The Swing Voter Paradox: Electoral Politics in a Nationalized Era

The Swing Voter Paradox:
Electoral Politics in a Nationalized Era

A dissertation presented

by

Shiro Kuriwaki

to

The Department of Government

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Political Science

Harvard University
Cambridge, Massachusetts

April 2021

Dissertation Advisors:
Stephen Ansolabehere and James M. Snyder, Jr. Shiro Kuriwaki

The Swing Voter Paradox: Electoral Politics in a Nationalized Era

Abstract

With each successive election since at least 1994, congressional elections in the United States have transitioned toward nationalized two-party government. Fewer voters split their tickets for different parties between President and Congress. Regional blocs and incumbency voting — a key feature of U.S. elections in the latter 20th century — appear to have given way to strong party discipline among candidates and nationalized partisanship among voters. Observers of modern American politics are therefore tempted to write off the importance of the swing voter, defined here as voters who are indifferent between the two parties and thus likely to split their ticket or switch their party support.

By assembling data from historical elections (1950 – 2020), surveys (2008 – 2018), and cast vote record data (2010 – 2018), and through developing statistical methods to analyze such data, I argue that although they comprise a smaller portion of the electorate, each swing voter is disproportionately decisive in modern American politics, a phenomenon I call the *swing voter paradox*. Historical comparisons across Congressional, state executive, and state legislative elections confirm the decline in aggregate measures of ticket splitting suggested in past work. But the same indicator has not declined nearly as much in county legislative or county sheriff elections (Chapter 1). Ticket splitters and party switchers tend to be voters with low news interest and ideological moderate. Consistent with a spatial voting model with valence, voters also

become ticket splitters when incumbents run (Chapter 2). I then provide one of the first direct measures of ticket splitting in state and local office using cast vote records. I find that ticket splitting is more prevalent in state and local elections (Chapter 3). This is surprising given the conventional wisdom that party labels serve as heuristics and down-ballot elections are low information environments.

A major barrier for existing studies of the swing voter lies in the measurement from incomplete electoral data. Traditional methods struggle to extract information about subgroups from large surveys or cast vote records, because of small subgroup samples, multi-dimensional data, and systematic missingness. I therefore develop a procedure for reweighting surveys to small areas through expanding poststratification targets (Chapter 4), and a clustering algorithm for survey or ballot data with multiple offices to extract interpretable voting blocs (Chapter 5). I provide open-source software to implement both methods.

These findings challenge a common characterization of modern American politics as one dominated by rigidly polarized parties and partisans. The picture that emerges instead is one where swing voters are rare but can dramatically decide the party in power, and where no single demographic group is a swing voter. Instead of entrenching elections into red states and blue states, nationalization may heighten the role of the persuadable voter.

Contents

List of Figures

List of Tables

Acknowledgments

Ten years ago in the Spring of 2011, I was taking my first statistics class, my first lecture for Introduction to American Politics, and my first political science seminar. I owe all that I have learned since that semester, including this dissertation, to the people who have helped me and enriched my life along the way.

My main dissertation advisors were formidable. From Steve Ansolabehere, with whom I took my first American Politics class at Harvard in my first semester, I learned how to formulate long term research while being excited to come to work every day. From Kosuke Imai, I learned to push myself just the right amount to meet higher standards and the importance of helping fellow students while learning immensely from them in turn. From Jim Snyder, I learned how to spot an interesting research question and go about turning that into a paper, and an appreciation for hours of data collection to find a single statistic. From all three of them, I learned how to think clearly and write elegantly.

Over the course of my time at Harvard, I also benefited from the advice from many other faculty and teachers, including Matt Blackwell, Amy Catalinac, Ryan Enos, Andrew Hall, Gary King, Dan Levy, Xiao-Li Meng, Susan Pharr, Jon Rogowski, Maya Sen, Dan Smith, Teddy Svoronos, Dustin Tingley, and Ista Zahn. Even in their busiest moments, their comments led to one discovery after another.

One of the best experiences as a graduate student at the Government department was the time spent with amazing peers. I learned immensely from Pam Ban, Peter

Bucchianeri, Dan De Kadt, Connor Huff, Chris Lucas, Michael Morse, Daniel Moskowitz, and Melissa Sands, who all guided my path as I entered the program. Those who entered at the same time or after me also were sources of inspiration and fun, including Jake Brown, Chris Chaky, Angelo Dagonel, Naoki Egami, Shusei Eshima, Max Goplerud, Jaclyn Kaslovsky, Chris Kenny, Hanno Hilbig, Michael Olson, Yon Soo Park, Hunter Rendleman, Albert Rivero, Tyler Simko, Diana Stanescu, and Michael Zoorob. Outside of the department I had the good fortune of knowing Alexander Agadjanian and Shun Yamaya. I am especially thankful to Andy Stone and Soichiro Yamauchi, whose friendship kept me going on more days than they probably realized. Conversations with both lightened up any day. In addition, two of the chapters in this dissertation would not have been possible without Soichiro's substantial help.

In my personal and pre-graduate school life, I would first like to thank Susan Fiske, my undergraduate advisor and one of my academic heroes. Christina Davis, who taught that first political science seminar and wrote my first letter of recommendation, has been a generous advisor. I am grateful for Youngsuk Chi for believing in my future in academia and life in the United States. After graduation I was lucky to join the Analyst Institute, through which I started to learn from Jonathan Robinson, Josh Rosmarin, Aaron Strauss, and Annie Wang, and others. Special appreciation goes to Yi Yang for her warm support as well as to Jacky Cheng and Chris Cochran, with whom I have effectively lived together for nearly a decade. I credit Chris for many of the life habits that helped me stay health and happy through graduate school.

I have failed to list many others who have helped me through this process. My final and deepest thanks goes to my family, especially my mother. Her commitment to my education and personal happiness came with sacrifices to her time and career, I cannot overemphasize my appreciation of how much that has shaped my trajectory into graduate school and beyond.

Introduction

Changes in the vote choice of a small subset of the electorate can dramatically swing election results, making voters who deviate in any way from a straight party ticket a perennial interest for social scientists and political campaigns alike. This volatile electorate are often called swing voters. Understanding the prevalence and characteristics of these swing voters is a cornerstone to understanding why strategic parties and re-election seeking politicians take the positions they do. This dissertation collects data from multiple sources — historical election results, survey data, and cast vote records — to provide a sytematic and in-depth accounting of the prevalence and voting pattenrs of swing voters.[1] Of particular interest is whether (or in which eras and elections) the groups of voters I identify as swing voters is consequential for who wins an election.

In this introductory chapter, I provide a definition for the swing voter and a theoretical rationale for how that latent concepts relates to observable behavior such as ticket splitting and party switching. Without a clear definition, a "swing voter" is a rather elusive term that social science disciplines and journalistic coverage define in differing ways. For example, some include decisiveness as part of the definition of the swing voter, while in other definitions a swing voter is a easily persuadable voter regardless of whether they are pivotal or not. I define the swing voter in a spatial voting framework where voters choose candidates by comparing the utility they derive from

[1] Data and code to replicate some of this analysis is provided in Kuriwaki (2021c).

candidates, whose positions are defined on a left-right scale.

The notion of swing voters has been a focus in the study of electoral behavior and in theoretical models of political economy (Burden and Kimball 2002; Persson and Tabellini 2000, ch.8). Indeed I argue that this simple framework is appropriate because the definitions are clear, widely applicable, and illuminating in its own right. The framework justifies the use of *ticket splitting* and *party switching* as the key behavior I study empirically in the rest of this dissertation. Being a swing voter is a latent quality, but these two voting patterns are logical implications of a swing voter in a general spatial voting framework.

This formalization clarifies the difference with other ways scholars have defined the swing voter, as well as where the logic overlaps. One of the latest formalizations of the swing voter is by Mayer (2008), who focused on identifying the swing voter in the context of Presidential elections. His definition relies on a survey instrument often called the feeling thermometer, where respondents indicate how favorable or unfavorable they feel towards each presidential candidate. Mayer operationalizes the swing voter as voters who give the exact or approximately same value for both candidates, and argues that other related measures, such as party switching, being undecided, or independents, are less desirable mainly for measurement and interpretability. In a similar spirit, Hillygus and Shields (2009) investigated the characteristics of the persuadable voter, operationalized as a political independent or having issue positions that conflict with the party platform.

While the general strategy of this body of work is reasonable, taken literally its generalizability is limited. For example, both focus on a single office and the theoretical framework provides only suggestive guidance about modeling voting across multiple offices. The measurement strategy of a feeling thermometer is of limited applicability; most political surveys measure vote choice and partisanship but few consistently provide a feeling thermometer. And while, as Mayer argues, other measures such as being a moderate or undecided have their own measurement challenges, a black-or-white statement about whether these are correct ways to measure the swing voter makes it difficult to aggregate evidence across multiple classic studies, including V.O. Key's final work on floating voters (Key 1966).

In what follows, I provide a definition and measurement strategy for swing voters in a simple spatial voting framework. The general idea of the definition is consistent with what scholars such as Mayer (2008) have outlined, i.e. that swing voters are generally *indifferent* to the two alternatives. The spatial voting framework posits a left-right ideological spectrum and voters who make choices based on a cardinal measure of utility. This utility framework is in fact quite similar to the feeling thermometer measure. There is less conflict between my definition and Mayer as it might seem at first. The spatial voting framework can flexibly incorporate factors other than ideology. Here I consider two: a valence term and uncertainty.

In sum, the spatial voting framework turns out to be a simple and fruitful model to operationalize the concepts and justify why ticket splitting is a reasonable measure to operationalize the swing voter. As with all models, the goal of the model is not to fully predict behavior or describe the psychological process by which voters make the choices they do. Yet the model is illuminating because it logically shows how being a swing voter relates to ticket splitting or party switching. Perhaps more useful is that it also shows how the propensity for a voter to split a ticket is increasing in the ratio of two well-known factors in electoral politics: a candidate's valence advantage and the spatial distance, or polarization, between the two candidates.

Notation and Setup

The classic model of vote choice serves as a fundamental building block in defining a swing voter. Here voters are indexed by i, and have ideal points z_i on a left to right continuum. Candidates are members of a party $a \in \{\text{L}, \text{R}\}$ and are indexed by their office j. They have two attributes: their policy position x_j^a and valence v_j^a. Valence is any attribute that all voters prefer more of to less. In the US context, it may include incumbency and relevant experience for the job. We consider a case where candidates are nationalized, where we can simplify $x_j^a = x_{\tilde{j}}^a \;\; \forall a, j, \tilde{j}$. In other words, candidates of the same party have the same spatial position but may still have different valences.

This setup is non-strategic and non-dynamic. The main finding of valence models is not new (Ansolabehere and Snyder 2000; Groseclose 2007) but this description applies its insights to the case of ticket splitting.

There is only one player: the voter i. Voters have ideal point $\mathbf{x}_i^* \in \mathbb{R}$. They make binary choices between candidates in J offices. Candidates are members of a party $a \in \{\text{L}, \text{R}\}$ and run for one office j. Candidates have two fixed attributes: their policy position $\mathbf{x}_j^a \in \mathbb{R}$ and valence $v_j^a \in \mathbb{R}$. Both their policy and valence are fixed, for example by constraints in the primary election or national politics.

