606, 0.5843]	0.5270	0.5821	0.5532
Demographics	1972	0.7058	0.7098	[0.6564, 0.7592]	0.7692	0.2655	0.3947
Demo. + PID	1972	0.8616	0.8044	[0.7564, 0.8466]	0.7429	0.6903	0.7156
Demo. + PID + Issues	1972	0.8793	0.8139	[0.7666, 0.8552]	0.7455	0.7257	0.7354
All Observables	1972	0.5180	0.5237	[0.4671, 0.5798]	0.3623	0.4425	0.3984
Demographics	1968	0.6566	0.6264	[0.5517, 0.6968]	0.6600	0.3929	0.4925
Demo. + PID	1968	0.9082	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
Demo. + PID + Issues	1968	0.8562	0.8077	[0.7428, 0.8622]	0.7882	0.7976	0.7929
All Observables	1968	0.6199	0.6209	[0.5461, 0.6916]	0.5824	0.6310	0.6057
Demographics	1964	0.6643	0.6441	[0.5773, 0.7071]	0.6859	0.8733	0.7683
Demo. + PID	1964	0.8744	0.8378	[0.7826, 0.8838]	0.8562	0.9133	0.8839
Demo. + PID + Issues	1964	0.8598	0.7793	[0.7189, 0.8320]	0.8217	0.8600	0.8404
All Observables	1964	0.5129	0.5000	[0.4324, 0.5676]	0.6857	0.4800	0.5647
Demographics	1960	0.5910	0.5866	[0.5107, 0.6595]	0.5714	0.6742	0.6186
Demo. + PID	1960	0.8858	0.8045	[0.7387, 0.8599]	0.7547	0.8989	0.8205
Demo. + PID + Issues	1960	0.8752	0.8101	[0.7448, 0.8647]	0.7723	0.8764	0.8211
All Observables	1960	0.6551	0.6313	[0.5561, 0.7020]	0.6386	0.5955	0.6163
Demographics	1956	0.5745	0.6008	[0.5376, 0.6616]	0.5152	0.1667	0.2519
Demo. + PID	1956	0.9019	0.8538	[0.8041, 0.8949]	0.7778	0.8922	0.8311
Demo. + PID + Issues	1956	0.9093	0.8577	[0.8085, 0.8983]	0.7895	0.8824	0.8333
All Observables	1956	0.8489	0.7826	[0.7266, 0.8318]	0.7196	0.7549	0.7368
Demographics	1952	0.5539	0.5691	[0.5047, 0.6319]	0.4615	0.1748	0.2535
Demo. + PID	1952	0.8757	0.8171	[0.7630, 0.8633]	0.7458	0.8544	0.7964
Demo. + PID + Issues	1952	0.8787	0.8211	[0.7674, 0.8669]	0.7521	0.8544	0.8000
All Observables	1952	0.7472	0.7195	[0.6589, 0.7747]	0.6417	0.7476	0.6906

Table A.3: Performance Metrics, Presidential Vote Choice, Logit, ANES 1952–2016

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics	2016	0.6751	0.6238	[0.5806, 0.6656]	0.6250	0.6985	0.6597
Demo. + PID	2016	0.8007	0.8004	[0.7634, 0.8339]	0.7414	0.9485	0.8323
Demo. + PID + Issues	2016	0.8535	0.8023	[0.7655, 0.8357]	0.7752	0.8750	0.8221
All Observables	2016	0.9519	0.9271	[0.9013, 0.9479]	0.9398	0.9191	0.9294
Demographics	2012	0.7240	0.6714	[0.6385, 0.7032]	0.8522	0.5431	0.6634
Demo. + PID	2012	0.8082	0.8399	[0.8133, 0.8641]	0.7977	0.9800	0.8795
Demo. + PID + Issues	2012	0.8909	0.8554	[0.8298, 0.8786]	0.8921	0.8617	0.8767
All Observables	2012	0.9504	0.9391	[0.9207, 0.9543]	0.9590	0.9379	0.9483
Demographics	2008	0.8011	0.7362	[0.6831, 0.7846]	0.8647	0.7171	0.7840
Demo. + PID	2008	0.7849	0.8436	[0.7981, 0.8824]	0.8285	0.9659	0.8919
Demo. + PID + Issues	2008	0.7727	0.7980	[0.7487, 0.8415]	0.8295	0.8780	0.8531
All Observables	2008	0.8884	0.8827	[0.8414, 0.9165]	0.8894	0.9415	0.9147
Demographics	2004	0.6639	0.6460	[0.5668, 0.7196]	0.7200	0.4557	0.5581
Demo. + PID	2004	0.8780	0.8758	[0.8147, 0.9224]	0.7980	1.0000	0.8876
Demo. + PID + Issues	2004	0.8983	0.8944	[0.8363, 0.9373]	0.8605	0.9367	0.8970
All Observables	2004	0.9712	0.9503	[0.9044, 0.9783]	0.9277	0.9747	0.9506
Demographics	2000	0.5892	0.5625	[0.4948, 0.6285]	0.5794	0.6186	0.5984
Demo. + PID	2000	0.7910	0.7946	[0.7358, 0.8456]	0.7338	0.9576	0.8309
Demo. + PID + Issues	2000	0.7910	0.7946	[0.7358, 0.8456]	0.7338	0.9576	0.8309
All Observables	2000	0.8396	0.8438	[0.7895, 0.8887]	0.8217	0.8983	0.8583
Demographics	1996	0.6553	0.6262	[0.5563, 0.6925]	0.7216	0.5833	0.6452
Demo. + PID	1996	0.8274	0.8495	[0.7932, 0.8954]	0.8112	0.9667	0.8821
Demo. + PID + Issues	1996	0.8274	0.8495	[0.7932, 0.8954]	0.8112	0.9667	0.8821
All Observables	1996	0.8238	0.8301	[0.7717, 0.8787]	0.7815	0.9833	0.8708
Demographics	1992	0.6430	0.5963	[0.5351, 0.6553]	0.6339	0.7342	0.6804
Demo. + PID	1992	0.7289	0.7556	[0.6998, 0.8056]	0.7277	0.9304	0.8167
Demo. + PID + Issues	1992	0.7289	0.7556	[0.6998, 0.8056]	0.7277	0.9304	0.8167
All Observables	1992	0.8367	0.8333	[0.7834, 0.8758]	0.8122	0.9304	0.8673
Demographics	1988	0.6353	0.6513	[0.5870, 0.7117]	0.8537	0.3125	0.4575
Demo. + PID	1988	0.7554	0.7395	[0.6789, 0.7941]	0.6524	0.9554	0.7754
Demo. + PID + Issues	1988	0.7554	0.7395	[0.6789, 0.7941]	0.6524	0.9554	0.7754
All Observables	1988	0.8878	0.8866	[0.8392, 0.9239]	0.9126	0.8393	0.8744
Demographics	1984	0.5819	0.6473	[0.5876, 0.7037]	0.8750	0.1826	0.3022
Demo. + PID	1984	0.8886	0.8582	[0.8113, 0.8972]	0.8167	0.8522	0.8340
Demo. + PID + Issues	1984	0.7211	0.7527	[0.6974, 0.8026]	0.8133	0.5304	0.6421
All Observables	1984	0.9024	0.8582	[0.8113, 0.8972]	0.7923	0.8957	0.8408
Demographics	1980	0.5963	0.6322	[0.5559, 0.7039]	0.7000	0.2763	0.3962
Demo. + PID	1980	0.8820	0.8391	[0.7759, 0.8903]	0.7500	0.9474	0.8372
Demo. + PID + Issues	1980	0.8820	0.8391	[0.7759, 0.8903]	0.7500	0.9474	0.8372
All Observables	1980	0.8300	0.8161	[0.7504, 0.8707]	0.7895	0.7895	0.7895
Demographics	1976	0.6534	0.5909	[0.5290, 0.6508]	0.5756	0.7388	0.6471
Demo. + PID	1976	0.7224	0.7159	[0.6574, 0.7695]	0.6545	0.9328	0.7692
Demo. + PID + Issues	1976	0.7224	0.7159	[0.6574, 0.7695]	0.6545	0.9328	0.7692
All Observables	1976	0.8036	0.8182	[0.7663, 0.8628]	0.7905	0.8731	0.8298
Demographics	1972	0.6779	0.7192	[0.6663, 0.7680]	0.6875	0.3894	0.4972

