

Measuring Swing Voters with a Supervised Machine Learning Ensemble

Christopher Hare¹ and Mikayla Kutsuris²

¹Department of Political Science, University of California, Davis, CA, USA. Email: cdhare@ucdavis.edu

²Desmond, Nolan, Livaich & Cunningham, Sacramento, CA, USA. Email: mkutsuris@dnlc.net

Abstract

Theory has long suggested that swing voting is a response to cross-pressures arising from a mix of individual attributes and contextual factors. Unfortunately, existing regression-based approaches are ill-suited to explore the complex combinations of demographic, policy, and political factors that produce swing voters in American elections. This gap between theory and practice motivates our use of an ensemble of supervised machine learning methods to predict swing voters in the 2012, 2016, and 2020 U.S. presidential elections. The results from the learning ensemble substantiate the existence of swing voters in contemporary American elections. Specifically, we demonstrate that the learning ensemble produces well-calibrated and externally valid predictions of swing voter propensity in later elections and for related behaviors such as split-ticket voting. Although interpreting black-box models is more challenging, they can nonetheless provide meaningful substantive insights meriting further exploration. Here, we use flexible model-agnostic tools to perturb the ensemble and demonstrate that cross-pressures (particularly those involving ideological and policy-related considerations) are essential to accurately predict swing voters.

Keywords: voting behavior, swing voters, supervised machine learning, ensemble models

If you want to pick cherries, go where the cherries is.

Matthew Reese, Executive Director of Citizens for Kennedy

1 Introduction

Scholars and pundits alike have long been captivated by “swing voters”: a niche group of voters who are undecided or otherwise could be persuaded to switch their support between candidates or parties. Pollsters routinely tag these voters using a colorful taxonomy: soccer moms, Reagan Democrats, and so forth. Likewise, research into the political behavior of swing voters has spanned the entire course of modern political science, beginning with the pioneering voter panel studies of the Columbia school (Lazarsfeld, Berelson, and Gaudet 1944). In short, both campaigns and political scientists have a keen interest in identifying the “cherries” of the electorate.

Although they have long captured the political imagination, swing voters have remained an enigmatic player in American politics. At times, swing voters are idealized as model citizens—a thoughtful, fair, open-minded, and dispassionate force for reasoned discourse and moderation in American democracy. However, one can also point to a legion of findings documenting depressed levels of political information and interest among swing voters. Recent work also emphasizes the role of partisan polarization in clarifying the major party brands and reducing the number of swing voters in contemporary American elections (e.g., Gelman *et al.* 2016; Smidt 2017).¹

¹ However, greater clarity in policy differences between the parties may also serve to highlight salient cross-pressures that would otherwise go unnoticed (Box-Steffensmeier *et al.* 2015).

Political Analysis (2023)
vol. 31: 537–553
DOI: [10.1017/pan.2022.24](https://doi.org/10.1017/pan.2022.24)

Published
17 October 2022

Corresponding author
Christopher Hare

Edited by
Jeff Gill

© The Author(s), 2022. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Before we can learn about swing voters, we must first be able to identify them. More than 70 years of theory has established cross-pressuring as a key motivator of swing voting (Berelson, Lazarsfeld, and McPhee 1954; Campbell *et al.* 1960; Key Jr. 1966). These cross-pressures are complex entities that involve friction between partisan, demographic, social, and attitudinal elements. Recent work demonstrates that cross-pressures remain pervasive among voters, even in the face of partisan polarization (e.g., Brader, Tucker, and Theriault 2014; Hillygus and Shields 2008).

In voting behavior models, these cross-pressures are manifest as (often high-order) interaction terms that are difficult to detect using standard regression-based approaches. Instead, we argue that an ensemble machine learning approach—one that combines insights from a variety of supervised methods—provides a more flexible and theoretically superior means of modeling the complex mechanisms driving swing voter behavior.

We test the effectiveness of a learning ensemble in predicting swing voter propensity using longitudinal survey data between 2011 and 2019. Our analysis reveals two main findings: first, the ensemble model identifies an individual-level disposition toward swing voting that is predictive (1) for out-of-sample respondents, (2) across related behaviors such as split-ticket voting, and (3) of swing voting in later elections. We attribute the learning ensemble's success to its ability to identify consequential cross-pressures among voters. In short, swing voters exist, and this propensity can be measured with an appropriate modeling strategy.

Second, we find that policy-related considerations (specifically, measures of ideological identification, issue preferences, and issue salience) are essential for accurately predicting swing voters. The results indicate that policy attitudes are a key component of swing voting behavior, even if their influence often cannot be characterized in simple linear terms. Swing voting arises not only among voters who lack partisan attachments or political sophistication, but also those who experience conflict between the parties on salient policy issues (see also Green 2020).

2 Swing Voters and Cross-Pressures

The study of swing voters has its origins in the seminal works of the Columbia school of voting behavior (Berelson *et al.* 1954; Lazarsfeld *et al.* 1944).² One pioneering aspect of these studies concerns their discussion of “cross-pressures”: competing factors or considerations that introduce ambivalence into voters' decisionmaking processes. Voters facing cross-pressures exhibit greater volatility their vote intentions: they are more likely to change their minds and remain undecided over the course of the campaign. Consequently, cross-pressured voters' final decisions often turn on short-term factors or personal salience—which candidate or party aligns with the set of considerations the voter considers most important (Petrocik, Benoit, and Hansen 2003; RePass 1971; Zaller 2004).

More recent work also indicates that demographically (Brader *et al.* 2014) and ideologically (Treier and Hillygus 2009) cross-pressured voters hold more ambivalent political preferences. Such cross-pressures are also widespread in the American electorate: about two-thirds of partisans disagree with their party on at least one issue they consider important, with 40% of partisans disagreeing on two or more personally salient issues (Hillygus and Shields 2008; see also Lupton, Myers, and Thornton 2015). A sizable proportion of American voters also hold incongruent symbolic and operational ideological preferences (Ellis and Stimson 2012). The diversity of issue publics in the American electorate provides yet another source of tension in the multidimensional lattice of voters' policy attitudes (Klar 2014; Krosnick 1990).

² Like Hill (2017), we use the term “swing voter” to refer to regular voters whose preferences oscillate between elections (Converse 1962; Key Jr. 1966) rather than voters whose turnout decisions produce compositional changes in the electorate (e.g., “surge and decline” theory; Campbell 1960). In Appendix C in the Supplementary Material, we show that the learning ensemble produces highly correlated predictions across a range of swing voter operationalizations.

A number of factors—most importantly, party identification and political sophistication—serve to inoculate voters from the effects of cross-pressures and promote stability in voting behavior. Indeed, much of the literature on swing or “floating” voters focuses on these voters’ lack of partisan attachments and political knowledge to explain their susceptibility to cross-pressuring (Converse 1962; Mayer 2007). However, these factors also interact with each other and the political environment in complex ways to influence swing voter behavior. For instance, intensely polarized campaigns can harden partisan identities, but can also serve to inform or remind cross-pressured partisans of inconsistencies in their political choices (Box-Steffensmeier *et al.* 2015). In fact, campaign efforts to prime specific considerations are most effective among voters with moderate and especially high levels of political awareness in the United States (Claassen 2011) and comparative contexts (Dassonneville and Dejaeghere 2014; Weghorst and Lindberg 2013). Accordingly, although cross-pressured partisans tend to hold lower levels of political information (Claassen and Highton 2009), cross-pressures are likely most consequential for knowledgeable partisans—especially when the issue is salient (Carsey and Layman 2006).

Determining the extent to which policy considerations underlie swing voter propensity has meaningful normative implications. If swing voters are politically alienated and lack meaningful policy views, then they may simply be “electoral drifters or rolling stones, who cast their vote almost at random” (Katz, Rattinger, and Pedersen 1997, 87). On the other hand, policy preferences may systematically influence swing voter propensity, but do so in nuanced (interactive) ways whose detection require a more sophisticated modeling approach. If so, swing voters may be more pervasive and receptive to policy-based appeals than commonly believed.