Definiton

Following convention I define a *swing voter* as voters for whom $U_i(z, x^{\text{L}}) \approx U_i(z, x^{\text{R}})$. In other words, these are voters who are largely indifferent between two parties. Indifference could also be defined with respect to particular candidates, such that a voter places equal utility for both candidates (their valence baked in). But formal theory models of the swing voter typically reserves such non-spatial attributes as potential shocks. They can include characteristics specific to each candidate or the distributional consequences of a particular party winning power, which all induce variation and uncertainty in election outcomes.

By this definition I distinguish between a swing voter and a pivotal voter. Some articles use "swing voter" to mean a voter that is both indifferent between the two choices *and* pivotal, in that they are the median voter and cast the deciding vote (Pesendorfer and Feddersen 1996). The swing voter in the definition I adopt need not be pivotal, and examining how swing and pivotality interact often leads to important insights (see e.g. Enns and Wohlfarth 2013, in the case of the Supreme Court). In chapters 1 and 2, I study how often the group of swing voters can be pivotal.

Model of Vote Choice

Each of a voter's outcome $y_{ij} \in \{\text{L}, \text{R}\}$ refers to a party choice by voter i in office j. Voters prefer candidates with closer policy positions, but they may also prefer more valence over less. Therefore in contest j a voter considers a quadratic utility for each party a with random measurement error:

$$\begin{aligned} U_i(\mathbf{x}_j^{\text{L}}, v_j^{\text{L}}) &= -\left(\mathbf{x}_i^* - \mathbf{x}_j^{\text{L}}\right)^2 + \theta v_j^{\text{L}} + \ell_{ij} \\ U_i(\mathbf{x}_j^{\text{R}}, v^{\text{R}}) &= -\left(\mathbf{x}_i^* - \mathbf{x}_j^{\text{R}}\right)^2 + \theta v_j^{\text{R}} + r_{ij} \end{aligned} \tag{1}$$

The key parameter of interest is $\theta \in \mathbb{R}$, which indicates the weight voters value valence relative to party. If $\theta = 0$, then voters only ignore valence and only vote party. If $\theta > 0$, some voters may defect from their party allegiance to vote for a high quality candidate. For simplicity θ is left constant for all voters, but we now let it vary by i in order to identify the model (see end of this section).

Decision Rule For tractability let both errors have a Normal distribution with the same mean, and let the variance of the difference of the two distributions be 1, i.e., $(\ell_{ij} - r_{ij}) \sim \text{Normal}(0, 1)$.

Then voter i votes for the Democrat (candidate L) in office j if

$$\begin{aligned}
\Pr(y_{ij} = \text{L}) &= \Pr\left(U_i(\mathbf{x}_j^{\text{L}}, v_j^{\text{L}}) > U_i(\mathbf{x}_j^{\text{R}}, v_j^{\text{R}})\right) \\
&= \Pr\left(r_{ij} - \ell_{ij} < \left(-\left(\mathbf{x}_i^* - \mathbf{x}_j^{\text{R}}\right)^2 + \theta_i v_j^{\text{R}}\right) - \left(-\left(\mathbf{x}_i^* - \mathbf{x}_j^{\text{L}}\right)^2 + \theta_i v_j^{\text{L}}\right)\right) \\
&= \Phi\left(2(\mathbf{x}_j^{\text{L}} - \mathbf{x}_j^{\text{R}})'\left(\mathbf{x}_i^* - \frac{(\mathbf{x}_j^{\text{L}} + \mathbf{x}_j^{\text{R}})}{2}\right) + \theta_i \cdot (v_j^{\text{L}} - v_j^{\text{R}})\right)
\end{aligned}$$

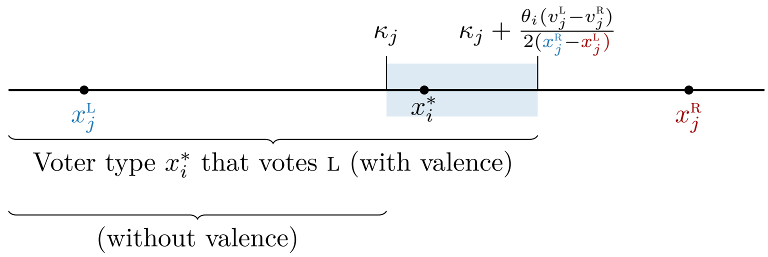
where $\Phi(\cdot)$ is the cumulative density function of a standard Normal distribution. In other words, a voter's choice for a particular election depends on the spatial differential $\Delta\mathbf{x}_j \equiv x_j^{\text{L}} - x_j^{\text{R}}$ combined with the cutpoint $\left(\kappa_j \equiv \frac{(\mathbf{x}_j^{\text{L}} + \mathbf{x}_j^{\text{R}})}{2}\right)$, a valence differential $\Delta v_j \equiv v_j^{\text{L}} - v_j^{\text{R}}$:

$$\Pr(y_{ij} = \text{L}) = \Phi\left(2(\Delta\mathbf{x}_j)'\left(\mathbf{x}_i^* - \kappa_j\right) + \theta_i \cdot \Delta v_j\right) \tag{2}$$

Valence differentials can therefore lead voters to split their ticket, which in this model corresponds to choosing a candidate that would not have been chosen based on the spatial cutpoint κ_j. The impact of valence is determined by its size relative to the spatial differential. To see this, consider the one-dimensional case and solve for $\Pr(y_{ij} = \text{L}) > 0.5$. Rearranging terms and assuming $x_j^{\text{L}} < x_j^{\text{R}}$ without loss of generality,

$$\begin{aligned}
\text{Vote for L if: } & 2(x_j^{\text{L}} - x_j^{\text{R}})\left(x_i^* - \kappa_j\right) + \theta_i(v_j^{\text{L}} - v_j^{\text{R}}) > 0 \\
\Rightarrow & x_i^* < \kappa_j + \frac{\theta_i(v_j^{\text{L}} - v_j^{\text{R}})}{2(x_j^{\text{R}} - x_j^{\text{L}})}
\end{aligned} \tag{3}$$

The intuition for this result is that considering valence moves the cutpoint in L's favor from the original (κ_j) by an additive factor that is increasing in θ_i and $(v_j^{\text{L}} - v_j^{\text{R}})$, and decreasing in the positive difference $(x_j^{\text{R}} - x_j^{\text{L}})$. The figure below sketches out an example where the voter chooses L only because of L's valence advantage, with the contribution of valence highlighted in blue.



Valence as Clarity of in Ideal Points We can also consider how uncertainty factors into this decision by treating the candidate position $\mathbf{x}_j^a$ as a random variable. For simplicity consider the one-dimensional case and let

$$\mathbb{E}(x_j^a) = \mu_j^a, \quad \mathrm{Var}(x_j^a) = \eta_j^a > 0.$$

Then the voter's expected utility from party a, $\mathbb{E}(U_i(x_j^a, y_j^a))$, becomes

$$\begin{aligned}
&- \mathbb{E}\left(x_j^a - x_i^*\right)^2 + \theta v_j^a \\
&= -\mathbb{E}\left(x_j^a - \mu_j^a + \mu_j^a - x_i^*\right)^2 + \theta v_j^a \\
&= -\underbrace{\mathbb{E}(x_j^a - \mu_j^a)^2}_{=\,\eta_j^a} - \mathbb{E}(\mu_j^a - x_i^*)^2 + 2\mathbb{E}((x_j^a - \mu_j^a)(\mu_j^a - x_i^*)) + \theta v_j^a \\
&= -\eta_j^a - \mathbb{E}(\mu_j^a - x_i^*)^2 + 2\left(\underbrace{\mathrm{Cov}\big((x_j^a - \mu_j^a), (\mu_j^a - x_i^*)\big)}_{=\,0} + \underbrace{\mathbb{E}(x_j^a - \mu_j^a)}_{=\,0}\mathbb{E}(\mu_j^a - x_i^*))\right) + \theta v_j^a \\
&= -\underbrace{\mathbb{E}(\mu_j^a - x_i^*)^2}_{\text{Spatial}} + \underbrace{\theta v_j^a - \eta_j^a}_{\text{New Valence}}
\end{aligned}$$

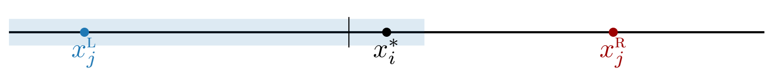
Therefore, a large variance η_j^a can be interpreted as having the opposing effect as high valence θv_j^a. In other words, high uncertainty in policy position effectively lowers valence.

Connection with Ticket Splitting and Party Switching Thus far, I have limited the exposition to a case where a voter makes one office, rather than voting on the long ballot. The theoretical results make the extension straightforward if we fix the national, top of the ticket office to be polarized and contested by candidates equally matched in valence. In essence, vote choice is a function of the relative difference between *candidates*, not offices. With a simialr logic, we can capture party switching overtime in this model by considering offices at time 0 and time 1 as two choices the same voter makes.

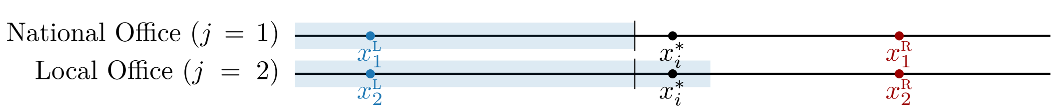
In the figure below, we show three cases: the first case repeats the single-office above, which can be thought of as a local office. In the second case, we introduce a "national" office in $j = 1$ where candidates are equally separated and assumed to have equal valence.

In the ticket splitting case, the voter's ideology x_i^* remains the same across offices, because the voter is voting for the two offices in the same ballot. As before, the tick-mark is the cutpoint between candidates (κ) and the blue region indicates the type $\left[0, \frac{\theta(v_j^{\mathrm{L}} - v_j^{\mathrm{R}})}{2(x_j^{\mathrm{R}} - x_j^{\mathrm{L}})}\right]$ which is all the possible values of x_i^* under which he will vote for the Left candidate.

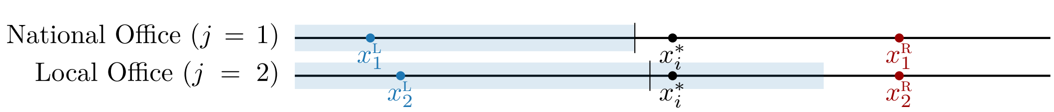
1. Crossing the party line: polarized parties with Democrat valence advantage



2. Ticket Splitting: Democrat in local office has valence advantage



3. Ticket Splitting: In addition to 2, local Democrat is more moderate



The figure reinforces the point that ticket splitting for a candidate a is increasing in the valence advantage of candidate a and decreasing in the degree of moderation (or convergence) between the two candidates.

To model party switching, we can replace the diagrams above with offices at two different time periods. An additional moving piece in the party switching model is that the voter's ideal point itself may shift over time. This adds an additional degree of complication, but still can be thought of in the same framework.