Demo. + PID	1972	0.6116	0.7161	[0.6630, 0.7651]	0.8485	0.2478	0.3836
Demo. + PID + Issues	1972	0.8397	0.7634	[0.7127, 0.8091]	0.6638	0.6814	0.6725
All Observables	1972	0.8244	0.7855	[0.7362, 0.8294]	0.8689	0.4690	0.6092
Demographics	1968	0.6014	0.6319	[0.5574, 0.7020]	0.7297	0.3214	0.4463
Demo. + PID	1968	0.8469	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
Demo. + PID + Issues	1968	0.8469	0.8352	[0.7731, 0.8859]	0.7700	0.9167	0.8370
All Observables	1968	0.9181	0.8791	[0.8227, 0.9227]	0.8690	0.8690	0.8690
Demographics	1964	0.5834	0.6757	[0.6098, 0.7368]	0.6893	0.9467	0.7978
Demo. + PID	1964	0.6713	0.7748	[0.7141, 0.8280]	0.7632	0.9667	0.8529
Demo. + PID + Issues	1964	0.7675	0.7883	[0.7286, 0.8401]	0.8366	0.8533	0.8449
All Observables	1964	0.8862	0.8649	[0.8127, 0.9069]	0.8448	0.9800	0.9074
Demographics	1960	0.4900	0.4972	[0.4218, 0.5728]	0.4968	0.8764	0.6341
Demo. + PID	1960	0.7848	0.7765	[0.7084, 0.8353]	0.7025	0.9551	0.8095
Demo. + PID + Issues	1960	0.7848	0.7765	[0.7084, 0.8353]	0.7025	0.9551	0.8095
All Observables	1960	0.9361	0.9218	[0.8723, 0.9566]	0.8947	0.9551	0.9239
Demographics	1956	0.5613	0.5889	[0.5256, 0.6502]	0.4833	0.2843	0.3580
Demo. + PID	1956	0.8941	0.8261	[0.7737, 0.8707]	0.7544	0.8431	0.7963
Demo. + PID + Issues	1956	0.8961	0.8261	[0.7737, 0.8707]	0.7685	0.8137	0.7905
All Observables	1956	0.8834	0.8419	[0.7910, 0.8846]	0.7981	0.8137	0.8058
Demographics	1952	0.5795	0.6098	[0.5457, 0.6711]	0.5778	0.2524	0.3514
Demo. + PID	1952	0.7865	0.7561	[0.6975, 0.8084]	0.6387	0.9612	0.7674
Demo. + PID + Issues	1952	0.7865	0.7561	[0.6975, 0.8084]	0.6387	0.9612	0.7674
All Observables	1952	0.8602	0.8293	[0.7763, 0.8741]	0.7607	0.8641	0.8091