3 A Learning Ensemble Strategy to Identify Swing Voters

Existing work demonstrates that swing voters are a heterogeneous group whose choices are motivated by complex interactions between a set of factors—namely, demographic and attitudinal cross-pressures, issue salience, political sophistication, strength of partisan ties, cross-cutting social affiliations and identities, and ideological moderation. The forces driving swing voter propensity are likely to take the form of complex nonlinear and conditional patterns—patterns that could be easily overlooked by researchers employing traditional regression models. In such cases, relaxing functional form assumptions can enhance inference and reveal new possibilities for investigation (Wager and Athey 2018).

This consideration motivates our use of an alternative, flexible supervised machine learning-based estimation strategy to detect the pairings of attributes and attitudes that theory suggests underlie swing voter behavior. These methods allow us to test the possibility that nonlinear and interactive relationships predict swing voting. Social scientists are more frequently adopting machine learning to discover patterns and identify meaningful relationships in data—especially those involving complex patterns in the data generating process.³

Below we detail our use of a learning ensemble—one that combines insights from multiple supervised machine learning methods—to study swing voter behavior. Ensemble methods have revolutionized the fields of predictive modeling and statistical learning over recent decades, advanced through theoretical work (Schapire 1990) alongside the development of specific procedures such as stacking (Wolpert 1996), bagging (Breiman 1996), boosting (Freund and Schapire 1997), and super learners (van der Laan, Polley, and Hubbard 2007). Ensembling is based on a key insight sometimes referred to as the “wisdom of crowds”: a diverse set of component methods, each with their own assumptions about the underlying data generating process, can capture different aspects of the relationship between predictor and response variables (Ali and Pazzani 1996). Aggregating their results into an ensemble smooths over idiosyncratic errors produced by

3 See, for example, Grimmer, Roberts, and Stewart (2021) and Gill (2021).

the component methods (or base learners), producing a lower variance estimator. Hence, the goal of ensemble methods is to find the combination of base learners that optimizes generalizable (i.e., out-of-sample) predictive performance.

To accomplish this, we follow the approach developed in Grimmer, Messing, and Westwood (2017) to estimate the component weights for a series of supervised machine learning methods trained to predict swing voting. Formally, let i index the observations ($i = 1, \dots, n$) and m index the component methods ($m = 1, \dots, M$). We first randomly divide the N observations into D ($d = 1, \dots, D$) folds, and generate predictions for observations in each fold d by fitting the M methods on observations in the remaining $D - 1$ folds. This produces an $N \times M$ matrix of out-of-sample predictions (\hat{Y}_{im}) for each observation across the component methods. We then estimate the component weights (w_m) by fitting the constrained regression problem:

$$Y_i = \sum_{m=1}^M w_m \hat{Y}_{im} + \epsilon_i, \quad (1)$$

where Y_i is the observed response by respondent i and ϵ_i is a stochastic error term, with the constraints that $\sum_{m=1}^M w_m = 1$ and $w_m \geq 0$. Finally, we fit each of the component methods to the complete dataset and weight the predictions using the M -length vector of estimated component weights $\hat{\mathbf{w}}$. The weighted combination of predictions comprise the ensemble estimates.

In our analysis, we use an ensemble of eight component methods (or base learners) to model the relationship between swing voting and a series of predictor variables. These include a standard additive probit regression model, k -nearest neighbors, group-lasso interaction network, support vector machine (SVM), neural networks, random forests, gradient-boosted decision trees, and extremely randomized trees.⁴ We describe the component methods and weights in greater detail in Appendix B in the Supplementary Material, but it is worth emphasizing that these procedures range in complexity and each bring strengths to the ensemble. For instance, SVMs and neural networks excel in fitting nonlinear mappings between X and Y by estimating various transformations of the input data whereas the group-lasso interaction network and tree-based methods are robust to irrelevant predictors (Hastie, Tibshirani, and Friedman 2009, 350–352; Kuhn and Johnson 2013, 488–490, 550).

Given our theoretical expectations about the role of cross-pressures on swing voting, it is also important to emphasize that the tree-based methods provide a natural way to model interactive relationships, since the effect of any given variable is considered conditionally based on previous splits in the tree. This helps to explain why decision trees are so effective in uncovering interactive effects and modeling the behavior of heterogeneous groups like swing voters (Lampa *et al.* 2014; Montgomery and Olivella 2018). The ensemble harnesses the strengths of each base learner in a single prediction function, one that can be validated and probed as described below.

4 Predicting Swing Voters in Presidential Elections

We train our component methods and construct our learning ensemble using data from the 2012 Cooperative Campaign Analysis Project (CCAP; Jackman *et al.* 2012).⁵ The 2012 CCAP interviewed a group of 43,998 respondents with a rolling cross-section design at three points during the 2012 campaign: a baseline survey in December 2011, a weekly interview between January and November 2012, and a post-election survey. Beyond its large sample size, the 2012 CCAP is ideal for our purposes, because it allows us to fit the learning ensemble to only the variables collected during the baseline survey, then evaluate the model's performance in predicting subsequent

4 We use the *caret* package in **R** to tune and estimate each model (Kuhn and Johnson 2013).

5 Replication materials are available at Hare and Kutsuris (2022).

behavior during and immediately following the 2012 election. Four-thousand seven-hundred and fifteen of these respondents are also panelists in a long-term survey conducted by the Democracy Fund (the VOTER Study Group; Democracy Fund Voter Study Group 2020), allowing us to validate the model using vote choice/intention measures from the 2016 and 2020 presidential elections.

The 2012 CCAP asked a wide set of political questions concerning past turnout and vote choice; candidate evaluations and ideological and party identification; issue preferences; political knowledge and interest; racial resentment; economic evaluations; and personal issue salience. We also construct a measure of ideological inconsistency based on Federico and Hunt's (2013) operationalization. In all, we include 45 predictor variables (detailed in Appendix A in the Supplementary Material) that theory suggests are related to swing voting and were included in the baseline wave of the 2012 CCAP.

To construct our measure of swing voting, we look for manifest behaviors that serve as indicators of a voter's latent propensity to switch between electoral choices. As Mayer (2008, 2) defines the concept: "In simple terms, a swing voter is, as the name implies, a voter who could go either way: a voter who is not so solidly committed to one candidate or the other as to make all efforts at persuasion futile." Following Weghorst and Lindberg (2013), we include two commonly used markers of swing voting: party switching and indecision (i.e., undecided voters).⁶ In line with Mayer's (2008) definition, both behaviors indicate that a voter's preferences meet a meaningful threshold of volatility. By providing the learning ensemble multiple instances of the concept under investigation, it has greater leverage to uncover the complex patterns underlying swing voter propensity.⁷

Specifically, our response variable codes swing voters as those respondents who fully or temporarily defected from their 2008 presidential vote choice in the following ways:

1. **Full defectors:** respondents who voted for Obama (McCain) in the 2008 presidential election, but switched to the other party and voted for Romney (Obama) in 2012. This corresponds to the standard operationalization of floating voters.
2. **Temporary defectors:** respondents who voted for the same party in the 2008 and 2012 presidential elections, but stated that they were undecided or intended to switch to the other party's candidate at some point during the 2012 campaign.

This operationalization produces 2,294 respondents whom we classify as swing voters, compared to the 14,881 respondents who expressed party-consistent preferences across the 2008 and 2012 presidential elections coded as non-swing voters. Of these respondents, 12,914 have complete, non-missing profiles across the predictor and response variables. We select a random 65% (8,395) of these respondents to create the training set, leaving 4,519 respondents to comprise the validation set. The training set is used to estimate the learning ensemble, whereas the validation set is used to assess the model's out-of-sample performance.