Candidates who run in local offices may be different from national elections because there is less information about them in the media. Recall that uncertainty in the information about candidate b's position also effectively enters the decision rule as negative valence for candidate b. This implies that candidate a being more well-known (perhaps through news coverage, advertising, incumbency) than candidate b in a local office also draws ticket splitting towards candidate a. All this can happen even if voters cared equally about party across offices (θ is constant across offices) and *candidate's* positions were nationalized ($x_j^{\text{L}} = x_j^{\text{L}}$ for all j, and the same for R).

Identification Although I do not attempt to estimate the parameters of the model from the data in this dissertation, it is worth noting how that would be possible. On its own, this model is not identified; an additional constraint is needed to distinguish between the contribution of the valence differential and the spatial differential in the data. To see why, following Bailey and Maltzman (2008), consider adding and subtracting some arbitrary $\bar{\theta}\Delta y_j$ to the right-hand side of equation 2. Then we would be able to rewrite the equation as $\Phi\left(2(\Delta\mathbf{x}_j)'(\mathbf{z}_i - \tilde{\kappa}_j) + \tilde{\theta}(\Delta v_j)\right)$ where $\tilde{\kappa}_j = \kappa_j + \frac{\bar{\theta}\Delta v_j}{\beta_j}$ and $\tilde{\theta} = (\theta_i - \bar{\theta})$, making it indistinguishable from the original equation despite having a different coefficient on Δv_j. Bailey and Maltzman get around this problem by adding Members of Congress to the dataset and assuming that their equivalent of θ is 0 for those individuals. In my case, I can take the voters who choose the party lever (and therefore do not consider candidate-specific valence) and set $\theta_i = 0$ for those individuals.

Operationalization

We cannot observe a voter's utility for each alternative, let alone the specific component of the utility that is about the party bundle. At best, we only observe the revealed preference of the voter. Scholars have used two measures in particular: vote switching and ticket splitting. Both measures involve the same voter voting for both parties, either across time, across office, or both.

Party Switching: Surveys and panels of respondents can ask respondents to recall their party choice for two different elections, often spaced across time. This corresponds to the common definition of swing districts. V.O. Key's seminal study of electoral change focuses on this operationalization as well. Through analyzing a series of national surveys Key (1966) showed how voters transitioned between the Republican and Democratic presidential candidates during 1940 - 1960, finding that as much as 15 to 20 percent of voters in this era switched parties to align their vote with their policy views on key issues of the day. This is both an observable implication and a operationalization of being a swing voter. If the same person switches parties, that suggests voters derive similar utilities from both parties generically.

Ticket Splitting: The pair of elections to be compared can also be across offices within the same election. The American federalist system abounds with direct elections of various elections. The majority of these are partisan elections (i.e. candidates have party labels on the ballot) in the same general election cycle. Beck et al. (1992) examine ticket splitting in state executive offices and Burden and Kimball (2002) examine ticket splitting in the US House election. But as noted by Burden and Helmke (2009), a general definition of ticket splitting applies to what others will call vote switching.

Persuadability is another operationalization of the swing vote that conforms with one way the term is used colloquially. In this definition, a voter is a swing voter if they are one of the first voters to switch their vote under the counterfactual that some factor, like candidate valence, candidate ideology, or national information shocks, changed. Unfortunately, this counterfactual quantity is difficult to measure reliably. Work in estimate heterogeneous treatment effects have been applied successfully in turnout (Imai and Strauss 2011), but less in candidate vote choice, perhaps because effects are noisy and experimental samples for vote choice are smaller than turnout experiments. Coppock, Hill, and Vavreck (2020) report a large collection of survey experiments to examine heterogeneity and subgroup effects, and report little heterogeneity in persuadability. However, their settings focus on the 2016 Presidential election and therefore do not test variation in candidate attributes. Other machine learning methods may be able to reliably identify subgroups with high treatment effects, but this is an evolving methodological field (Künzel et al. 2019) and beyond the scope of most of the evidence presented in this dissertation.

1 | The Swing Voter Paradox and the Emergence of Two-Party Competition, 1950 – 2020

Abstract

One interpretation of the nationalization and polarization of U.S. politics is that voters have become more loyal to a single party across multiple offices. I argue against this characterization by studying the trends in ticket splitting across a wider range of offices and years than previous work, and comparing two-party voteshares with the Presidential vote in the same constituency. Although this aggregate metric is not an exact measure of the proportion of voters who split their ticket, it is close to a lower bound. Therefore while average Congressional election in 2016 and 2020 differed from the Presidential voteshare by 2 to 3 percentage points, the proportion of ticket splitting this implies was large enough to have reversed party control in Congress. I also find that statewide executive and state legislative elections trend the same as Congress. However, gubernatorial and countywide elections do not show the same trend, or have larger discrepancies from the Presidential vote. This suggests that the swing voter is not a single bloc, but varies by the office and candidate.

* I thank Jim Snyder, Gary Jacobson, Carl Klarner, Steven Rogers, Michael Zoorob, Chris Warshaw, and Justin de Benedictis-Kessner for sharing their electoral data, most of it unpublished or before public release. I also thank those at Daily Kos for their values of Presidential voteshare at legislative districts that enable many of these comparisons.

All states in the 2016 U.S. Senate Election produced "straight-ticket" delegations with Presidential results. In all states where a Republican candidate won the U.S. Senate race, Donald Trump, the Republican, won. And in all states where a Democratic candidate won the U.S. Senate race, Hilary Clinton, the Democrat won. The 2020 elections were similar. Except for Senator Susan Collins (R-ME) winning re-election, no state with a U.S. Senate election that year delivered a split delegation between President and Senate. Earlier, Jacobson (2015) noted that "[w]ith little fanfare, the electoral advantage enjoyed by U.S. representatives has fallen over the past several elections to levels not seen since the 1950s," a pattern of a declining personal vote that would be consistent with congressional election results increasingly mirroring Presidential support.

Observers took this as evidence that ticket splitting was less consequential, that allegiance to national parties now dominated electoral behavior, and differences across candidates and offices were increasingly minor (Stein 2016; Enten 2016). This became a pattern for the next few elections in 2018 and 2020 (Skelley 2018; Rakich and Best 2020). Moreover, the decline in ticket splitting has co-occurred with the increase in polarization between parties' voting behavior in Congress. Putting these trends together, one might expect highly disciplined political parties with extensive degrees of partisanship found among voters, cutting across different levels of government (Drutman 2018).

In this chapter I document and interrogate these overtime trends from the lens of the swing voter. Disentangling the contribution of parties, candidates, and voters from a set of election results is a source of numerous debates in political science (Fiorina and Abrams 2008; Ansolabehere, Snyder, and Stewart 2001). Large electoral trends that decide which parties and politicians win elections and set policies should be traceable to specific blocs of voters. But simple statistics often mask this interpretation. For example, the statistic that "no states split their Senate-presidential vote for the first time ever" (Stein 2016) is only loosely indicative of the prevalence of ticket splittings in the electorate because it only accounts for who won an election – masking the margins.

Another potential pitfall when inferring voting behavior from these election returns in Congress is that it generalizes across offices too easily. The United Stats features a "long ballot" (Key 1963), where offices ranging from President to Sheriff to County Council are up for re-election at once, often affiliated with a major party. Have ticket splitting rates in these state and local elections also declined? Because of the limited availability of state and local election data, only recently have scholars tracked rates across the long ballot.

Through my analysis, I argue for two modifications to this common interpretation of modern elections. First, while ticket splitting indeed has likely declined, this does not mean that ticket splitting has become electorally irrelevant. In fact, the modern era of nationalized Congressional elections is also an electoral politics where ticket splitting has even more disproportionate electoral influence to determine the party control over elections. I label this the *swing voter paradox* because one might think that as swing voters decline, electoral results become more predictable. The basic idea of this pattern is that ticket splitting has declined but parties have been equally balanced, and, as we see in Chapter 2, ticket splitting is *not* concentrated in any particular district. Unstable majorities have been highlighted in past work as well, but my focus on the low rates of ticket splitting does not hinge on the argument (as in Fiorina 2017) that a large portion of voters are moderate.

Second, I collect electoral data in offices other than Congress, and show varying degrees of decline. If swing voters are a single group that is indifferent to either party, one might expect to see ticket splitting for all offices declining across for state and local offices as well. I do find decline in ticket splitting rates in offices such as Governor elections and statewide legislative elections, but the drop is not as precipitous. Moreover, the evidence is inconclusive in county offices such as Sheriff and county council. There is value in comparing electoral results from different offices because it can start to disentangle a secular trend that applies to all voters, or particular aspects of the candidates and the districts they run in (for a similar design, see Ansolabehere and Snyder 2002). Under the spatial voting framework with valence considerations, I would expect that whether or not a voter casts a split ticket is not something inherent to the voter, but something that varies across offices and candidates.

The goal of this chapter is to be broad and represent electoral behavior with a common metric that can be compared across different offices and different decades. On the other hand, the simple measure has its flaws. I use the difference in a pair of vote shares to measure the prevalence of swing voters. This is a classic ecological inference problem. In the two-party case I focus on here, the difference in two vote shares almost certainly underestimates the proportion of voters who split their ticket in the population. Nevertheless, by using a simple measure allows for a more transparent comparison across different datasets. Moreover, because the direction of mismeasurement is largely known, the evidence is sufficient to make an important observation in support of the paradox I propose: Even by a conservative estimate, the proportion of ticket splitters in Congressional elections in the modern era has been sufficient to reverse the party control of both chambers of Congress.

1.1 Data and Measures

To provide a broad picture of ticket splitting, I collected and pooled together existing datasets to cover as wide a set of offices and years as possible. Table 1.1 summarizes the coverage and source of the data. I cover federal, state, and local elections, spanning a period of 70 years in some offices. Election results are public record but there exists no central repository for results going back multiple decades and covering state legislative and local elections. Another hurdle for my purposes is that to measure ticket splitting, I must compare the electoral results with another election, like that for President, in the same constituency. In fact, most of the electoral data I use in this chapter come from unpublished or forthcoming work generously provided by researchers studying these offices.

Results for Congressional elections are reported by Congress and standardized by James Snyder. Ticket splitting in U.S. House elections is more difficult to compute than the Senate because Presidential election results are often not reported by congressional districts, which often cross county lines in complex ways. Here I use figures of the Presidential vote by Gary Jacobson, who has pieced together the Presidential vote by congressional district. Statewide executive elections cover the offices of Governor, Lieutenant Governor (if elected separately), Attorney General, Secretary of State, Treasurer, and Auditor, and are provided by James Snyder, extending published results from Ansolabehere and Snyder (2002) and Eggers et al. (2015). State legislative elections are provided by Rogers (2021) and Klarner (2021), with Presidential vote at the state legislative district computed by Rogers or the organization *Daily Kos*. For county Sheriff elections, I rely on a dataset of historical Sheriff elections collected by Zoorob (2020). For county legislative elections, I rely on an ongoing data collection effort by de Benedictis Kessner and Warshaw, which collect results from municipal and county election results from 2000 to 2018, including those published in de Benedictis-Kessner and Warshaw (2020). I do not use municipal elections in this analysis because most of these offices are nominally non-partisan. That is, these candidates do not participate in a party primary to appear on the general election ballot, and party affiliation is not listed on the ballot.