Table A.4: Performance Metrics, Presidential Vote Choice, CART, ANES 1952–2016

A.3 ROC Curves

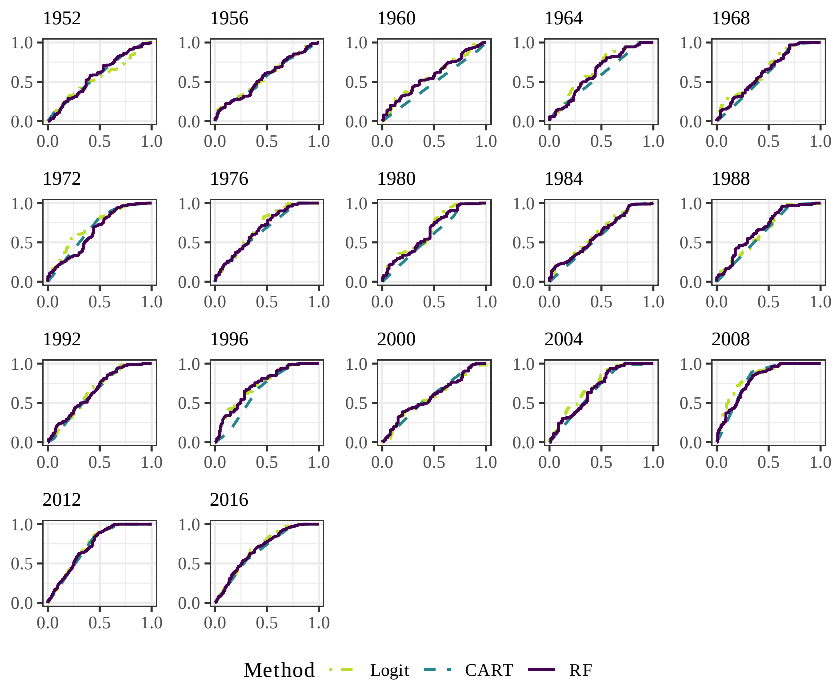


Figure A.1: ROC Curves for Presidential Vote Choice, Demographics Only, Comparison Between Logit, CART, and RF

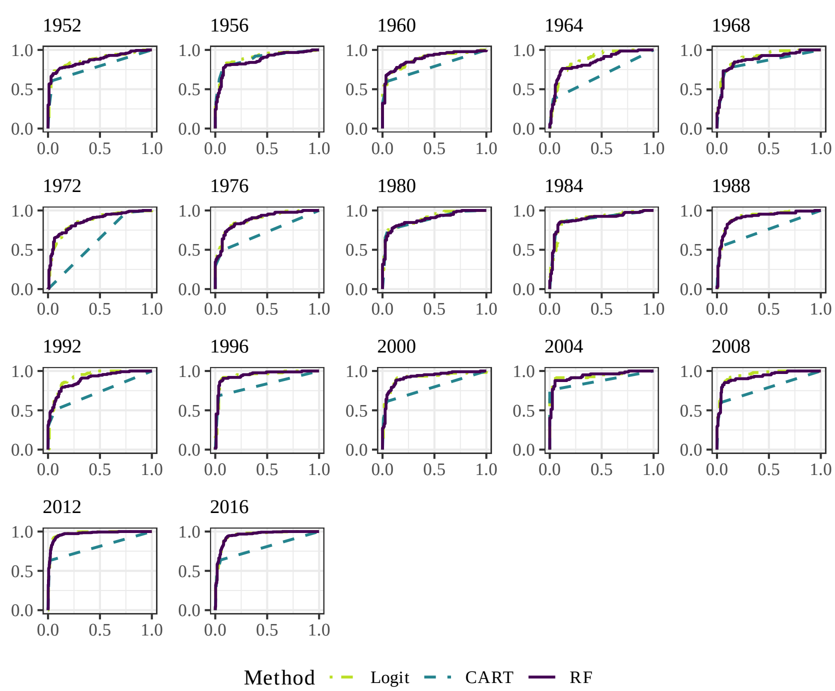


Figure A.2: ROC Curves for Presidential Vote Choice, Demo. + PID, Comparison Between Logit, CART, and RF

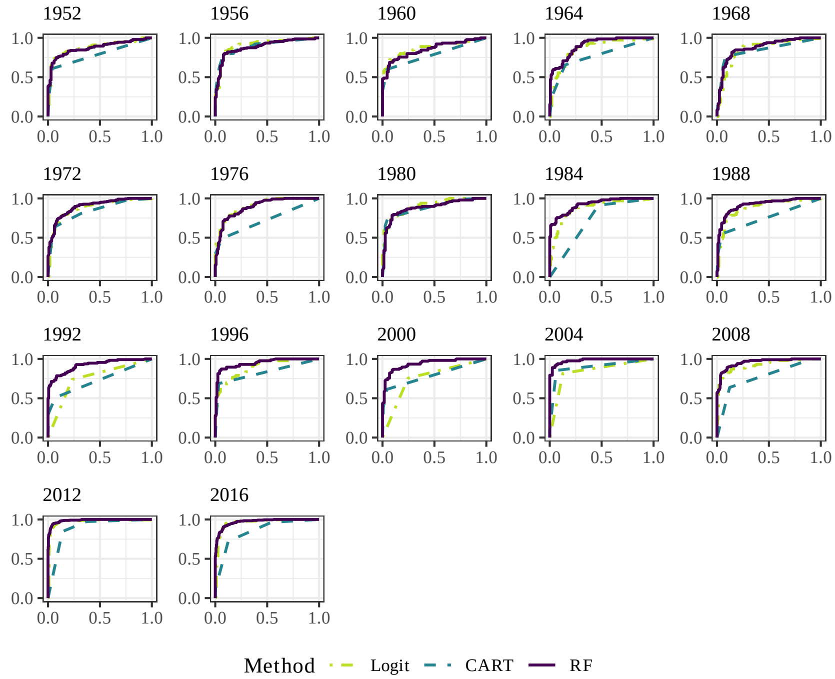


Figure A.3: ROC Curves for Presidential Vote Choice, Demo. + PID + Issues, Comparison Between Logit, CART, and RF