We estimate and tune the component models using 10-fold cross-validation (repeated five times) using area under the receiver operating characteristic curve (AUC-ROC) as our performance metric. The learning ensemble selects four methods (with component weights in parentheses): extremely randomized trees ($w = 0.32$), gradient-boosted trees ($w = 0.27$), group-lasso interaction network ($w = 0.23$), and the SVM classifier ($w = 0.17$). Hence, the ensemble scores are weighted sums of the predictions from these methods. In the next section, we assess the generalizability of

- 6 Although reporting oneself as "undecided" may overstate a voter's level of indecision, past work demonstrates that there are meaningful differences between early and late deciders in elections according to this measure. Namely, the eventual vote choices of late deciders (those who report they are undecided until the end of the campaign) are less predictable and more heavily influenced by short-term factors (Box-Steffensmeier *et al.* 2015). We also limit the analysis to CCAP respondents who completed the mid-election wave after the end of April 2012, when Mitt Romney became the presumptive Republican presidential nominee.
- 7 We test alternate operationalizations of the response variable and find that each set of out-of-sample predictions is highly correlated ($r \geq 0.9$) with our results. Additional details are provided in Appendix C in the Supplementary Material.

the learning ensemble, specifically how well it predicts swing voting among different respondents and across different electoral contexts. This is of course key to the quality of any substantive inferences that can be drawn from the model concerning general patterns of swing voting in the electorate.

4.1 Validating the Swing Voter Scores

In this section, we exclusively use the 4,519 observations in the validation set to assess the out-of-sample predictive accuracy of the learning ensemble. We also compare the ensemble's performance to a baseline model, using a generalized additive model (GAM) using splines to generate a second set of swing voter predictions.⁸ We are interested in how well each model predicts swing voting as defined by our measure *and* as evidenced in related behaviors such as split-ticket voting.

As detailed in Table 1, we use a variety of fit metrics to evaluate classifier performance. Given the large class imbalance (only 13% of the respondents in both the training and validation sets are classified as swing voters), we focus on metrics other than simple accuracy that measure different desirable qualities of a prediction model. While AUC-ROC provides an indicator of a model's *discriminatory power* (how well it separates predictions for observations in each class), Brier skill scores (BSSs) are used to assess *calibration*: the consistency between the predicted and observed class proportions. Both discrimination and calibration should be considered when evaluating the quality of predictive models (e.g., Bansak 2019). The final two metrics—the Matthews correlation coefficient (MCC) and cross-entropy (CE) loss—both describe the *strength of association* between the model predictions (the swing voter propensity scores) and the response variable.

The performance metrics for our main measure of swing voting are presented in Table 2. Alongside the learning ensemble and baseline GAM, we also include the individual base learners that comprise the ensemble. Across the four key aggregate metrics (AUC-ROC, BSS, MCC, and CE loss), Table 2 shows that the ensemble outperforms both the GAM and its component methods. The learning ensemble especially outperforms the baseline GAM in terms of its ability to detect the swing voters in the sample (sensitivity rates of 0.57 vs. 0.22). While two of the base learners (ERT and GBM) have greater sensitivity than the ensemble, they also tend to overpredict swing voting and achieve lower specificity rates. The learning ensemble balances both goals, creating a more discriminatory *and* better calibrated set of swing voter propensity scores.⁹

This smoothing effect is also seen in Figure 1, which isolates the ROC and calibration curve plots for the learning ensemble and GAM. Two calibration plots are shown: the top panel uses the raw predictions from each model, whereas the bottom panel uses Platt scaling to calibrate the predictions with a logistic transformation (Platt 2000).¹⁰ Focusing first on the ROC curves, we find the largest gap between the learning ensemble and GAM occurs in the high-specificity (low false positive rate) region. The GAM is committing many Type II (false-negative) errors, overly classifying respondents as non-swing voters because it misses patterns that are predictive of swing voting.

The calibration plots in Figure 1b in the respondents based on their swing voter propensity scores and show the percentage of observed swing voters within each bin. The bin midpoints are plotted on the x-axis and the observed swing voter proportions on the y-axis. A perfectly

8 We estimate the GAM model by regressing the swing voter measure onto the same set of predictor variables for the same set of training observations as used to estimate the learning ensemble. Additional details are provided in Appendix B in the Supplementary Material.

9 Appendix I in the Supplementary Material provides a sensitivity analysis of these results, showing that the performance gap between the learning ensemble and GAM persists (particularly according to the AUC-ROC, BSS, and MCC fit metrics) even when the predictions are severely perturbed with random noise.

10 Platt scaling transforms the model predictions \hat{p} (or more generally, model output such as $\text{logit}(\hat{p})$) using the inverse logit function $(1 + \exp(-\beta_0 - \beta_1 \text{logit}(\hat{p})))^{-1}$, where β is estimated by regressing the swing voter response variable on $\text{logit}(\hat{p})$. Because this is a monotonic transformation, it does not affect the AUC-ROC statistic.

Table 1. Description and interpretation of model predictive performance metrics.

Metric	Description	Range	Notes
Accuracy	Proportion of correctly classified cases.	[0,1] Higher values preferred.	
Sensitivity	Proportion of positive cases (swing voters) correctly classified.	[0,1] Higher values preferred.	Also referred to as the true positive rate (TPR).
Specificity	Proportion of negative cases (non-swing voters) correctly classified.	[0,1] Higher values preferred.	Also referred to as the true negative rate (TNR).
Area under the receiver operating characteristic curve (AUC-ROC)	Measures classifier discriminatory power, indicating how well a model balances the trade-off between sensitivity and specificity across the range of possible classification thresholds. ROC curves trace the pairs of sensitivity and (reversed) specificity values across thresholds; integrating the ROC curve produces the AUC-ROC.	[0,1] A value of 0.5 indicates no improvement over a null classifier.	Equivalent to the Mann–Whitney U statistic/Wilcoxon rank-sum test. Represents the median difference in predictions between observations in each class.
Brier skill score (BSS)	Brier scores measure how well a set of predictions is calibrated with the observed outcome. In a well-calibrated model, $x\%$ of observations with a predicted swing voter probability of $x\%$ will be swing voters. The BSS rescales the Brier score for better interpretability.	$(-\infty, 1]$ A value of 0 indicates no improvement over a null classifier.	Brier scores represent the mean squared error of a binary classifier: $\frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2$ BSSs are calculated as $1 - \text{Brier score}_{\text{model}} / \text{Brier score}_{\text{null}}$.
Matthews correlation coefficient	Measures the strength of (binary) association between the observed and predicted classes. Incorporates information from all four cells of the confusion matrix and, as a symmetric metric, is appropriate when class imbalance is present (Chicco and Jurman 2020).	$[-1, 1]$ Higher values preferred.	Equivalent to the ϕ coefficient. Calculated as $\frac{\text{Cov}(y, l)}{\sigma_y \sigma_l}$, where l is the vector of predicted classes.
Cross-entropy loss	Provides an information-theoretic metric to measure the strength of association between the two distributions p and y . Quantifies the amount of uncertainty (i.e., entropy) in y reduced by p .	$[0, \infty)$ Lower values preferred.	Equivalent to the (negative) log likelihood of the model (Murphy 2012). Calculated as $-\sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i)$.

Notes: y = response variable, p = predicted probabilities (positive class), l = predicted classes.

Table 2. Model performance on the out-of-sample (validation) data.