Table 1.1: Overview of Election Data

Office	Years	Areas	Unique Elections	Primary Source
U.S. Senate	1950–2020	All 50 states	1,179	Snyder
U.S. House	1952–2020	All 50 states	11,921	Jacobson
Statewide Executive	1950–2020	All 50 states	3,883	Snyder
State Legislatures	2000–2018	47 states	29,848	Rogers and Klarner
County Sheriff	1960*–2018	Around 20 states	2,386	Zoorob (2020)
County Legislature†	1990–2018	595 counties	1,438	de Benedictis-Kessner and Warshaw

Note: Because the goal of this chapter is to compare each offices' vote share to the contemporaneous or nearest *Presidential* vote share in the same district, I show elections where such a comparison is available.

* Sheriff data in some states reach to 1960, but only in less than five states a year before 1998. I start my analysis in 2000 where there is more coverage of states.

† County Council data collected by de Benedictis-Kessner and Warshaw is part of an ongoing project and an extension of earlier work (2020). The 2020 data covers roughly 300 mid-to-large counties which together holds about 50 percent of the U.S. population. Here I only use at-large races to compare them to Presidential vote share.

For any election for office j in constituency i, I compute the simple vote share difference from the Presidential vote,

$$\text{Vote Share Difference}_i = \left| \frac{D_{ij}}{D_{ij} + R_{ij}} - \frac{D_{i,\text{President}}}{D_{i,\text{President}} + R_{i,\text{President}}} \right| \tag{1.1}$$

where D_{ij} indicates the number of votes for the Democratic candidate in constituency i for office j, and R_{ij} indicates the number of votes for the Republican candidate. The overall goal of this section is to provide good enough measures for as wide a historical range as possible.

Throughout this section, I use the Presidential vote as the reference office because the President - Vice President ticket is the only office elected by the entire county. It is also conveniently comparable: always contested by the two parties, giving each voter in every state the same two choices. Because a goal of the analysis is to compare different congressional, state, and local offices with each other, I fix the reference category in all these comparison of pairwise differences. There are exceptions to this general pattern: strong third party candidates in 1960, 1968, and 1992 make the two-party presidential vote share in those states harder to interpret.

Therefore, high values of the Vote Share Difference indicate that one candidate has outperformed expectations relative to how a national candidate in the Presidential race did in that same district. The lowest value of 0 indicates that the Democratic and Republicans got exactly the same two-party voteshare as their respective Presidential candidates. For an example of this measure used in the literature, see Darr, Hitt, and Dunaway (2018). Moskowitz (2020) shows that this measure correlates highly from actual ticket splitting rates estimated from ballots.[1]

Of course, even if it is correlated with the quantity of interest, the vote share difference measure suffers from an ecological inference problem in measuring the degree of ticket splitting or party switching. The three simplified cases in Figure 1.1 illustrate what the difference measure can and cannot measure. Case A and Case B in the Figure show the same 2 percentage point difference in vote share between the President and U.S. Senate, yet in truth Case B happens to have over 10 times more ticket splitters than Case A. In a simple case with only two options in each office, the difference in vote shares measures the *net* ticket splitting that the winning candidate receives. This can be seen in Case C, where the vote share difference is 0 because the 4 percent of ticket splitters cancel each other out in the difference of vote shares. If the two offices are elections at two different times, we must further assume at least that the proportion of ticket splitters are equal.

The subsequent chapters improve upon this vote share difference by using individual-

[1] For usage in the popular press, see *FiveThirtyEight*, "Split-Ticket Voting Hit a New Low in 2018 Senate and Governor Races" (November 19, 2018). https://perma.cc/9Y75-3J9R.

Case A

	Vote for President		
Senator	D	R	Sum
D	50	**2**	52
R	**0**	48	48
Sum	50	50	100

Voteshare difference: **2** (52 - 50)
Actual ticket splitters: **2**

Case B

	Vote for President		
Senator	D	R	Sum
D	40	**12**	52
R	**10**	38	48
Sum	50	50	100

Voteshare difference: **2** (52 - 50)
Actual ticket splitters: **22**

Case C

	Vote for President		
Senator	D	R	Sum
D	48	**2**	50
R	**2**	48	40
Sum	50	50	100

Voteshare difference: **0** (50 - 50)
Actual ticket splitters: **4**

Figure 1.1: Difference in Vote Shares as a Lower Bound for Ticket Splitting Rates

Note: The schematic illustrates, in a simple constituency of 100 voters, how the aggregate difference in vote shares captures the degree of ticket splitting or party switching. In our data, only the row and column sums are observed. The vote share difference from these pairs of races does not accurately track the actual (unobserved) number of ticket splitters, but the difference gives a lower bound of the number of ticket splitting.

level data such as surveys and cast vote records. The simple vote share is nevertheless useful for historical comparisons where survey data do not exist. Moreover, in the two-party case, the direction of the measurement error is known. Even this lower bound measure is sufficient to demonstrate the point that there are enough ticket splitting voters to reverse the party control in modern Senate and House elections.

1.2 The Rise and Decline of Ticket Splitting in Congress

Figure 1.2 shows the trend of vote share differences comparing each chamber of Congress with the Presidency. The overall trends confirm that ticket splitting and party switching, at least measured through our vote share difference measure, has been steadily declining since the 1970s, reaching a low of 2.3 percentage points in both the 2020 U.S. Senate and U.S. House elections.

Because in a midterm year there is no Presidential election and therefore no single pair of candidates that the nation as a whole votes on, I use the Presidential election

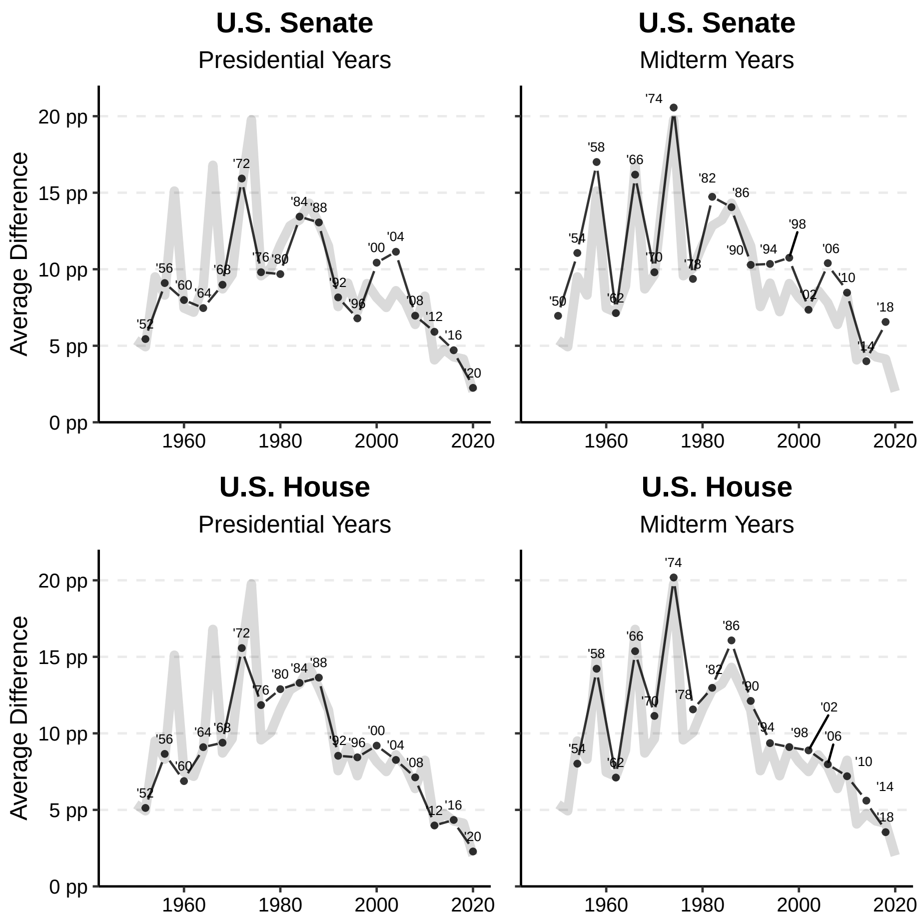


Figure 1.2: The Rise and Decline of Ticket Splitting in Congressional Elections, 1950 - 2020

Note: Each point represents the average absolute difference between two vote shares, in percentage points (pp): the Democratic vote share of a U.S. Senate candidate, and the Democratic vote share of the Presidential candidate, in a given state. Only states where a Republican and Democratic candidate contested the Senate seats are used.

Data: James M. Snyder Jr. (U.S. Senate, 1950–2016), Shiro Kuriwaki (U.S. Senate, 2018–2020), Gary Jacobson (U.S. House, 1952–2020) result two years prior in the same constituency. This introduces some complications in interpretation. Voters loyal to the President's party may stay at home in a midterm year, hurting that party's Democratic Senate candidate up for election rather than reflecting an increase in ticket splitters in the population. That is at least partially why the average difference is higher in midterm elections than presidential elections.

For this reason, I separate the midterm elections into a different panel. Midterm elections also show a steady decline in the past three to four decades. The difference in the two panels is at least not driven by different states being in different panels: A Senate seat is up every six years so all states cycle through having a Senate election in a Presidential year or midterm year.

Why did ticket splitting reach its height in the 1970s and drop precipitously thereafter? There is extensive literature on the causes of realignment and nationalization, so I highlight only the key points that I can test with my data. In my spatial voting framework ticket splitting is a function of both (a) the candidate's spatial positioning and valence, as well as (b) the voter's spatial position and the relative weight they value valence attributes. One possibility is that voters became more moderate in the 1970s and then increasingly divided into staunch partisan camps thereafter. If the story was entirely voter-driven, we should see patterns of moderates or independents rise and fall with similar patterns as the trends in ticket splitting.