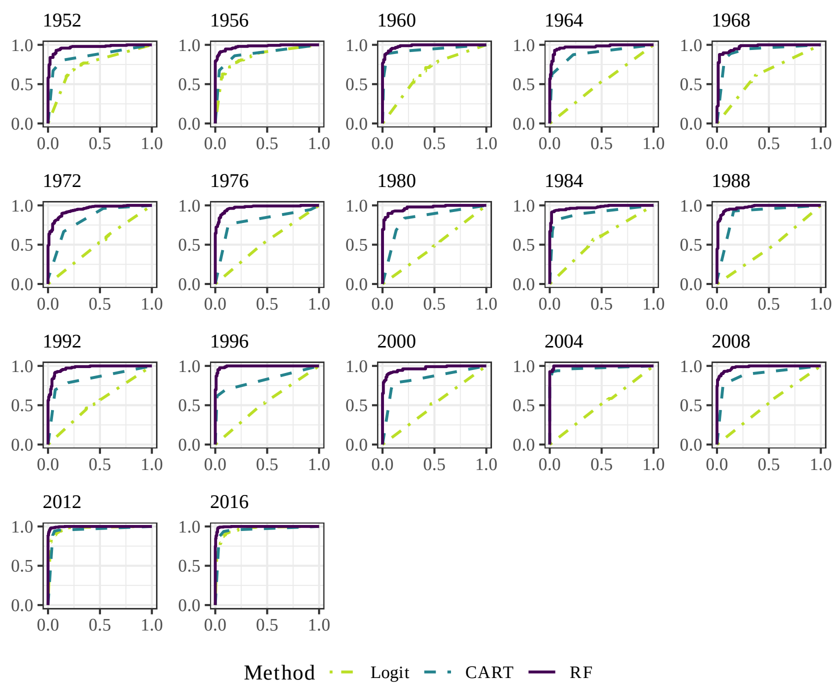


Figure A.4: ROC Curves for Presidential Vote Choice, All Covariates, Comparison Between Logit, CART, and RF

Note that logit's performance becomes very poor in the saturated model (more variables than number of observations) throughout years 1960–2008.

Appendix B Model Outputs and Performance: CCES

B.1 Accuracy Range Comparison Between Methods

Variable Specification	Year	Logit	CART	RF
Demographics Only	2018	[0.6515, 0.6721]	[0.6292, 0.6500]	[0.6535, 0.6740]
Demo. + PID	2018	[0.9259, 0.9369]	[0.8151, 0.8317]	[0.9223, 0.9336]
Demo. + PID + Issues	2018	[0.9562, 0.9647]	[0.9260, 0.9370]	[0.9533, 0.9621]
All Covariates	2018	[0.5212, 0.5429]	[0.9299, 0.9406]	[0.9608, 0.9688]
Demographics Only	2016	[0.6442, 0.6650]	[0.6222, 0.6432]	[0.6359, 0.6567]
Demo. + PID	2016	[0.8919, 0.9051]	[0.8579, 0.8728]	[0.8892, 0.9026]
Demo. + PID + Issues	2016	[0.9242, 0.9354]	[0.8218, 0.8382]	[0.9201, 0.9316]
All Covariates	2016	[0.9471, 0.9565]	[0.8875, 0.9009]	[0.9409, 0.9509]
Demographics Only	2014	[0.6382, 0.6591]	[0.6225, 0.6435]	[0.6353, 0.6563]
Demo. + PID	2014	[0.9048, 0.9173]	[0.8155, 0.8321]	[0.9019, 0.9146]
Demo. + PID + Issues	2014	[0.9188, 0.9304]	[0.8266, 0.8429]	[0.9160, 0.9278]
All Covariates	2014	[0.9436, 0.9534]	[0.9070, 0.9193]	[0.9412, 0.9511]
Demographics Only	2012	[0.6211, 0.6430]	[0.5851, 0.6073]	[0.6268, 0.6486]
Demo. + PID	2012	[0.9095, 0.9221]	[0.8271, 0.8439]	[0.9094, 0.9220]
Demo. + PID + Issues	2012	[0.9480, 0.9577]	[0.8956, 0.9091]	[0.9446, 0.9546]
All Covariates	2012	[0.9623, 0.9706]	[0.9163, 0.9285]	[0.9619, 0.9702]
Demographics Only	2010	[0.6332, 0.6527]	[0.6136, 0.6334]	[0.6404, 0.6598]
Demo. + PID	2010	[0.9134, 0.9246]	[0.8358, 0.8506]	[0.9109, 0.9222]
Demo. + PID + Issues	2010	[0.9403, 0.9497]	[0.9035, 0.9153]	[0.9369, 0.9465]
All Covariates	2010	[0.9497, 0.9583]	[0.9332, 0.9431]	[0.9511, 0.9596]
Demographics Only	2008	[0.6233, 0.6509]	[0.5953, 0.6234]	[0.6218, 0.6495]
Demo. + PID	2008	[0.8943, 0.9114]	[0.8087, 0.8308]	[0.8903, 0.9077]
Demo. + PID + Issues	2008	[0.9362, 0.9496]	[0.8743, 0.8928]	[0.9293, 0.9434]
All Covariates	2008	[0.9740, 0.9825]	[0.9607, 0.9713]	[0.9761, 0.9842]

Table B.5: Accuracy Range Comparison, Presidential Vote Choice, CCES 2008–2018

B.2 Full Performance Tables

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.7221	0.6638	[0.6535, 0.6740]	0.6492	0.6129	0.6305
Demo. + PID	2018	0.9699	0.9281	[0.9223, 0.9336]	0.9262	0.9197	0.9230
Demo. + PID + Issues	2018	0.9913	0.9578	[0.9533, 0.9621]	0.9561	0.9537	0.9549
All Covariates	2018	0.9932	0.9650	[0.9608, 0.9688]	0.9675	0.9573	0.9624
Demographics Only	2016	0.7141	0.6463	[0.6359, 0.6567]	0.6329	0.5452	0.5858
Demo. + PID	2016	0.9530	0.8960	[0.8892, 0.9026]	0.8846	0.8894	0.8870
Demo. + PID + Issues	2016	0.9781	0.9260	[0.9201, 0.9316]	0.9226	0.9155	0.9190
All Covariates	2016	0.9885	0.9461	[0.9409, 0.9509]	0.9404	0.9421	0.9413
Demographics Only	2014	0.7159	0.6459	[0.6353, 0.6563]	0.6354	0.4762	0.5444
Demo. + PID	2014	0.9659	0.9084	[0.9019, 0.9146]	0.8860	0.9110	0.8983
Demo. + PID + Issues	2014	0.9743	0.9221	[0.9160, 0.9278]	0.9194	0.9038	0.9115
All Covariates	2014	0.9898	0.9463	[0.9412, 0.9511]	0.9408	0.9382	0.9395
Demographics Only	2012	0.6914	0.6378	[0.6268, 0.6486]	0.6197	0.6953	0.6554
Demo. + PID	2012	0.9697	0.9158	[0.9094, 0.9220]	0.9087	0.9228	0.9157
Demo. + PID + Issues	2012	0.9892	0.9497	[0.9446, 0.9546]	0.9537	0.9443	0.9490
All Covariates	2012	0.9950	0.9662	[0.9619, 0.9702]	0.9664	0.9654	0.9659
Demographics Only	2010	0.7105	0.6502	[0.6404, 0.6598]	0.6245	0.7506	0.6818
Demo. + PID	2010	0.9669	0.9167	[0.9109, 0.9222]	0.9149	0.9186	0.9167
Demo. + PID + Issues	2010	0.9835	0.9418	[0.9369, 0.9465]	0.9402	0.9434	0.9418
All Covariates	2010	0.9887	0.9555	[0.9511, 0.9596]	0.9493	0.9623	0.9557
Demographics Only	2008	0.6806	0.6357	[0.6218, 0.6495]	0.6871	0.4663	0.5556
Demo. + PID	2008	0.9559	0.8993	[0.8903, 0.9077]	0.8863	0.9106	0.8983
Demo. + PID + Issues	2008	0.9836	0.9366	[0.9293, 0.9434]	0.9207	0.9522	0.9362
All Covariates	2008	0.9981	0.9805	[0.9761, 0.9842]	0.9775	0.9826	0.9801