	Accuracy	AUC-ROC	Sensitivity (TPR)	Specificity (TNR)	BSS	MCC	CE loss
Learning ensemble	0.90	0.90	0.57	0.95	0.33	0.55	0.26
Generalized additive model	0.87	0.82	0.22	0.97	0.17	0.28	0.32
Extremely randomized trees	0.88	0.89	0.67	0.92	0.18	0.54	0.32
Gradient boosting machine	0.89	0.88	0.58	0.93	0.27	0.51	0.28
Group-lasso interaction network	0.87	0.85	0.19	0.98	0.23	0.27	0.29
Support vector machine	0.89	0.88	0.55	0.94	0.27	0.51	0.29

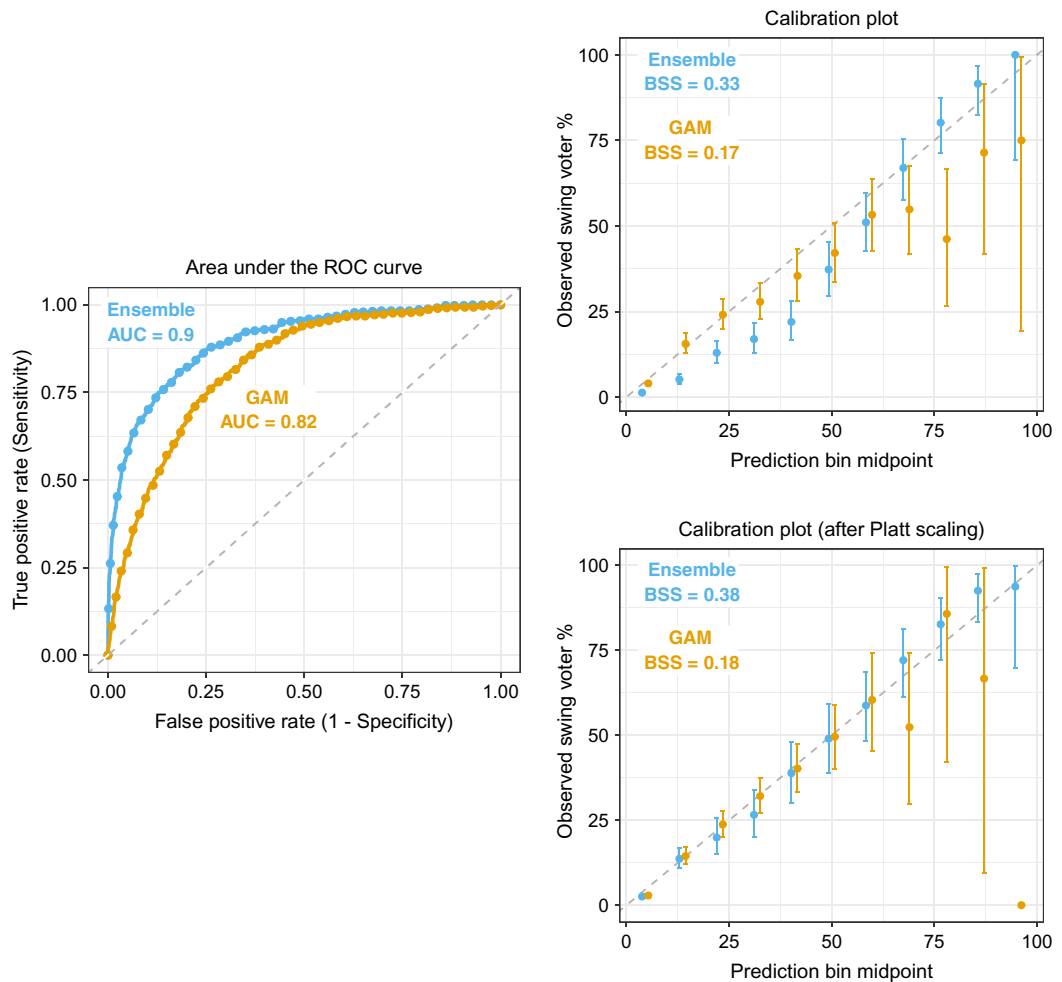


Figure 1. Comparison of the learning ensemble with the baseline generalized additive model (GAM) on the out-of-sample (validation) data. *Notes:* AUC, area under the ROC curve; BSS, Brier skill score. The bars show 95% confidence intervals for the estimated proportions.

calibrated model will have a curve that overlays the 45° reference line (e.g., the bin prediction midpoints will exactly match the observed proportions). The top-right panel of Figure 1 shows that the GAM predictions are well calibrated for respondents with low propensity scores, but are noisy and poorly calibrated at higher predicted values. Conversely, the learning ensemble exhibits a sigmoid calibration curve, falling slightly below (above) the reference line at lower (higher) predicted values.

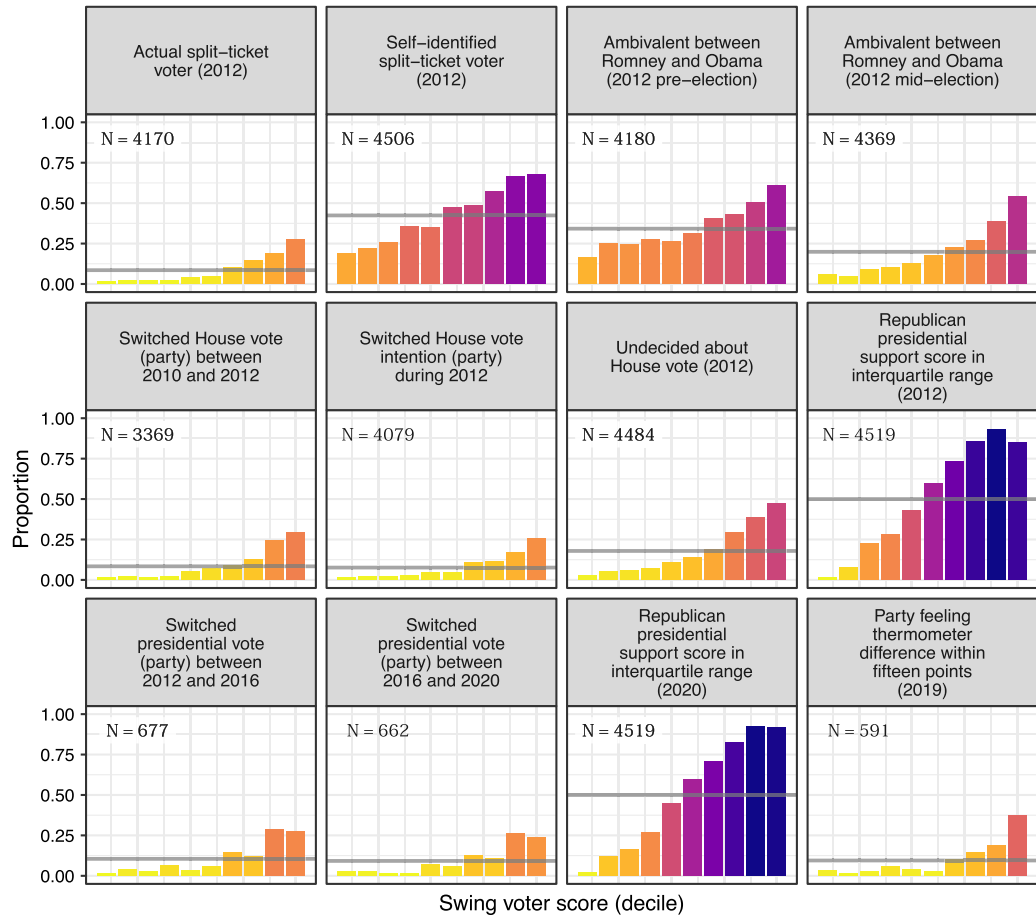


Figure 2. Predictive performance on additional indicators of swing voter propensity using out-of-sample (validation) data. The horizontal bars show the overall proportion of respondents satisfying the corresponding indicator. Ambivalence defined as placing the candidates within one point of each other on a five-point favorability scale. Republican presidential support scores are calculated by estimating ensemble models of 2012 presidential vote choice and 2020 presidential vote intention. Additional details are provided in Appendices D and E in the Supplementary Material.