However, trends in the independents are not large enough to explain the substantial shifts in ticket splitting. Figure 1.3 presents the time trends from well-known national surveys on the proportion of those who identify as some form of Independent. The gray line shows estimates that Pew Research Center (2015) estimated from their own data and historic Gallup polls. Most of these independents still vote consistently for a single parties' candidates, at least in the high profile national and Congressional races surveys can measure, and identify as Independents mainly due to their distaste

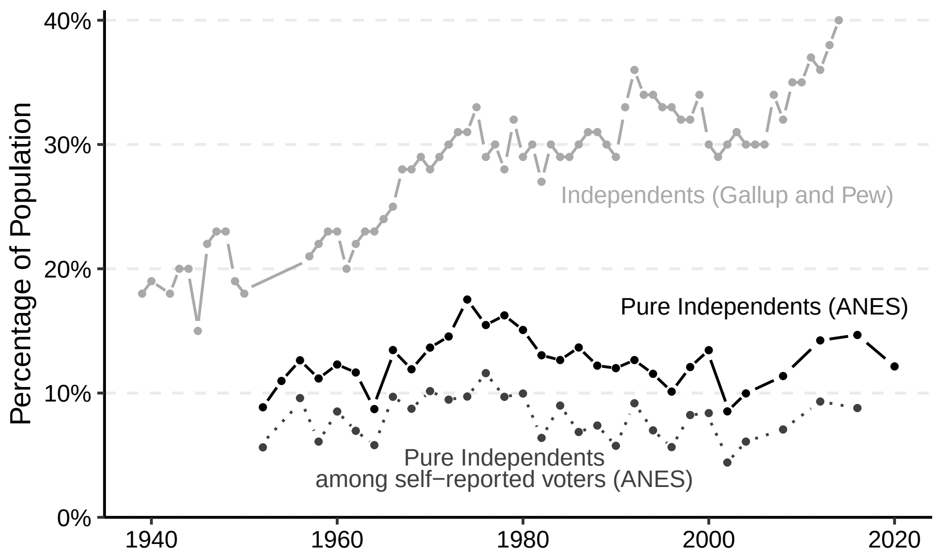


Figure 1.3: Voters Identifying as Independents, 1940–2020

Note: Each line represents different ways to measure Independents. The gray line is a wider measure that includes Independents who lean towards a particular party. The black lines estimate the proportion of "pure" Independents who do not lean towards a particular party. The dotted line restricts the sample to the turnout electorate. The trends do not correspond to the rise and fall of ticket splitting in Congressional Elections.

Source: Pew Research Center (2015), ANES Cumulative File (1948-2016), using the variable `VCF0305`. For 2020, I use the preliminary release of the time series dataset.

of traditional parties (Klar and Krupnikov 2016). A better measure that isolates the indifference to parties on the voter-side may be the Independents that do not lean towards a particular party, often called "Pure Independents." I use the American National Election Study (ANES) cumulative file to show the proportion of such Pure Independents in the general adult population, I also show this estimate among those who turn out to vote as measured by self-report, in order to form a better comparison with election results. The proportion of pure independents hovers around 10 to 15 percent in the modern era, and there is no consistent trend since the 1970s. It is therefore unlikely that voter's partisanship alone explains the change in historic patterns.

A more likely possibility includes changes in candidate-related attributes, such as the ideological positioning of Democratic and Republican candidates and the incumbency advantage. Hopkins (2018) also argues that voter choices nationalized as party brands nationalization, and I reaffirm the general mechanism he lays out. Existing scholarship in Congress identifies two periods of transition. Schickler (2016) traces the party realignment over racial policy in the 1960s back to the New Deal era and the incorporation of the Congress of Industrial Organization (CIO) in the Democratic coalition. Schickler argues that much of the realignment that Lyndon Johnson represented in his passage of the Civil Rights Act had already been set in motion decades ago, with both Democratic voters and state party activists overwhelmingly supporting Civil Rights and Southern Democrats entering the conservative wing of the Republican party. It does not appear, however, that this rising fissure in the two parties materialized in Congressional elections at the same time. Although split delegations were apparent in Eisenhower's election, the divergence between Presidential and Congressional elections kept increasing past the 1960s.

The next important time period in the history of the two parties was the 1980 election. Lee (2016) traces the polarization of parties to this time period and specifically this election, where Republicans unexpectedly won a majority of the U.S. Senate. At this time, the Democrats still appeared to have a solid grasp of the U.S. House. But once the control of a chamber was in play, Lee argues, party leaders had additional incentive to differentiate their platforms from the other party. Congressional Republicans therefore increased their use of messaging bills and Congressional hardball instead of emphasizing consensus building. The strategy also required the entire caucus to stick to the party line. The legislative dynamics that Lee traces fits the trends in electoral results rather well. As Congressional candidates repositioned themselves to align more closely with the President's party, the spatial voting framework would predict that voters will accordingly be less likely to split their ticket.

Similar trends show up in the rise and fall of the incumbency advantage in the U.S. House. Various estimates suggest that by the 1980s, incumbents gained an additional 8-10 percentage points above and beyond the normal party allegiance (Levitt and Wolfram 1997), that this increase was similar in state executive and state legislative office as well (Ansolabehere and Snyder 2002), and the advantage slowly declined thereafter. The incumbency advantage is closely tied to ticket splitting because it often involves some voters defecting from their general party identification to vote for the incumbent, which happens to be of a different party (Jacobson 2015). It is therefore hard to say whether the declining incumbency is a *cause* of ticket splitting. A declining incumbency advantage is often a manifestation of an increase in ticket splitting, which can be seen both in overall trends, the theoretical model outlined in the introduction of this appendix, and the results in the rest of this dissertation.

Other work points to the realignment of the former Confederate states, increased centralization of power to the federal government, and the nationalization of the media as contributing factors to the decline in ticket splitting and the nationalization of politics (Hopkins 2018). Testing each of these factors is beyond the scope of this chapter on electoral politics. From a broad view of Congressional elections reviewed here so far, a change in party brands appears to be an important factor. This trend is not exclusively driven by Southern states. Although the South underwent drastic realignment through the 1960s, voting for the Republican party in presidential elections but continuing to re-elect conservative Democrats to Congress, the overall trends hold in other states as well. I discuss these findings in the last analyses of this chapter. Realignment took place in Congressional elections across the country. California and New England Congressional delegations becoming more Democrat as the South realigned to the conservative Republican party.

1.3 The Closeness of Elections

For the rest of the analysis, it would be useful to set some benchmarks about how much of a swing vote is large enough to be politically important. I propose one benchmark: the magnitude of uniform swing it takes to reverse the party that wins the most seats in the legislature.

That is, the "minimum swing", $\underline{\delta}$ for a given U.S. House election is given by

$$\underline{\delta}^H = \min_{\delta<0} |\delta| \quad \text{such that} \quad \sum_{s=1}^{435} \mathbb{1}\left(V_s + \delta < 0.5\right) < 218 \tag{1.2}$$

where $s \in \{1, ..., S\}$ indexes seats, V_s indicates the two-party vote share in seat s of the party that ultimately won the majority of seats in the election. Therefore, a $\underline{\delta} = -0.01$ indicates that a 1 percentage point uniform swing against the majority party would have cost the winning party enough seats to cost them the majority. Another way to think of this value is the majority party's lead in the tipping point district.

Only a third of U.S. Senate seats up for election in a general election, so the formula for swing in the Senate takes into account the number of seats that the winning party already has locked in and not defending. The seats required for a majority also depend on the party of the Vice President, who casts a tie-breaking vote. Because every marginal seat matters for the majority, independents are coded as essentially belonging to the party they caucus for.[2] Uncontested elections and seats in California and Washington where two candidates of the same party win the primary are coded as safe seats for that party.

This measure of swing has several advantages over existing measures of competitiveness that are easier to find. Typically, analysts use a party's seat share or the pop-

[2] Joe Lieberman (CT), Bernie Sanders (VT), Angus King (ME), Harry Byrd Jr. (VA), Wayne Morse (OR), are coded as Democrats, and James Buckley (NY) as a Republican.

ular vote to track how close a party is from capturing a majority. But using seat share may mask the heterogeneity in the vote share of pivotal seats. And the popular vote is ambiguous because an electoral system's swing ratio varies by context. The ratio has ranged from 2 to 3 in the modern era (Linzer 2012), meaning that a 1 percent increase in the popular vote can translate to roughly a 2 to 3 point increase in the proportion of seats. In this application, we care about the seat swings directly. Moreover, the presence of some uncontested seats and seats contested by two candidates of the same party makes the computation of the popular vote ambiguous.

Figure 1.4 shows the values by year. Each point represents a general election year and labeled by the year, colored by the party that ultimately won the majority in the chamber.

The Congress of the late 1950s, 1960s, and 1970s were marked by Democratic dominance, and the Figure shows that their electoral dominance was far-reaching indeed. In the U.S. House, this trend seems to have grown till the 1988 election. Even though the Democrat's won fewer seats in 1988 (260) compared to 1958 (283), their win was more comfortable than in 1958. The tipping point seat for the Democrats was won by 8.7 points, whereas it was won by a comfortable 12 points in 1988.

The distinctiveness of elections in the modern era is clear from the historical comparison. In the early 2000s to 2020, control of both chambers could have been reversed by less than a percentage point swing against the Democratic Party. One would need to go back to the 1950s to find a time when that small a swing would have been decisive for legislative control. The historical trends highlight two prior turning points where deepening Democratic control reversed – the Gingrich Revolution of 1994 which ushered in a two-decade stretch of near-Republican control of the U.S., and the Republican takeover of the Senate in 1980. Lee (2016) finds that this 1980 Republican victory was surprising even for Republican operatives, who then started to reorient

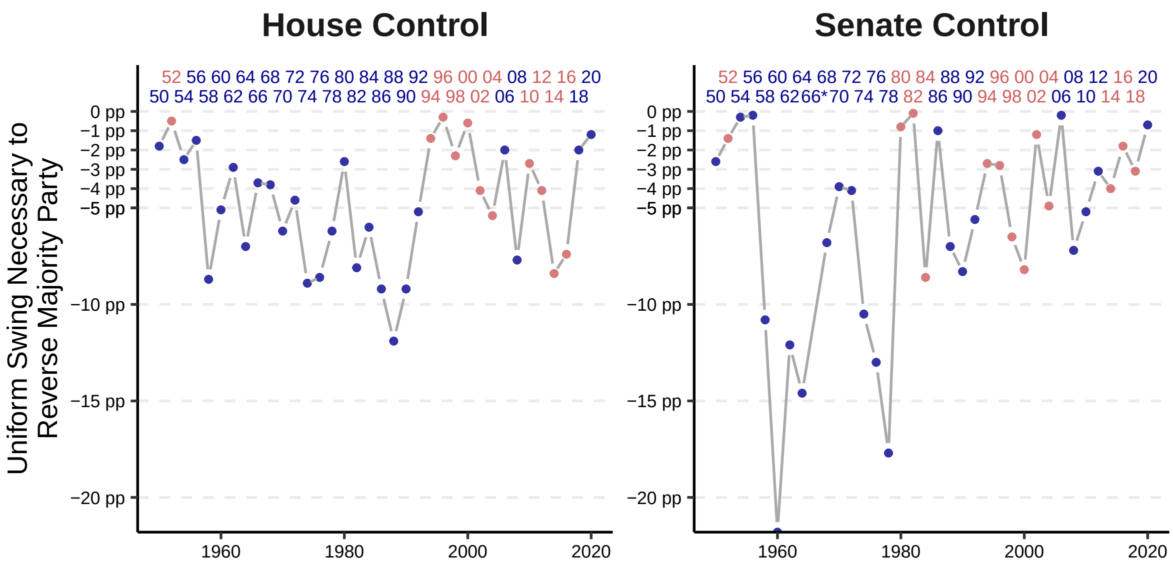


Figure 1.4: Increasing Competitiveness over Control of Congress.

Note: Each point represents the smallest uniform swing against the party that eventually won control of the chamber necessary to reverse the majority party. In 2020, for example, an 1 percentage point uniform swing against the Democratic party would have reversed their majorities in both the House and the Senate. In 1990, in contrast, it would have required a 9 point swing for the House and a 7 point swing in the Senate to reverse the Democratic majority. * In 1966, the Democratic Senate majority held 67 votes and only 20 were up for election. Therefore, no amount of swing would have been sufficient to vote Democrats out of their majority.

their party strategy towards distinguishing the party from the Democratic party.