Table B.6: Performance Metrics, Presidential Vote Choice, Random Forests, CCES 2008–2018

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.7190	0.6619	[0.6515, 0.6721]	0.6328	0.6613	0.6467
Demo. + PID	2018	0.9724	0.9315	[0.9259, 0.9369]	0.9252	0.9288	0.9270
Demo. + PID + Issues	2018	0.9923	0.9606	[0.9562, 0.9647]	0.9621	0.9534	0.9577
All Covariates	2018	0.5002	0.5321	[0.5212, 0.5429]	0.6000	0.0008	0.0016
Demographics Only	2016	0.7100	0.6546	[0.6442, 0.6650]	0.6185	0.6452	0.6315
Demo. + PID	2016	0.9549	0.8986	[0.8919, 0.9051]	0.8783	0.9043	0.8911
Demo. + PID + Issues	2016	0.9804	0.9299	[0.9242, 0.9354]	0.9246	0.9224	0.9235
All Covariates	2016	0.9901	0.9519	[0.9471, 0.9565]	0.9465	0.9488	0.9477
Demographics Only	2014	0.7129	0.6487	[0.6382, 0.6591]	0.5983	0.6369	0.6170
Demo. + PID	2014	0.9679	0.9112	[0.9048, 0.9173]	0.8824	0.9232	0.9023
Demo. + PID + Issues	2014	0.9760	0.9248	[0.9188, 0.9304]	0.9138	0.9171	0.9155
All Covariates	2014	0.9885	0.9487	[0.9436, 0.9534]	0.9418	0.9426	0.9422
Demographics Only	2012	0.6876	0.6321	[0.6211, 0.6430]	0.6082	0.7228	0.6606
Demo. + PID	2012	0.9732	0.9160	[0.9095, 0.9221]	0.9051	0.9276	0.9162
Demo. + PID + Issues	2012	0.9910	0.9530	[0.9480, 0.9577]	0.9545	0.9505	0.9525
All Covariates	2012	0.9952	0.9666	[0.9623, 0.9706]	0.9669	0.9656	0.9663
Demographics Only	2010	0.7076	0.6430	[0.6332, 0.6527]	0.6147	0.7633	0.6810
Demo. + PID	2010	0.9702	0.9191	[0.9134, 0.9246]	0.9084	0.9321	0.9201
Demo. + PID + Issues	2010	0.9857	0.9451	[0.9403, 0.9497]	0.9427	0.9477	0.9452
All Covariates	2010	0.9877	0.9541	[0.9497, 0.9583]	0.9493	0.9593	0.9543
Demographics Only	2008	0.6794	0.6372	[0.6233, 0.6509]	0.6616	0.5263	0.5862
Demo. + PID	2008	0.9601	0.9031	[0.8943, 0.9114]	0.8898	0.9149	0.9022
Demo. + PID + Issues	2008	0.9859	0.9432	[0.9362, 0.9496]	0.9350	0.9496	0.9423
All Covariates	2008	0.9786	0.9786	[0.9740, 0.9825]	0.9787	0.9774	0.9781

Table B.7: Performance Metrics, Presidential Vote Choice, Logit, CCES 2008–2018

Variable Specification	Year	AUC	Accuracy	CI	Precision	Recall	F1
Demographics Only	2018	0.6624	0.6397	[0.6292, 0.6500]	0.6067	0.6543	0.6296
Demo. + PID	2018	0.8136	0.8235	[0.8151, 0.8317]	0.9740	0.6401	0.7725
Demo. + PID + Issues	2018	0.9426	0.9316	[0.9260, 0.9370]	0.9220	0.9329	0.9274
All Covariates	2018	0.9402	0.9354	[0.9299, 0.9406]	0.9313	0.9306	0.9310
Demographics Only	2016	0.6598	0.6328	[0.6222, 0.6432]	0.5815	0.7113	0.6399
Demo. + PID	2016	0.8640	0.8655	[0.8579, 0.8728]	0.9367	0.7579	0.8379
Demo. + PID + Issues	2016	0.8265	0.8301	[0.8218, 0.8382]	0.8702	0.7401	0.7999
All Covariates	2016	0.9008	0.8943	[0.8875, 0.9009]	0.8731	0.9006	0.8866
Demographics Only	2014	0.6800	0.6331	[0.6225, 0.6435]	0.6059	0.4981	0.5467
Demo. + PID	2014	0.8063	0.8239	[0.8155, 0.8321]	0.9603	0.6297	0.7606
Demo. + PID + Issues	2014	0.8373	0.8349	[0.8266, 0.8429]	0.8045	0.8301	0.8171
All Covariates	2014	0.9216	0.9133	[0.9070, 0.9193]	0.8980	0.9080	0.9030
Demographics Only	2012	0.6083	0.5962	[0.5851, 0.6073]	0.5600	0.8628	0.6792
Demo. + PID	2012	0.8358	0.8356	[0.8271, 0.8439]	0.9794	0.6826	0.8045
Demo. + PID + Issues	2012	0.9028	0.9025	[0.8956, 0.9091]	0.9157	0.8847	0.8999
All Covariates	2012	0.9322	0.9226	[0.9163, 0.9285]	0.8994	0.9499	0.9240
Demographics Only	2010	0.6318	0.6235	[0.6136, 0.6334]	0.5790	0.9015	0.7051
Demo. + PID	2010	0.8463	0.8433	[0.8358, 0.8506]	0.9654	0.7117	0.8193
Demo. + PID + Issues	2010	0.9090	0.9095	[0.9035, 0.9153]	0.8921	0.9314	0.9113
All Covariates	2010	0.9470	0.9383	[0.9332, 0.9431]	0.9273	0.9509	0.9390
Demographics Only	2008	0.6108	0.6094	[0.5953, 0.6234]	0.7094	0.3391	0.4589
Demo. + PID	2008	0.8257	0.8200	[0.8087, 0.8308]	0.7382	0.9783	0.8415
Demo. + PID + Issues	2008	0.8854	0.8838	[0.8743, 0.8928]	0.8560	0.9162	0.8851
All Covariates	2008	0.9676	0.9663	[0.9607, 0.9713]	0.9673	0.9635	0.9654