The BSSs confirm that the learning ensemble provides better calibrated predictions than the GAM (BSS values of 0.33 vs. 0.17), but especially given the ensemble calibration curve’s sigmoid pattern, we can achieve further improvement in model calibration by applying Platt scaling to transform the predictions. The ensemble’s transformed predictions exhibit much greater correspondence between the observed and predicted proportions of swing voters for the learning ensemble (its BSS value jumps from 0.33 to 0.38 following Platt scaling). The improvement in calibration for the GAM model is only marginal (an increase in BSS from 0.17 to 0.18).

We next evaluate the ensemble by assessing its predictive performance on 12 indicators related to swing voter propensity. These include measures of attitudinal ambivalence, split-ticket voting, and swing voting in non-presidential elections (Mulligan 2011). This is a more rigorous test of the ensemble, since it is performed on variables distinct from the specific swing voter measure in the model. Because the 2012 CCAP is a part of a larger panel study (the VOTER Study Group), we can also evaluate how well the propensity scores predict behavior in later election cycles (e.g., swing voting in the 2016 and 2020 presidential elections).

Figure 2 divides the swing voter scores into deciles and shows the proportion of respondents in each percentile who fit the corresponding criteria. For instance, the top-left panel shows that virtually no respondents in the bottom deciles were split-ticket voters in 2012. If the model predictions tap into an underlying swing voter propensity, we should find a monotonic or near-monotonic

relationship in each panel. That is, higher scores should correspond to a greater likelihood of swing voter-type behaviors.

This is precisely what Figure 2 shows. Moving from left to right in each panel, we find a greater proportion of respondents engaging in split-ticket voting, being undecided or exhibiting volatile preferences in congressional elections, or possessing ambivalent attitudes, and switching between parties across later presidential elections. The predicted scores do not just predict the specific behavior defined by our measure, but other attitudes and choices corresponding to a more volatile electoral profile.

We also use the same predictor variables to estimate a learning ensemble that predicts respondents' probability of voting for the Republican candidate in the 2012 and 2020 elections. Major political campaigns routinely utilize these methods to generate such support scores for a universe of voters—frequently with the use of decision-tree-based approaches such as chi-square automatic interaction detection (CHAID) and random forests (Nickerson and Rogers 2014). Campaigns often then use those with support scores in the middle range of the scale to identify persuasion targets (Hersh 2015; Issenberg 2013). These are the voters that campaigns *perceive* as movable, and consequently they play an outsized role in shaping and implementing campaign strategy (Hersh 2015). We find that our swing voter model converges on many of the same voters that presidential campaigns are likely to view as persuadable (i.e., those with Republican presidential support scores in the interquartile range).

5 Identifying Predictors of Swing Voting

Another promising application of supervised machine learning is an exploratory tool in the model building and evaluation process. Owing to their flexibility, these methods can identify promising predictor variables whose effects have been overlooked either for theoretical reasons or because they are obscured by functional form complexity. That is, a given predictor may share a complex (but nonetheless meaningful) relationship with the response variable that escapes notice when using a series of linear additive terms in a regression model.

Machine learning is already widely used to detect heterogeneous treatment effects in experimental studies (e.g., Green and Kern 2012; Montgomery and Olivella 2018). Likewise, these methods can also alert us to meaningful predictors meriting further investigation when we have reasons to believe (as with swing voters) that the data generating process (DGP) is interactive in high-dimensional space. In such cases, Breiman (2001b) contends that feature importance is best measured using models with high levels of predictive accuracy.¹¹ The growing adoption of predictive modeling in the social sciences necessitates tools to probe the results from supervised machine learning algorithms (see generally Bicek and Burzykowski 2021; Molnar 2019).

In order to gain insight into the factors driving swing voter behavior, below we present two strategies for measuring feature importance in black-box prediction models.¹² We emphasize that both strategies are *model agnostic*: they can be used to interrogate any black-box prediction model (or ensemble of models). The unifying principle is that removing a feature X_j from a model breaks any meaningful relationship (however, simple or complex) it shares with the response variable y . Comparing model performance before and after removing X_j then provides an estimate of that feature's contribution to the ensemble that can be interpreted in information-theoretic terms.

11 Ensemble methods also serve to diminish the “Rashomon effect,” which arises when dissimilar models produce similar fits to the data and offer competing accounts of feature importance (p. 206).

12 Hill and Jones (2014) use a similar approach to assess the predictive power of factors driving state repression and human rights violations. As they note, “evaluating the ability of a model to predict state violence out-of-sample offers at least an addition, and perhaps a better, way of assessing the veracity of explanations for its occurrence” (p. 662). Our approach also closely parallels Carmines and Stimson (1980), who model vote choice as a function of different subsets of policy and partisan variables and use differences in predictive power as evidence of issue voting on easy and hard issues.

This basic strategy unifies a broad set of specific techniques for black-box model explanation, as discussed in (Covert, Lundberg, and Lee 2021).¹³

In the first approach, we proceed by estimating the learning ensemble using all features and generating model predictions. We then simulate each feature's removal by randomly permuting its values and estimating new model predictions. The difference in model fit (using the same four key metrics presented earlier) provides a permutation-based measure of that feature's contribution to the ensemble (Breiman 2001a).¹⁴ In the second approach (sometimes referred to as feature ablation or leave-one-covariate-out), we manually remove a feature (or subset of features) before estimating the ensemble. We then compare the predictive performance of each ablated model to the full model. Though computationally intensive, this technique allows us to assess whether the other predictors can compensate for the loss of information in the omitted features.¹⁵

Figure 3 shows the mean difference in fit metric for the 20 most influential predictor variables.¹⁶ The four metrics give similar estimates of feature importance (all pairs correlated at $r \geq 0.92$), and each identifies party identification as the most predictive characteristic of swing voters. Permutation on partisanship accounts for between 18% (CE loss) and 30% (AUC-ROC) of the total drop-off in ensemble performance.

After party identification, the learning ensemble derives a considerable portion of its predictive power from both symbolic and operational ideological factors. This includes ideological self-identification alongside specific policy attitudes (with economic and environmental issue preferences foremost among these). The general ideological inconsistency measure is also highly influential, but the model does not overly rely on this variable to parse out the specific policy cross-pressures driving swing voter propensity. Collectively, the policy-related variables supply between 34% (AUC-ROC) and 45% (CE loss) of the model's aggregate predictive power. This effect does not appear to be an artifact of projection (or a "follow the leader" effect) in which voters adopt a preferred candidate or party's issue stances (e.g., Lenz 2012), as the issue and ideological measures come from the pre-election, baseline wave of the 2012 CCAP.

The learning ensemble reveals several other influential predictors worth noting, including racial resentment and non-policy "nature of the times" assessments such as right/wrong track attitudes and collective (sociotropic) economic evaluations. This echoes Zaller's (2004) finding that floating voters place greater emphasis on performance-based considerations. Political knowledge, though influential, is not a dominant predictor in the model. Instead, the learning ensemble uses a broader range of features to make predictions and reveals that policy-related considerations are indeed meaningful factors underlying swing voter behavior.

Turning to feature ablation, we predict swing voting by estimating ensemble models (following the same process as above) with 12 combinations of the demographic, issue, and salience items. Table 3 shows the specific predictor variables included in each subset and the corresponding fit statistics. The results support prior claims that information available in public records—variables such as age, voting history, Census block-group statistics, and (in some states) race and party registration—poorly predicts persuadable voters (e.g., Endres 2016; Hersh 2015). The subset **D1** includes basic demographic variables, including race and political knowledge/interest (predictors

13 The removal-based approach also has a deep connection to a type of counterfactual reasoning in philosophy and psychology known as a *subtractive* counterfactual or Mill's method of difference (pp. 36–37).