Combining the findings on the prevalence of the swing voter and the closeness of party control produces Figure 1.5. This is the main relationship in the *swing voter paradox*: elections with fewer swing voters are the closest ones. The relationship is stronger for U.S. House elections, where all seats are up for election each year. Interpretation is more complex for the Senate, where roughly two thirds of the majority party seat share in any given year is actually not reflected in the ticket splitting estimates (which only use the year's 33 or so elections) represented on the horizontal axis. The slope coefficient in the U.S. House is 0.68 with a robust standard error of 0.14. This implies that every one percentage point decrease in the prevalence of swing vot-

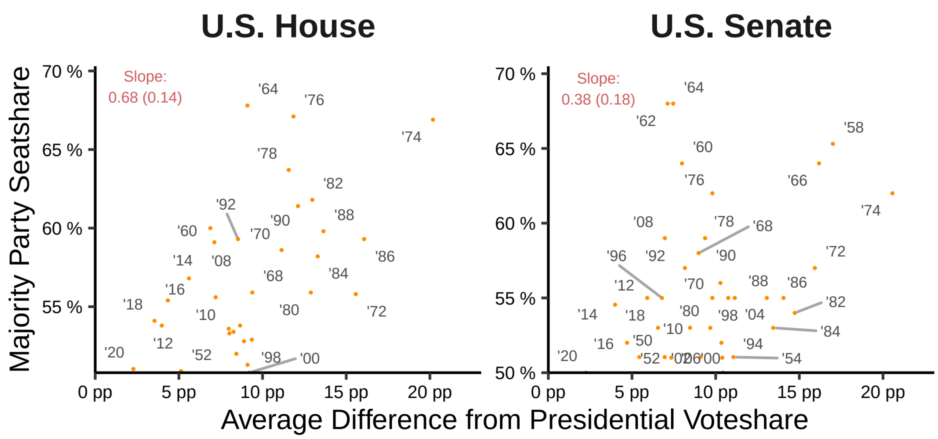


Figure 1.5: The Swing Voter Paradox

Note: The horizontal axis shows the values used in Figure 1.2, and tracks the degree of net ticket splitting from the President that year. The vertical axis shows the share of seats the majority party won in the election. Slope coefficients show coefficients and robust standard errors of the bivariate relationship. The two measures roughly correlate, especially for the House, where all seats are up for re-election. Somewhat paradoxically, the decline of ticket splitting and the increase in party polarization has occurred as any swing voter has become more likely to be pivotal.

ers is associated with a 0.68 percentage point decrease in the winning party's margin, thus making their party's victory more precarious.

These statistics of minimum swing provide useful context for subsequent analysis of the swing voters because it establishes in a concrete way how much their vote is electorally important. Swing voters that comprise three percent of the electorate would not be electorally important in the U.S. Senate election of 1966, where the Democratic party majority was so large that no amount of swing could have put them out of their majority. But the same three percent constituency *is* more decisive in a year like 2020, which could have made the difference between a Democratic trifecta and a Republican trifecta (The tipping point state in the 2020 Presidential election is either Wisconsin or Pennsylvania, where Biden won by less than a percentage point of the two-party vote).

1.4 Patterns in State and Local Offices

Almost all of past work on ticket splitting has focused on comparisons of Congressional races with the President. This is not sufficient evidence to attribute that the decline in ticket splitting is due to changes in the voter's preferences towards generic parties. This is why comparison across multiple levels of government is important. If the generic swing voter has declined, we would expect ticket splitting and party switching to decline across all offices similarly, because a single voter often votes for multiple partisan offices on the same ballot. On the other hand, if patterns differ by office even in the same state and same election, this suggests declines in ticket splitting may be particular to the office, and perhaps shaped by the ideological and valence candidates that run in those elections.

Data limitations have prevented even simple comparisons, however, below the statewide level of Governor. In this section, I walk through each set of offices outlined in Table 1.1. I retain the Presidential race as the reference category for the vote share difference.

Moving to state and local offices changes the interpretation of ticket splitting in one important respect, because these offices do not deliberate on the same legislation with the President. Because Congress and the President negotiate over the same legislation, it may be the voters prefer to balance their vote, ensuring that no party gains a trifecta (Fiorina 1992). Concerns about weak, unaccountable party government (American Political Science Association 1950) are also largely predicated on the common legislative setting. These concerns are not at play, for example, when a voter is voting for a Governor and a President. However, it is precisely because of this lack of a legislative connection that makes nationalization and party polarization of particular concern (Hopkins 2018). If voters cast their ballot for a single party regardless of the particularities of each level of government, this is one (but certainly not a sufficient) indication that partisan voting occurs without considerations of important qualifications.

Governors and Statewide Executives

Figure 1.6 starts by comparing the offices of statewide offices. Past research shows that compared to Senators, there are a few more Governors who come from opposite parties as their Presidential candidates (Sievert and McKee 2018). The data in this project allows a broad comparison of all executive offices.

Some care is necessary to provide a informative comparison of the rates found here with that of Congress, because election cycles vary. Statewide executive elections occur every four years in almost all states. About two thirds of these states hold elections in midterm years, about a third in Presidential years, and two states, New Jersey and Virginia, in odd years. Therefore, the difference between the year-by-year averages of Governor and Congress in Figure 1.2 may be due to the comparison of different states. To avoid this confounding, for each year I subset the Congressional statistics to the states which hold a Governor (or other statewide executive) election that year. Sometimes there is no Senate election in that year, so I use the U.S. House values as well. I generate a Congressional average baseline by upweighting the U.S. Senate values with the number of Congressional Districts the state has that year as a rough measure of population.

Figure 1.6 shows that statewide executive offices follow the same trend as Congress, but the office of Governor is more resilient in the modern era to the trend. In 2016, the average Democratic vote share in a Governor race (weighted roughly by a state's population size) differed from the state's Presidential election by 6 percentage points, while a Congressional race in those same states was about 3.5 percentage points. In

2018, with a different set of states with Governor elections in midterm years, the rates were 4.5 percentage points and 4 percentage points, respectively. This may in part be due to the fact that Governors in most states are elected in midterm years, therefore facing different electorates. Governors may also be able to free themselves from national party discipline as executives of their own states, dealing with a different set of issues.

The remaining statewide executive offices shown in Figure 1.6 include races for Attorneys General, Secretaries of State, Auditors, and State Treasurer. In the "Lottery of the Long Ballot," Key (1963) expressed concern that the separate elections of these offices would lead to state executive cabinets of different party members, and dampen accountability. At least in the modern era, these offices tend to align even closely with the national party lean of the state.

State Legislative Offices

Ticket splitting in the average state legislative election appears to have undergone a similar, if slightly modest, decline. Figure 1.7 shows the same average statistic in the years available by data provided by Steven Rogers.

Prior work on state legislative elections would lead us to predict that state legislative elections in the modern era should move in similar ways as the national offices. Rogers (2016) shows how partisan outcomes of the state house rise and fall along with the electoral results of the U.S. House, and that the electoral fortune of state legislators are shaped by in approval of the President and his party. On the other hand, other work focuses on how some state legislators may not be completely aligned with their party's positions in Congress, perhaps to adapt to their district's preferences (Shor and McCarty 2011; Erikson, Wright, and McIver 1993; Polborn and Snyder 2017).

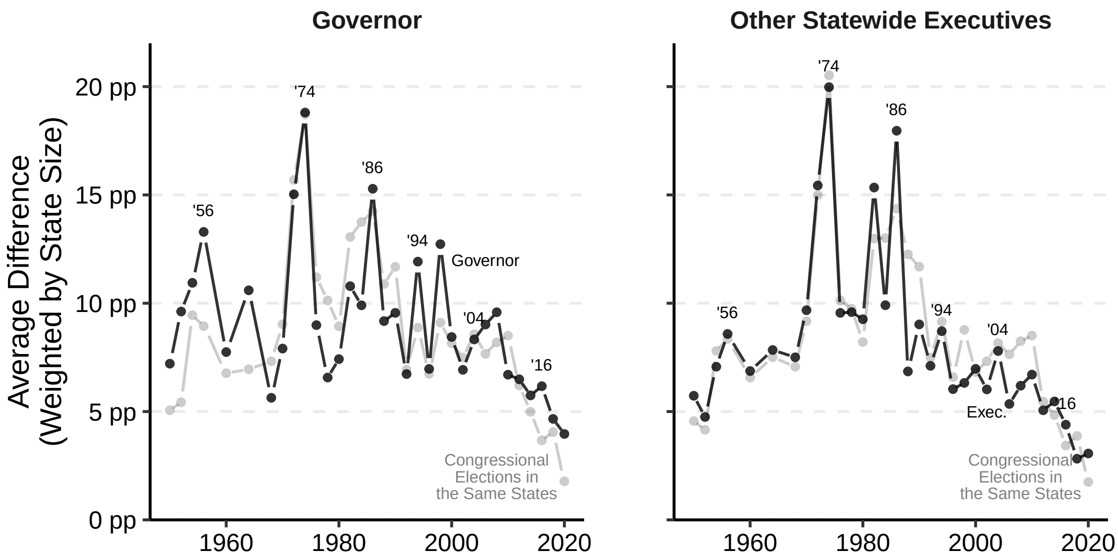


Figure 1.6: Difference from President: Governor and Other State Executive

Note: Points are shown in the same way as Figure 1.2 but separated by office. Because some states hold Governor elections in Presidential years and others hold them in Midterm Years, I recompute the Congressional average difference only among the states that have a Governor or statewide executive elections in that year and plot them as a comparison. Statewide executives average Lieutenant Governors (if separately elected), Attorneys General, Secretaries of State, Treasurers, and Auditors.

Source: James M. Snyder Jr. (for 1950–2016) and Shiro Kuriwaki (2018–2020)

Figure 1.7 shows that the average Democratic candidate for state legislature in a contested district had a voteshare that differed from the support for the Democratic Presidential party by around 5 to 10 percentage points. To compute this average, I took into account the population size of each district so that states with small districts are not overrepresented. Each New Hampshire state house district represents around 3,000 people. In contrast, every California state house district represents at least 450,000 residents. In computing a national average per year, I weight each district by the total number of votes cast in that state legislative election.

Like Congress, state legislative elections have become more similar to Presidential election results in the nearly two decades from 2000 to 2018. In the 2012 and 2016 elections, for example, the average estimate for ticket splitting from these data was only about a percentage point higher than for Congress. A direct comparison requires some caution because while the percentage points in the graphs are constructed to be comparable, the set of populations represented differ by office. First, state legislative elections do not happen in some years. To make a closer comparison, I use the same strategy as I used in the case for Governor and construct a Congressional reference average by taking only the same states that are used in each state legislative chamber and election year combination, upweighting U.S. Senate races by the number of districts. But state legislative offices are a harder case because they are not statewide races. Many state senate elections occur every four years, or only have a subset of the seats are up for re-election at any given cycle. Second, I also discard state legislative races that are uncontested, which I define following Rogers (2017) as an election where either the Republican or Democratic party wins less than 5 percent of the two party vote.