Table B.8: Performance Metrics, Presidential Vote Choice, CART, CCES 2008–2018

B.3 ROC Curves

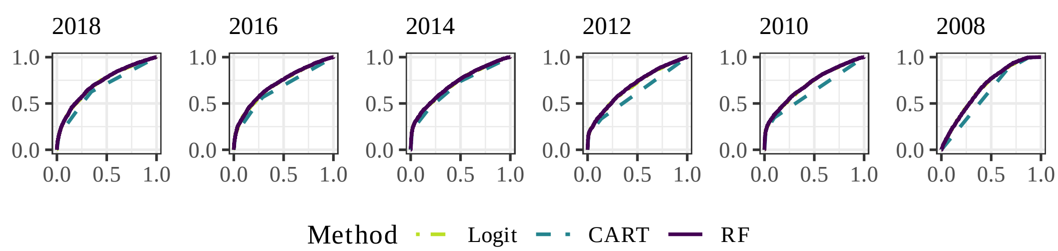


Figure B.5: ROC Curves for Presidential Vote Choice, Demographics Only, Comparison Between Logit, CART, and RF

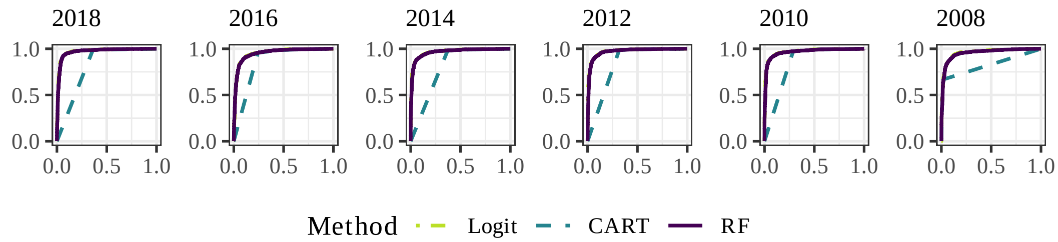


Figure B.6: ROC Curves for Presidential Vote Choice, Demo. + PID, Comparison Between Logit, CART, and RF

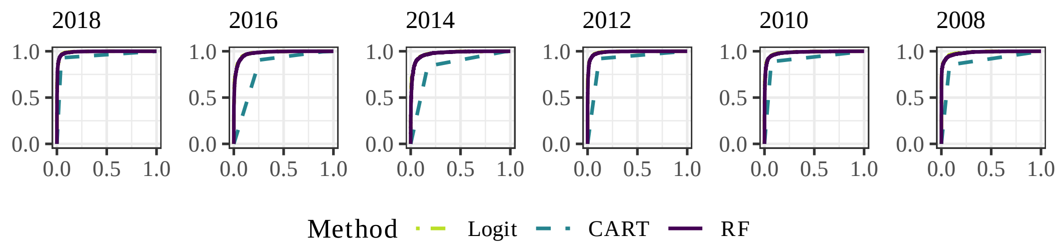


Figure B.7: ROC Curves for Presidential Vote Choice, Demo. + PID + Issues, Comparison Between Logit, CART, and RF

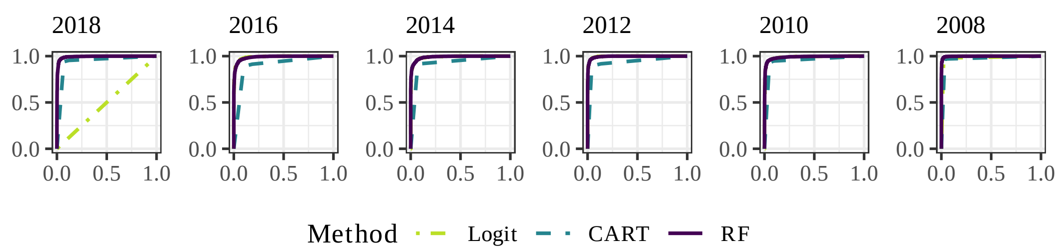


Figure B.8: ROC Curves for Presidential Vote Choice, All Covariates, Comparison Between Logit, CART, and RF