14 Note that this represents an unconditional permutation test. In Appendix F in the Supplementary Material, we present a conditional permutation scheme (Debeer and Strobl 2020) and show that it produces virtually equivalent results (across the four fit statistics, $r \geq 0.95$ for the conditional and unconditional sets of feature importance estimates).

15 Friedman and Popescu (2008) discuss two additional strategies to recover interactive relationships identified by black box models: Friedman's *H*-statistic and prediction rule ensembles. We show how these techniques can be used to identify promising subgroups for future investigation in Appendix G in the Supplementary Material. For instance, both methods suggest a heightened association between swing voter propensity and right/wrong track evaluations among ideologically inconsistent Democrats.

16 The results for all predictor variables are provided in Appendix F in the Supplementary Material.

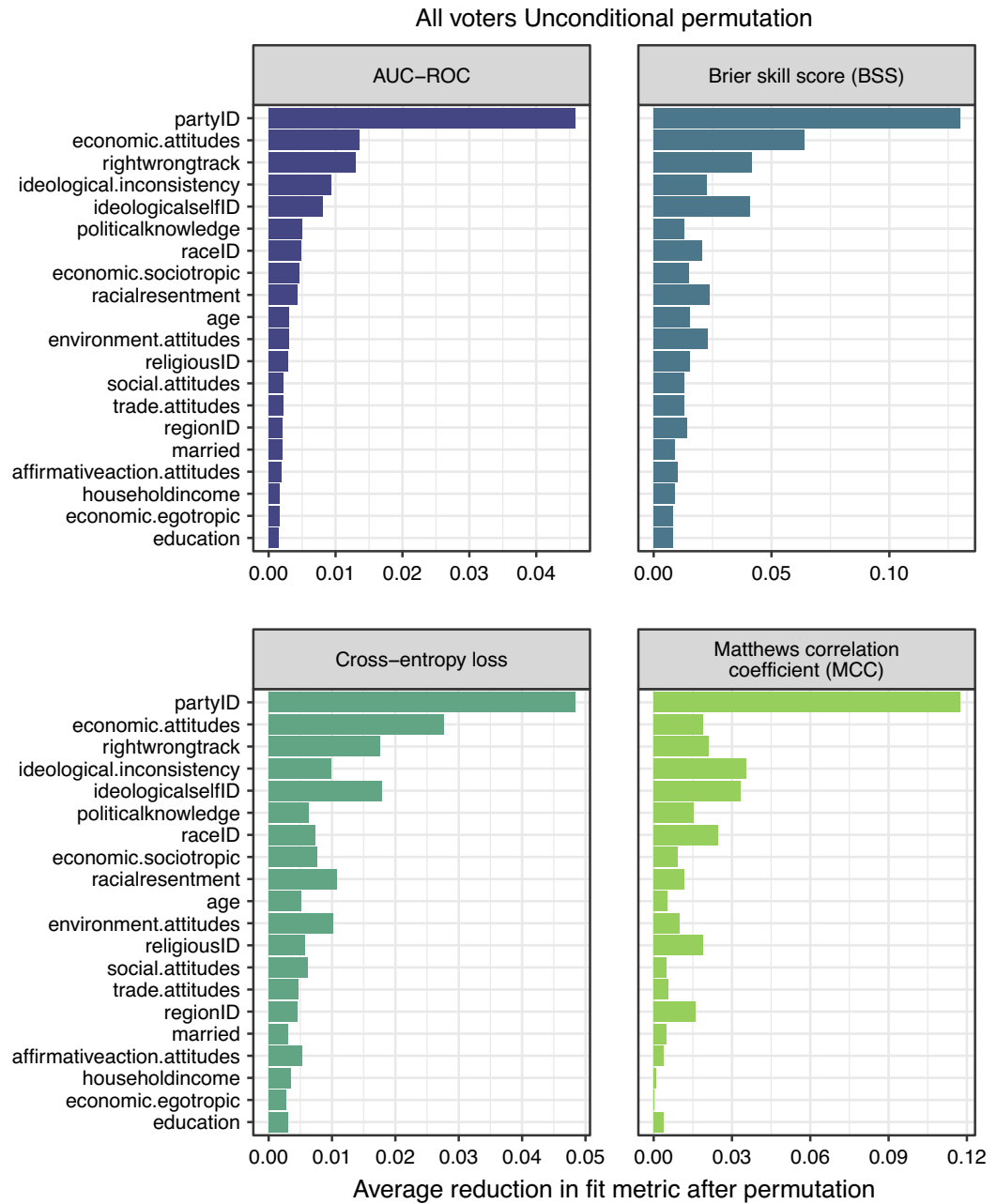


Figure 3. Feature importance estimates from the learning ensemble. Five-hundred random permutations conducted using out-of-sample (validation) data.

that proxy past voting turnout). With this restricted set of variables, the ensemble is both poorly discriminatory (an AUC-ROC of 0.74) and poorly calibrated (a BSS of -0.02). Adding an enhanced set of demographic variables in the subset **D2** (which includes predictors such as education and gun ownership that could be proxied with consumer data) increases performance only slightly.

The largest improvements in predictive performance occur once we add party identification and especially policy variables to the ensemble. For instance, subset **I4** (which removes demographic variables from the model) produces fit metrics that are in line with those obtained for the full learning ensemble. Consistent with the results from the permutation analysis, a combination of ideological dispositions, policy preferences, and issue salience attitudes are essential to accurately predict swing voters.

Table 3. Learning ensemble performance in predicting swing voters with subsets of predictor variables.

Fit metrics	D1	D2	D3	D4	D5	D6	I1	I2	I3	I4	S1	S2
AUC-ROC	0.74	0.80	0.83	0.86	0.87	0.82	0.84	0.87	0.87	0.88	0.79	0.85
BSS	-0.02	0.02	0.13	0.19	0.17	0.06	0.14	0.16	0.18	0.23	0.01	0.17
MCC	0.27	0.42	0.39	0.46	0.49	0.46	0.40	0.49	0.47	0.53	0.36	0.46
CE loss	0.39	0.38	0.33	0.32	0.32	0.37	0.33	0.32	0.31	0.30	0.38	0.32
Predictors	D1	D2	D3	D4	D5	D6	I1	I2	I3	I4	S1	S2
quarter	X	X	X	X	X	X	X	X	X	X	X	X
politicalknowledge	X	X	X	X	X	X	X	X	X	X	X	X
politicalinterest	X	X	X	X	X	X	X	X	X	X	X	X
regionID	X	X	X	X	X	X						
age	X	X	X	X	X	X						
female	X	X	X	X	X	X						
racelD	X	X	X	X	X	X						
householdincome	X	X	X	X	X	X						
married	X	X	X	X	X	X						
children	X	X	X	X	X	X						
education		X		X								
LGBT		X		X								
religiosity		X		X								
religiousID		X		X								
owngun		X		X								
unionmember		X		X								
havehealthinsurance		X		X								
economic.egotropic		X		X								
partyID			X	X					X	X		X
ideologicalselfID					X		X	X	X	X		
ideological.inconsistency					X		X	X	X	X		
economic.attitudes					X		X	X	X	X		
social.attitudes					X		X	X	X	X		
immigration.attitudes					X		X	X	X	X		
environment.attitudes					X		X	X	X	X		
deathpenalty.attitudes					X		X	X	X	X		
affirmaction.attitudes					X		X	X	X	X		
trade.attitudes					X		X	X	X	X		
issue.salience						X		X		X	X	X

6 Discussion

In 1960, a team of political scientists—armed with 130,000 survey responses, a 20-ton IBM 704, and a theory of cross-pressures—began the Simulmatics Project (Pool, Abelson, and Popkin 1964). The project simulated results for the 1960 presidential election by organizing respondents into voter types and estimating how partisan swings resulting from cross-pressures (especially among Protestant Democrats and Catholic Republicans) could affect the outcome. Their predictions were remarkably accurate, demonstrating how cross-pressure theory could be combined with raw computing power to improve models of voting behavior. Pool, Abelson, and Popkin (1964, pp. 181–182) reflected on the project:

[W]e are moving into an era where the limitation on data bank operations will no longer be computer capacity but only the imagination of the researcher and the extent to which raw data are available to him. . . The tools that we are now developing can handle models of a complexity adequate to the complexity of the political process itself.