County Offices

The decline of ticket splitting is not nearly as apparent in local offices such as county councils and county sheriff offices, at least with the set of elections analyzed here. For our last set of results, Figure 1.8 shows differences for county council and county sheriff elections in Presidential years.

Unlike the prior analyses which covered most states, these analyze only a hundred or so counties each year and so are not readily comparable with the figures with other federal and state offices. A better comparison is to hold the particular set of counties constant by recomputing differences in congressional elections with the same set of counties. The dotted lines in Figure 1.8 show the average difference between the Presidential vote share and the U.S. House vote share in the same set of counties used in

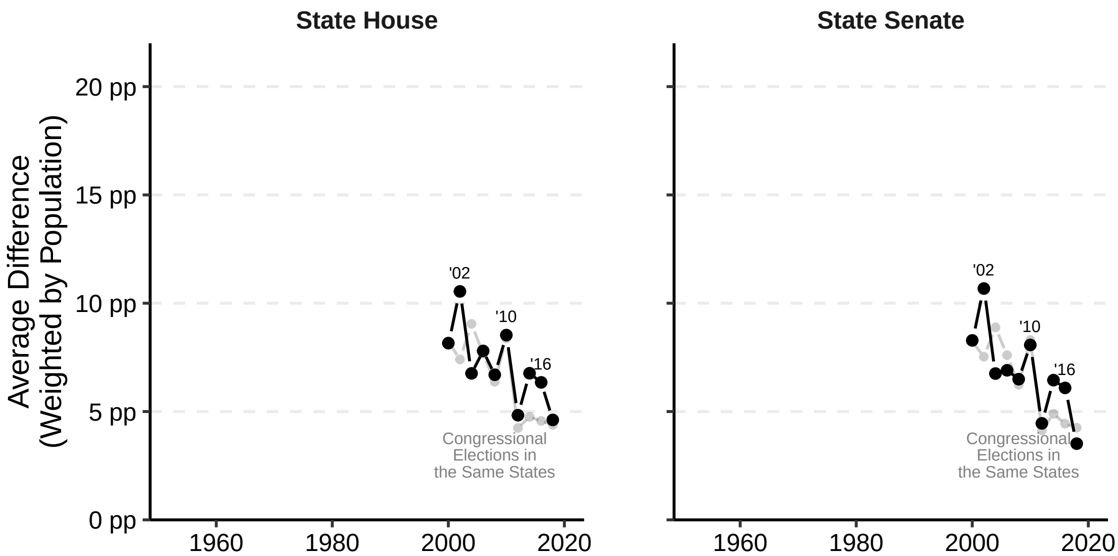


Figure 1.7: Difference from President: State Legislative Offices

Note: Following Figure 1.2, only contested elections are used each year. An average value is constructed by weighting each contested legislative race by the total number of votes cast in the race. As a comparison, I show in gray the average deviation of Congressional candidates from the Presidential race, only using Congressional races that are used at least once in the state legislative election data.

Data: Klarner (2021) and Rogers (2021), with Presidential voteshares in the state legislative districts originating from Rogers (2000), NCEC (2004, 2008), and Daily Kos (2012, 2016).

the county office average for that year.

The average contested, partisan Sheriff race during 2000–2016 saw vote shares that were on average around 15 percentage points different from the Presidential vote share in the same election, a remarkably large difference given the single-digit rate in Congressional elections both nationwide and in the same set of counties. The average difference in countywide county council elections hovered between 5 to 10 percent, showing no clear decline.

Caution is warranted when making the claim that the decline in ticket splitting has not materialized in state and local offices. The set of counties that have contested elections change every year and only a subset of those are shown here, so the lack of

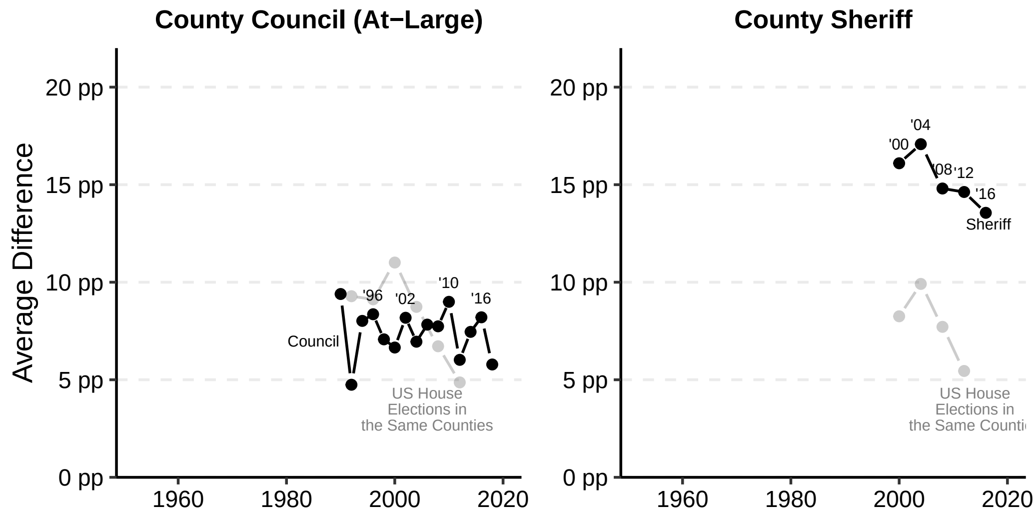


Figure 1.8: Difference from President: County Council and County Sheriff Elections

Note: Only countywide elections in the collected datasets are used and all comparisons are with the Democratic Presidential vote share. All races are fully contested with at least one Democratic candidate and one Republican candidate. For multi-member districts where more than one party can win, vote shares from the same party are summed. As a reference, I show the county-level average difference from the U.S. House election, in the *same counties* used in the local elections in that year.

Data: de Benedictis-Kessner and Warshaw (2020) and Zoorob (2020) (See Table 1.1)

a trend may be due to the particular set of counties we observe here. That said, the data covered here represent one of the best opportunities to date in examining this question. It opens the possibility that while voters are less likely to split their ticket in congressional races, the same partisan voter is willing to split their ticket in local offices on the same ballot, a proposition I test more thoroughly in Chapter 3.

1.5 Regional Differences

Regional and state-specific differences in these trends may also be an important factor, in addition to office, in explaining these trends. Mayhew (1986) showed varying levels of strength in party organization and machine politics across states. Strong party states may have been able to run and elect a party slate that differed in ideology from the party platforms in the Presidential elections. Regions such as the former Confederate South may also have been unique in resisting the convergence of state party platforms to national ones.

We can investigate the extent to which state and regional differences explain variation in ticket splitting through the analysis of statewide elections, where we have a long time series of over 70 years. Figure 1.9 shows the trends of each state separately. There are typically one to five statewide elections in each state and each general election. To reduce the variance from idiosyncratic races, I therefore take the simple average across statewide offices and across years within a decade. Each state - full decade observation, then, contains an average of 14 elections.

There are no clear regional differences in the statewide trends of the difference from the Presidential vote. Figure 1.9 shows the trends of all 50 states in light gray and highlights six illustrative states in color. All states show a general decline in the difference in voteshare starting from around the 1980s. The states that diverge from this general trend are in the South or appear to be rather idiosyncratic. Before the 1980s, voters in Southern states voted for Democratic statewide candidates at exceptionally higher rates than for Democratic Presidents; Mississippi and South Carolina are illustrative. But since the 1980s, voting behavior in these states aligned across offices. Then in the 1990s and 2010s, Southern states in fact ranked *lower* than other states in the degree of divergence. South Carolina is an important state for this dissertation because I focus on the state in Chapter 3 to compare state and local elections. The findings here suggest that South Carolina in the modern era is comparable to the rest of the nation when it comes to levels of ticket splitting. Massachusetts and California, two large states where a Republican base gave way to Democratic dominance

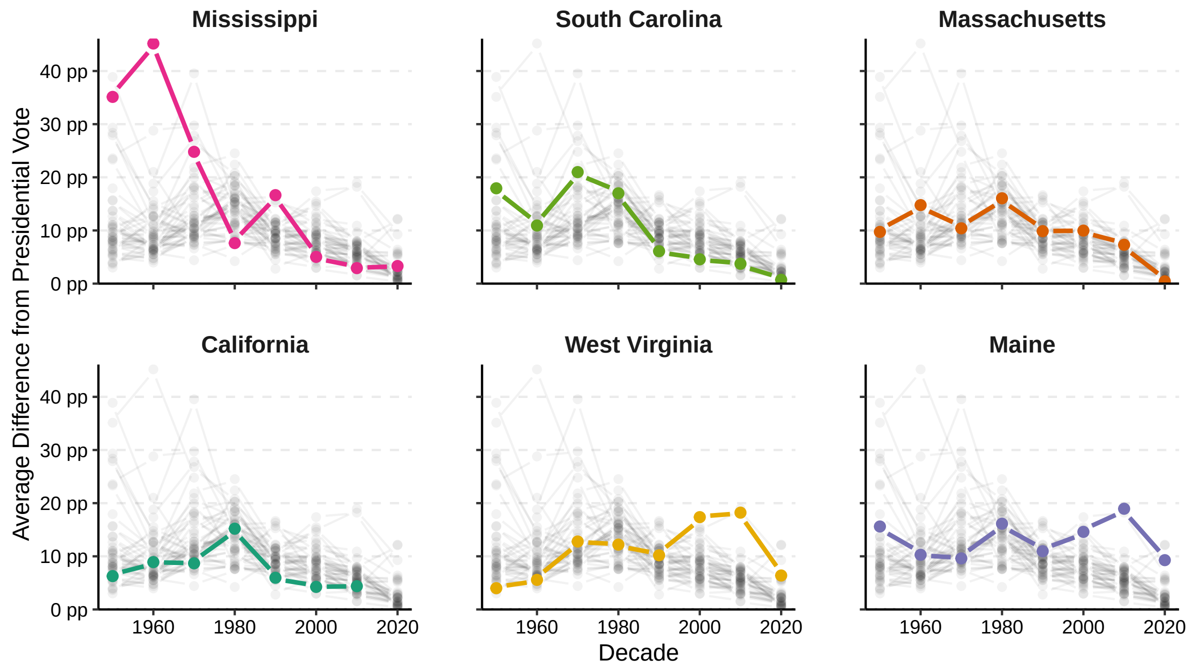


Figure 1.9: Trends in Ticket Splitting by State

Note: Gray lines show the Average Difference measure by state and decade, averaged across years and statewide offices (U.S. Senate, Governor, other Statewide Executive). Every value is then centered to the national average in that decade. Therefore, a zero indicates that the Average Difference in the state is right at the national average of the decade, higher values indicate a larger discrepancy in the state election than the national average, and lower values indicate a smaller discrepancy in the state election than the national average. Six states are highlighted for illustrative purposes.

over the course of these decades, show trends that are roughly in line with the aggregate trend.