		Accuracy					
Model and Specification	AUC	Point estimate	Low est.	High est.	Precision	Recall	F1
RF: Demographics Only	0.712	0.655	0.642	0.668	0.691	0.682	0.686
RF: Demo. + PID	0.945	0.888	0.879	0.896	0.891	0.908	0.899
RF: Demo. + PID + Issues	0.969	0.906	0.898	0.914	0.903	0.930	0.917
RF: All Covariates	0.975	0.917	0.909	0.925	0.911	0.942	0.926
CART: Demographics Only	0.695	0.652	0.638	0.665	0.694	0.661	0.677
CART: Demo. + PID	0.932	0.887	0.878	0.895	0.889	0.908	0.899
CART: Demo. + PID + Issues	0.940	0.906	0.898	0.914	0.906	0.926	0.916
CART: All Covariates	0.941	0.909	0.901	0.917	0.905	0.934	0.919
Logit: Demographics Only	0.710	0.650	0.637	0.663	0.676	0.705	0.690
Logit: Demo. + PID	0.941	0.886	0.877	0.895	0.892	0.904	0.898
Logit: Demo. + PID + Issues	0.969	0.909	0.901	0.917	0.906	0.932	0.919
Logit: All Covariates	0.975	0.919	0.912	0.927	0.919	0.937	0.928

Table C.9: Accuracy Range and Other Performance Metrics, Presidential Vote Intent (Nationscape, 2019-20), Random Forest, Logit, and CART