In the same spirit as the Simulatics Project, this paper leverages technological and methodological advancements to explore swing voter behavior in a novel way, using a suite of supervised machine learning methods to search the lattice of cross-pressures permeating the electorate and identifying those most predictive of swing voting. This strategy yields improved estimates of individual swing voter propensity, reiterating the benefits of ensembling and model averaging (e.g., Grimmer *et al.* 2017; Montgomery, Hollenbach, and Ward 2012). One possible application of these propensity scores is to improve polling estimates by better identifying swing voters.¹⁷

However, black-box models need not be limited to pure prediction tasks. Indeed, precisely because these methods are so predictively powerful, they also serve as a valuable exploratory tool—alerting us to untapped information in the data and suggesting new avenues for investigation. As (Grimmer *et al.* 2021, 10) note, “[r]esearchers often discover new directions, questions, and measures within quantitative data. . . and machine learning can facilitate those discoveries.”

In the application to swing voters, removal-based experimentation on the learning ensemble reveals the importance of policy attitudes in the prediction task. Policy attitudes (specifically, patterns of salient ideological cross-pressures) cannot simply be proxies for other non-policy factors such as partisanship, political sophistication, and demographic characteristics. If this were the case, then the learning ensemble should be able to compensate when policy-related features are removed. Instead, model performance noticeably deteriorates without the pre-election measures of policy attitudes. Using an explicitly predictive modeling strategy, the results demonstrate the difficulty of identifying swing voters without rich individual-level data on voters’ ideological orientations and policy attitudes (cf. Endres 2016; Hersh 2015). From a predictive standpoint, the caricature of swing voters as simply politically unaffiliated and uninformed “rolling stones” of the electorate neglects a suite of consequential cross-pressures underlying swing voter behavior.

While the approach presented here is primarily exploratory, future work should continue to unpack the learning ensemble and more formally test-specific high-order cross-pressures involving factors that the ensemble identifies as important.¹⁸ We think two methods are especially well suited to this task. The first involves the use of causal forests (Wager and Athey 2018) as a semiparametric means of identifying heterogeneous treatment effects for specific features. The second applies the game-theoretic concept of Shapley values—which are concerned with allocating credit to players in coalitional games—to the problem of estimating feature contributions to model predictions (Lundberg *et al.* 2020). Both methods are promising options to further enhance ensemble interpretability.¹⁹

In addition to the specific methodological tools presented here, we also hope that this exercise serves to demonstrate the utility of engaging the study of voting behavior from a predictive modeling standpoint. We can imagine several substantive settings where this approach could reveal other novel insights into existing questions. For instance, in an era of increasingly nationalized political behavior (Hopkins 2018; though see Kuriwaki 2020), are the same cross-pressures that motivate swing voter propensity at the national level also predictive of volatility in state and local

17 (Rentsch, Schaffner, and Gross 2019), for example, demonstrate how random forests and supplemental demographic data can be used to derive better estimates of turnout probabilities and screen likely voters in survey research.

18 For instance, information about interaction importance in predictive models could be used inform experimental designs that simulate the effects of cross-pressures on voting behavior (e.g., Endres and Panagopoulos 2019).

19 Although we obtain virtually equivalent results from the unconditional and conditional permutation tests of variable importance, we also note that accounting for feature dependencies when calculating importance remains an active area of research (e.g., Chan and Ratkovic 2020; Debeer and Strobl 2020).

elections?²⁰ The ensemble could also further leverage longitudinal data to distinguish between persistent and transitory cross-pressures by evaluating their predictive power across electoral cycles.

Acknowledgments

We would like to thank Bob Ellsworth, Ben Highton, Adrienne Hosek, Bob Lupton, participants in the UC Davis American Politics Reading Group, the editor, and three anonymous reviewers for their helpful feedback.

Data Availability Statement

The replication materials for this paper can be found at <https://doi.org/10.7910/DVN/CLQY60> (Hare and Kutsuris 2022).

Supplementary Material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2022.24>.

References

- Ali, K. M., and M. J. Pazzani. 1996. "Error Reduction through Learning Multiple Descriptions." *Machine Learning* 24 (3): 173–202.
- Bansak, K. 2019. "Can Nonexperts Really Emulate Statistical Learning Methods? A Comment on 'The Accuracy, Fairness, and Limits of Predicting Recidivism.'" *Political Analysis* 27 (3): 370–380.
- Berelson, B. R., P. F. Lazarsfeld, and W. N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.
- Bicek, P., and T. Burzykowski. 2021. *Exploratory Model Analysis: Explore, Explain, and Examine Predictive Models*. Boca Raton: CRC Press.
- Box-Steffensmeier, J., M. Dillard, D. Kimball, and W. Massengill. 2015. "The Long and Short of It: The Unpredictability of Late Deciding Voters." *Electoral Studies* 39: 181–194.
- Brader, T., J. A. Tucker, and A. Theriault. 2014. "Cross Pressure Scores: An Individual-Level Measure of Cumulative Partisan Pressures Arising from Social Group Memberships." *Political Behavior* 36 (1): 23–51.
- Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 24 (2): 123–140.
- Breiman, L. 2001a. "Random Forests." *Machine Learning* 45 (1): 5–32.
- Breiman, L. 2001b. "Statistical Modeling: The Two Cultures." *Statistical Science* 16 (3): 199–231.
- Campbell, A. 1960. "Surge and Decline: A Study of Electoral Change." *Public Opinion Quarterly* 24 (3): 397–418.
- Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes. 1960. *The American Voter*. New York: Wiley.
- Carmines, E. G., and J. A. Stimson. 1980. "The Two Faces of Issue Voting." *American Political Science Review* 74 (1): 78–91.
- Carsey, T. M., and G. C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate." *American Journal of Political Science* 50 (2): 464–477.
- Chan, Z., and M. Ratkovic. 2020. "Improving Variable Importance Measures." Presented at the 2020 Annual Meeting of the Society for Political Methodology (PolMeth XXXVII), University of Toronto, Toronto, CA, 14–17 July 2020.
- Chicco, D., and G. Jurman. 2020. "The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation." *BMC Genomics* 21 (6): 1–13.
- Claassen, R. L. 2011. "Political Awareness and Electoral Campaigns: Maximum Effects for Minimum Citizens?" *Political Behavior* 33 (2): 203–223.
- Claassen, R. L., and B. Highton. 2009. "Policy Polarization among Party Elites and the Significance of Political Awareness in the Mass Public." *Political Research Quarterly* 62 (3): 538–551.
- Converse, P. E. 1962. "Information Flow and the Stability of Partisan Attitudes." *Public Opinion Quarterly* 26 (4): 578–599.
- Covert, I., S. Lundberg, and S.-I. Lee. 2021. "Explaining by Removing: A Unified Framework for Model Explanation." *Journal of Machine Learning Research* 22 (209): 1–90.
- Dassonneville, R., and Y. Dejaeghere. 2014. "Bridging the Ideological Space: A Cross-National Analysis of the Distance of Party Switching." *European Journal of Political Research* 53 (3): 580–599.
- Debeer, D., and C. Strobl. 2020. "Conditional Permutation Importance Revisited." *BMC Bioinformatics* 21 (1): 1–30.

²⁰ We thank an anonymous reviewer for this suggestion.

- Democracy Fund Voter Study Group. 2020. *Views of the Electorate Research Survey, January 2019. Release 1: January 2020*. Washington: Democracy Fund. <https://www.voterstudygroup.org/>.
- Ellis, C., and J. A. Stimson. 2012. *Ideology in America*. Cambridge: Cambridge University Press.
- Endres, K. 2016. "The Accuracy of Microtargeted Policy Positions." *PS: Political Science & Politics* 49 (4): 771–774.
- Endres, K., and C. Panagopoulos. 2019. "Cross-Pressure and Voting Behavior: Evidence from Randomized Experiments." *Journal of Politics* 81 (3): 1090–1095.
- Federico, C. M., and C. V. Hunt. 2013. "Political Information, Political Involvement, and Reliance on Ideology in Political Evaluation." *Political Behavior* 35 (1): 89–112.
- Freund, Y., and R. E. Schapire. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* 55 (1): 119–139.
- Friedman, J. H., and B. E. Popescu. 2008. "Predictive Learning via Rule Ensembles." *The Annals of Applied Statistics* 2 (3): 916–954.
- Gelman, A., S. Goel, D. Rivers, and D. Rothschild. 2016. "The Mythical Swing Voter." *Quarterly Journal of Political Science* 11 (1): 103–130.
- Gill, J. 2021. "Political Science Is a Data Science." *Journal of Politics* 83 (1): 1–7.
- Green, D. P., and H. L. Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly* 76 (3): 491–511.
- Green, J. 2020. "Floating Policy Voters in the 2016 U.S. Presidential Election." *Electoral Studies* 67: 1–8.
- Grimmer, J., S. Messing, and S. J. Westwood. 2017. "Estimating Heterogeneous Treatment Effects and the Effects of Heterogeneous Treatments with Ensemble Methods." *Political Analysis* 25 (4): 413–434.
- Grimmer, J., M. E. Roberts, and B. M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24 (1): 1–25.
- Hare, C., and M. Kutsuris. 2022. "Replication Data for: Measuring Swing Voters with a Supervised Machine Learning Ensemble." Harvard Dataverse, V1. <https://doi.org/10.7910/DVN/CLQY60>.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edn. New York: Springer.
- Hersh, E. D. 2015. *Hacking the Electorate*. New York: Cambridge University Press.
- Hill, D. W., and Z. M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108 (3): 661–687.
- Hill, S. J. 2017. "Changing Votes or Changing Voters? How Candidates and Election Context Swing Voters and Mobilize the Base." *Electoral Studies* 48: 131–148.
- Hillygus, D. S., and T. Shields. 2008. *The Persuadable Voter: Wedge Issues in Presidential Campaigns*. Princeton: Princeton University Press.
- Hopkins, D. J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. Chicago: University of Chicago Press.
- Issenberg, S. 2013. *The Victory Lab*. New York: Crown.
- Jackman, S., J. Sides, J. Tesler, and L. Vavreck. 2012. *The 2012 Cooperative Campaign Analysis Project*. Palo Alto: YouGov/Polimetrix.
- Katz, R. S., H. Rattinger, and M. N. Pedersen. 1997. "The Dynamics of European Party Systems." *European Journal of Political Research* 31 (1): 83–97.
- Key, V., Jr. 1966. *The Responsible Electorate*. New York: Vintage.
- Klar, S. 2014. "A Multidimensional Study of Ideological Preferences and Priorities among the American Public." *Public Opinion Quarterly* 78 (S1): 344–359.
- Krosnick, J. A. 1990. "Government Policy and Citizen Passion: A Study of Issue Publics in Contemporary America." *Political Behavior* 12 (1): 59–92.
- Kuhn, M., and K. Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.
- Kuriwaki, S. 2020. "Ticket Splitting in a Nationalized Era." Working Paper. <https://osf.io/preprints/socarxiv/bvgz3/>.
- Lampa, E., L. Lind, P. M. Lind, and A. Bornefalk-Hermansson. 2014. "The Identification of Complex Interactions in Epidemiology and Toxicology: A Simulation Study of Boosted Regression Trees." *Environmental Health* 13 (1): 57.
- Lazarsfeld, P. F., B. R. Berelson, and H. Gaudet. 1944. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.
- Lenz, G. S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago: University of Chicago Press.
- Lundberg, S. M., et al. 2020. "From Local Explanations to Global Understanding with Explainable AI for Trees." *Nature Machine Intelligence* 2 (1): 56–67.
- Lupton, R. N., W. M. Myers, and J. R. Thornton. 2015. "Political Sophistication and the Dimensionality of Elite and Mass Attitudes, 1980–2004." *Journal of Politics* 77 (2): 368–380.
- Mayer, W. G. 2007. "The Swing Voter in American Presidential Elections." *American Politics Research* 35 (3): 358–388.
- Mayer, W. G. 2008. "What Exactly Is a Swing Voter? Definition and Measurement." In *The Swing Voter in American Politics*, edited by W. G. Mayer, 1–31. Washington: Brookings Institution Press.

- Molnar, C. 2019. *Interpretable Machine Learning*. <https://christophm.github.io/interpretable-ml-book>.
- Montgomery, J. M., F. M. Hollenbach, and M. D. Ward. 2012. "Improving Predictions using Ensemble Bayesian Model Averaging." *Political Analysis* 20 (3): 271–291.
- Montgomery, J. M., and S. Olivella. 2018. "Tree-Based Models for Political Science Data." *American Journal of Political Science* 62 (3): 729–744.
- Mulligan, K. 2011. "Partisan Ambivalence, Split-Ticket Voting, and Divided Government." *Political Psychology* 32 (3): 505–530.
- Murphy, K. P. 2012. *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press.
- Nickerson, D. W., and T. Rogers. 2014. "Political Campaigns and Big Data." *Journal of Economic Perspectives* 28 (2): 51–74.
- Petrocik, J. R., W. L. Benoit, and G. J. Hansen. 2003. "Issue Ownership and Presidential Campaigning, 1952–2000." *Political Science Quarterly* 118 (4): 599–626.
- Platt, J. C. 2000. "Probabilistic Outputs for Support Vector Machines." In *Advances in Large Margin Classifiers*, edited by A. J. Smola, P. L. Bartlett, B. Schölkopf, and D. Schuurmans, 61–74. Cambridge: MIT Press.
- Pool, I. d. S., R. Abelson, and S. Popkin 1964. *Candidates, Issues, and Strategies*. Cambridge: MIT Press.
- Rentsch, A., B. F. Schaffner, and J. H. Gross. 2019. "The Elusive Likely Voter." *Public Opinion Quarterly* 83 (4): 782–804.
- RePass, D. E. 1971. "Issue Salience and Party Choice." *American Political Science Review* 65 (2): 389–400.
- Schapiro, R. E. 1990. "The Strength of Weak Learnability." *Machine Learning* 5 (2): 197–227.
- Smidt, C. D. 2017. "Polarization and the Decline of the American Floating Voter." *American Journal of Political Science* 61 (2): 365–381.
- Treier, S., and D. S. Hillygus. 2009. "The Nature of Political Ideology in the Contemporary Electorate." *Public Opinion Quarterly* 73 (4): 679–703.
- van der Laan, M. J., E. C. Polley, and A. E. Hubbard. 2007. "Super Learner." *Statistical Applications in Genetics and Molecular Biology* 6 (1): 1–23.
- Wager, S., and S. Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–1242.
- Weghorst, K. R., and S. I. Lindberg. 2013. "What Drives the Swing Voter in Africa?" *American Journal of Political Science* 57 (3): 717–734.
- Wolpert, D. H. 1996. "The Lack of A Priori Distinctions between Learning Algorithms." *Neural Computation* 8 (7): 1341–1390.
- Zaller, J. 2004. "Floating Voters in U.S. Presidential Elections, 1948–2000." In *Studies in Public Opinion: Attitudes, Nonattitudes, Measurement Error, and Change*, edited by W. E. Saris and P. M. Sniderman, 166–212. Princeton: Princeton University Press.