Appendix C Nationscape

C.1 Full Performance Tables

C.2 ROC Curves

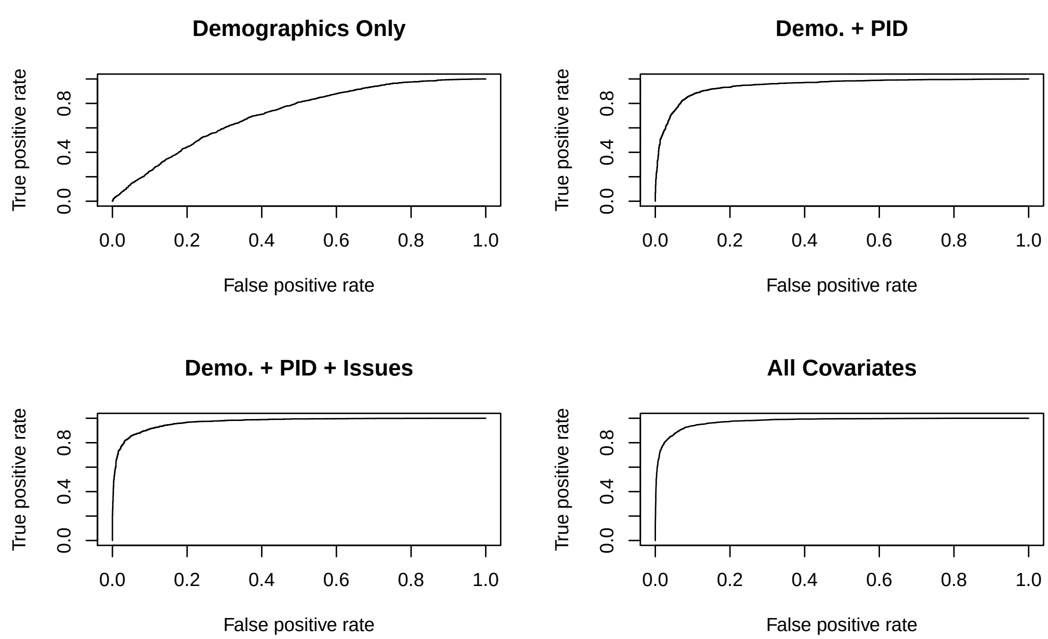


Figure C.9: ROC Curves for all specifications, Nationscape, Random Forests

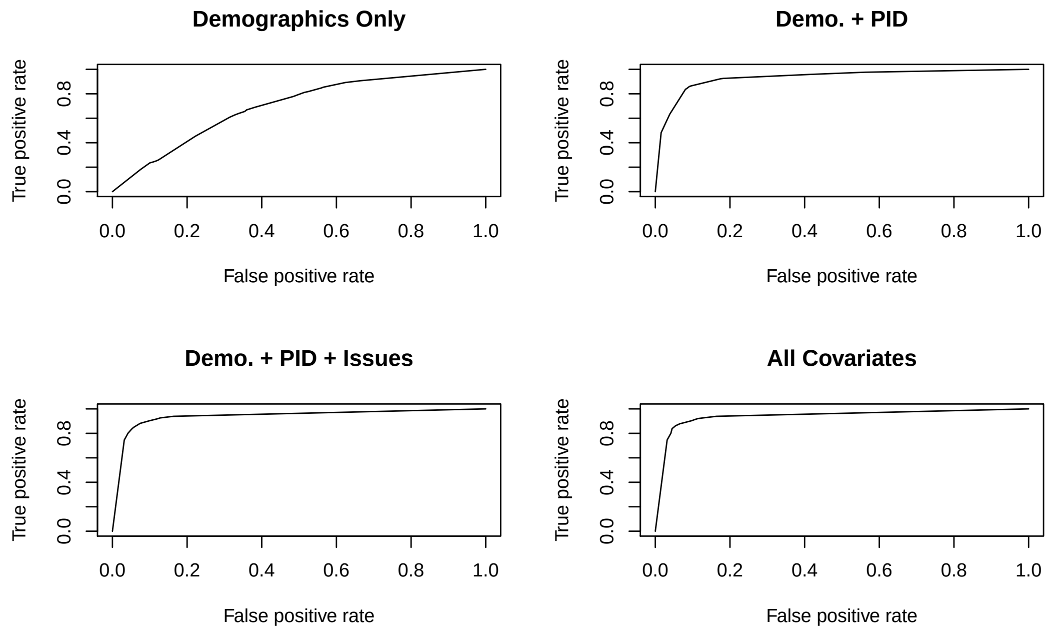


Figure C.10: ROC Curves for all specifications, Nationscape, CART

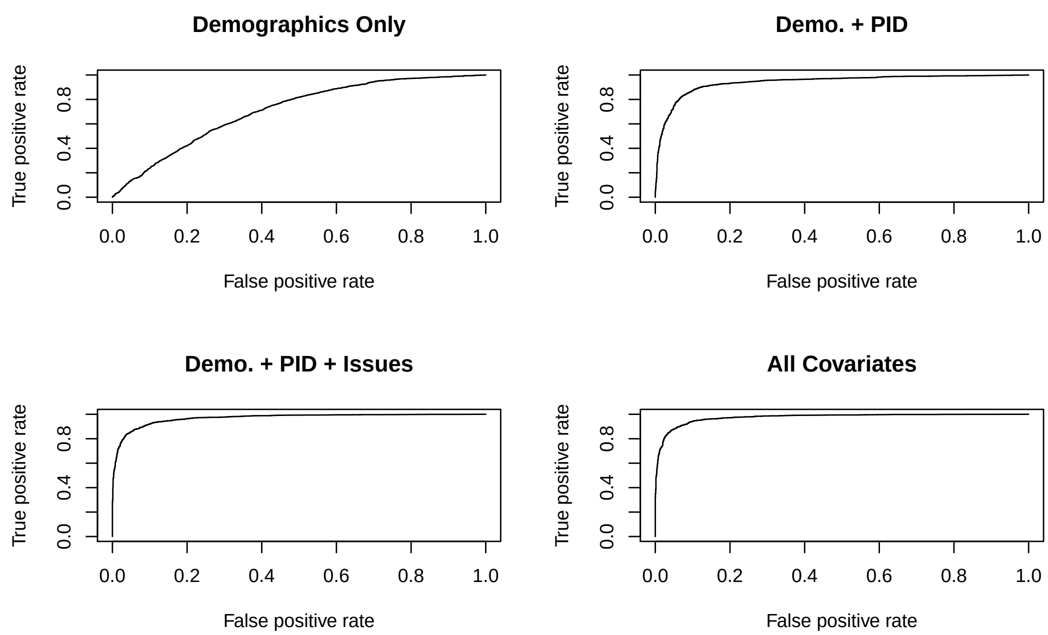


Figure C.11: ROC Curves for all specifications, Nationscape, Logit

Appendix D Alternative Specifications for Demographics Only

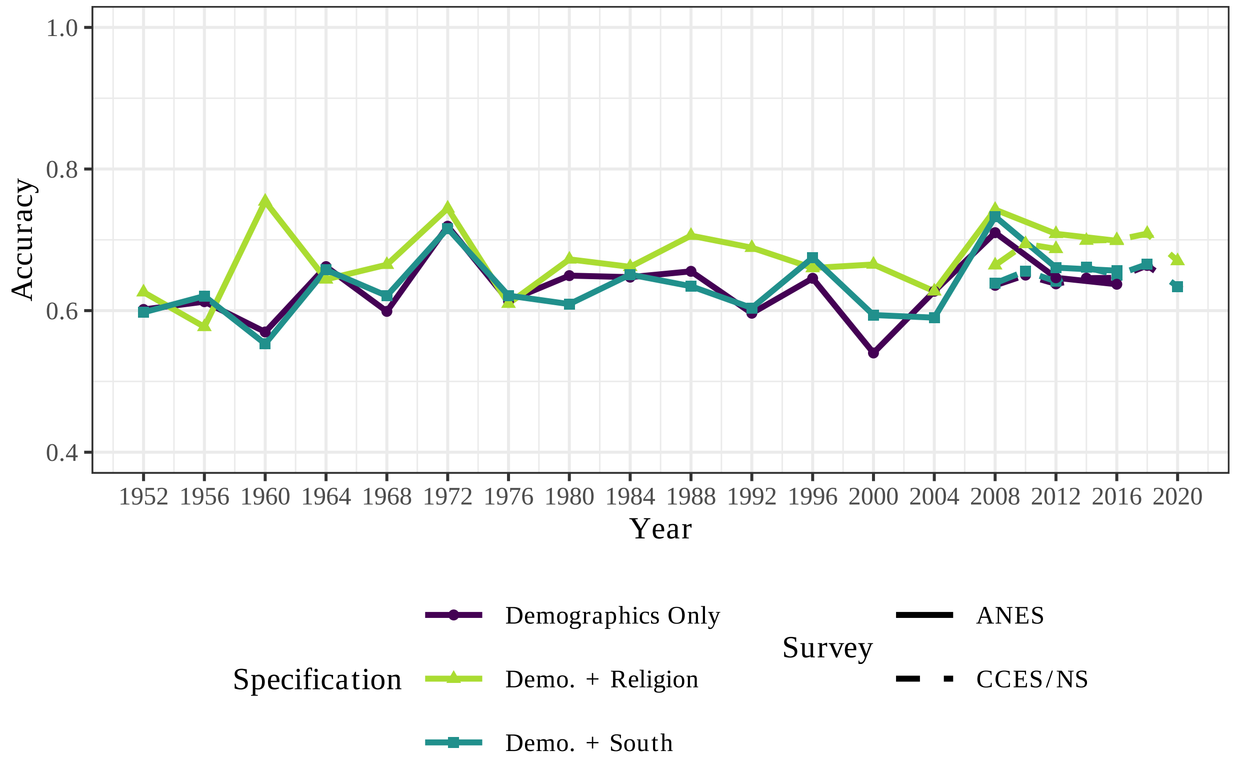


Figure D.12: Performance of Presidential Vote Prediction Over Time.
Comparison of 5-variable Demographics, Demographics + South, Demographics + Religion, All Surveys. The prediction algorithm employed is Random Forests.

Early ANES years do not contain information on respondents' states nor on their specific religion (e.g., Evangelical Christian). Therefore, to see the addition of geography, we added South vs. non-South variable (the 11-state definition). To see the addition of religion, we added a four-level categorical variable of (1) Protestant, (2) Catholic, (3) Jewish, and (4) all others. None of the slopes in Figure D.12 are, when regressed over time, statistically significant.

Appendix E Benchmarking Accuracy Rates

Election Year	Accuracy achieved by a naive guess (i.e., a classification rule where the popular vote winner is predicted for each respondent)
2020	52.27
2016	51.11
2012	51.97
2008	53.69
2004	51.25
2000	50.26
1996	54.73
1992	53.46
1988	53.90
1984	59.17
1980	55.31
1976	51.06
1972	61.79
1968	50.41
1964	61.34
1960	50.09
1956	57.75
1952	55.45
Average	54.17

Table E.10: Two-party vote share of the winner in U.S. presidential elections (1952-2020)