There are only two states where trends in divergence have bucked the rest of the country: West Virginia and Maine. West Virginia started out as having a close correspondence between statewide and Presidential vote, but gradually diverged, with Democrats winning statewide in 2010 but with the same voters voting for Republican presidential candidates. The same pattern can be said for Maine, although independent candidates (which are excluded in the voteshare calculation) complicate this picture. Overall, almost all states appear to have experienced a trend for convergence to the Presidential vote, a story consistent with theories of nationalization.

The South in particular is a region that stands out in Figure 1.9 as well as in the scholarship on realignment (Hopkins 2018; Aldrich 2000). In Figure 1.10 I show more clearly the difference between the South and non-South with the same statewide data. Here I group the differences by every pair of elections that use the same Presidential election (for example, 2016 and 2018), and compute differences separately by office. I compute the difference by regressing an election's average difference on a binary variable separating Southern vs. non-Southern states and show the regression coefficient on the binary variable. I use two operationalizations of the South: one with core Southern states including those that so diverged from the national Democratic party that they ran third party Presidential candidates, and a more expansive definition including more former Confederate states listed in the figure.

The differences reinforce how the South drove the aggregate trends before the 1980s, but this is no longer the case in the modern era. As we saw with Mississippi and South Carolina, by the 2000s Southern states tended to have lower divergence than non-Southern states across all statewide offices. In other words, there appears to be more party loyalty in statewide Congressional and executive elections in the South than in the non-South.

1.6 Limitations

There are at least three limitations to the evidence I have provided so far in making inferences about the prevalence of swing voters, ticket splitters, and party switchers. First, electoral data at the local data is far from complete. Electoral data before 2000 is still rare, and only recently have scholars pursued the collection of state and local election data even in the modern era. Continued data collection, including through new methods for scraping election results from historical newspapers, is

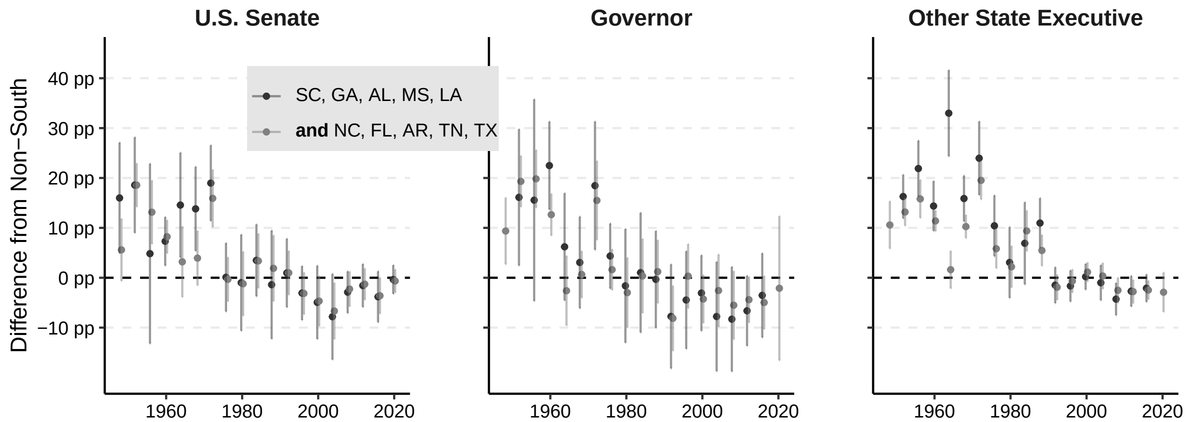


Figure 1.10: Difference between South and Non-South

Note: Each point shows the difference between the Average Difference measure between Southern states and Non-Southern states. It is the coefficient regressing the absolute difference by an indicator variable for the South, with 95 percent confidence interval. Positive values indicate Southern states had *higher* amounts of divergence between the Democratic votes in those offices from Presidential Democratic voteshare. For each year and office, I show two alternative definitions of the South: South Carolina, Georgia, Alabama, Mississippi, and Louisiana (in black lines), and adding to those five states, North Carolina, Florida, Arkansas, Tennessee, and Texas.

needed for a more representative understanding of how nationalization has affected local politics.

Another challenge in this area is one that cannot be solved with simply more data collection. Local elections for the same county council also vary in district magnitude, electoral timing, and party labels, all making it difficult to code and standardize candidate data. Local elections are more likely to be uncontested, so two-party vote shares are not observed. In practice, quantitatively summarizing measures of party support and ticket splitting in county elections in the same scale as U.S. Senate elections is largely out of reach.

A final class of limitations to the analysis is that the statistic used here, the absolute difference between two vote shares, is unsatisfying because it masks the true amount of ticket splitting among the electorate. Fortunately, this class of limitations is one that I overcome in the subsequent chapters. In Chapter 2 I use recent survey data that not only allows me to avoid the ecological inference problem, but also allows me to avoid the assumption of uniform swing. In Chapter 3 I use an untapped source of electoral data — cast vote records — which avoid the ecological inference problem *and* allows me to measure vote choice without error. Both findings advance the understanding of the electoral power of the swing voter as well as add more detail to the findings in this chapter, which instead provided relatively thin but historically and geographically wide-reaching coverage of the elected offices in the U.S. federal system.

1.7 Conclusion

In a well-known 1974 article, David Mayhew noticed that the close elections were on the decline in the past several elections. "Congressional Elections: The Case of the Vanishing Marginals," was consequential because it suggested the congressional elections were no longer competitive, which perhaps meant that holding incumbents accountable by voting them out of office was out of reach. This finding was valid at the time, where Mayhew examined U.S. House elections from 1956 to 1974. As I show in my uniform swing metric in Figure 1.4, the same party won a majority of seats almost with larger margins with each successive election.

But the trend of "vanishing marginals" appears to have gradually flattened out since 1980. As both parties became equally balanced, the control over Congress came well into play for both parties. Electoral considerations to stake out distinct party platforms intertwined with this trend and perhaps further facilitated the decline of ticket split voting in Congressional elections. Because the parties were equally balanced and no third parties emerged, modern electoral politics became one where parties have solidified their support across offices and yet have both parties have less of a grasp on the control over the power of lawmaking.

The main implication of this chapter is to establish this new equilibrium, which hews closely to the idealized notion of two-party government but is a break from most of the last past century. In this politics, fewer voters appear persuadable, but electoral results vary widely. This electoral landscape has several implications for the strategy of campaigns that seek to gain power, although many are not as clear-cut as they may appear at first. The dearth of swing voters might naturally imply that campaigns should focus instead of mobilizing their base to turn out. In a 50-50 election, every voter matters. But if the two parties have starkly polarized, partisans are less persuadable than the sliver of swing voters. It is also the case that persuading a swing voter to switch parties is twice as efficient for the electoral margin as persuading a core supporter to turn out to vote.

One reason it may not be obvious to characterize this regime as a two-party system is because, contrary to theories of spatial convergence in two-party elections, parties do not appear to have converged their platforms to the median voter. Recent research suggests that this is not a coincidence: minority parties have an incentive to distinguish themselves from the ruling party by exaggerating their differences, especially when their power is equally balanced (Lee 2016). These dynamics may in fact paint a normatively troubling picture of unproductive governance (Kustov et al. 2021), or what Drutman (2020) calls a "Two-Party Doom Loop" in which two polarized parties engage in a protracted back-and-forth of control without convergence.

Finally, this historical investigation across multiple offices has opened up new questions of its own. It refutes the idea that in this nationalized era, voters are either strong partisans regardless of the office or candidate. The modest trends in Governor and county-level elections open the possibility that being a swing voter is not a trait determined by individual partisanship alone, but by the choice setting and candidate. I directly address this question in the third chapter, providing one of the first estimates of the proportion of ticket splitters in state and local elections.

2 | The Fluid Swing Voter in U.S. House Elections

Abstract

I estimate individual and district level models of voters splitting their ticket or switching their party support between their Presidential vote and House support. Using a large survey dataset with samples in every Congressional District allows me to test two hypotheses from the previous chapters: that the swing vote can be explained by a spatial voting model with valence, that the degree of swing voters in each district is often large enough to flip the Congressional election results. I find that Independents, White Voters, and voters who do not frequently follow the news are more likely to be swing voters. In a panel regression of congressional districts, I also find that candidates who become incumbents net more votes through increases in ticket splitting from out-partisans. The district-specific levels of ticket splitters were enough in 2018 to reverse party control.

* I thank Jim Snyder, Santiago Olivella, Jeremy Bowles, the Imai Research Group, and participants at the American Politics Research Workshop for discussions on this paper.

2.1 Introduction

The first chapter of this dissertation showed a decline in the aggregate rate of ticket splitting from around the 1980. It also made the case that, even as the proportion of swing voters were declining, a switch in their vote choice was still likely to be pivotal, if note more so. By using a simple difference in aggregates for the measure of ticket splitters and using the assumption of uniform swing to measure pivotality, however, the chapter still did not definitively answer the seemingly straightforward question of how many swing voters there are in each legislative district.

How common are swing voters, why do they swing, and when are swing voters decisive? This chapter provides systematic answer to these questions in the U.S House. Conventional wisdom suggests swing voters are thought to have all but disappeared. My measurement approach does confirm that swing voters are a small portion of the electorate, about 3 to 5 percent. This level is within the margin of error in a typical survey and are therefore easy to gloss over or rount to zero. To avoid this, I combine extensive survey datasets with statistical adjustments for small area estimation, and give an accounting of why and to what extent some voters become swing voters. In particular, I study the US House and how candidate characteristics shape the degree to which voters cross party lines.

The search for swing voters poses a measurement challenge. Previous results have focused on national surveys (Hillygus and Shields 2009; Smidt 2017) or aggregate election results (Burden and Kimball 1998) largely because survey estimates of district-level quantities are unreliable without larger samples and statistical adjustments. Estimates of rare events and population also bring with it unique estimation concerns (King and Zeng 2001).

Published academic work and extensive data journalism using survey adjustments focus on the office of President or the generic congressional ballot, which do not allow inferences about what candidate characteristics affect vote choice (Cohn 2018; Cohn 2019b). But clearly, the policy positions, experience, and campaign strategies are often consequential in who wins elections (Canes-Wrone, Brady, and Cogan 2002; Hall 2015). Candidate visibility is in fact one of the key predictors of ticket splitting found in early work (Burden and Kimball 1998; Beck et al. 1992). Hall and Thompson (2018) is one exception, which tests the effect of US House candidate extremism on differential turnout at the congressional district level. However, the authors do not make small area estimation adjustments to their analysis of survey outcomes.

After reporting descriptive statistics from the measurement strategy, I conduct three sets of analyses to understand the drivers of the swing vote in the modern era. To estimate the causal effect of candidate quality — a key parameter in theoretic models — I employ a panel regression. Together, the evidence identifies a relatively small subset of the population of swing voters who respond to candidate quality, and may, especially in districts that are largely balanced in partisan strength, be decisive.

2.2 Models and Methods

The formalization of the concept of the swing voter in the Introduction provides justification for why ticket splitters and party switchers are reasonable approximations for the swing voter. We cannot measure latent preferences for a generic set of party candidates with survey or behavioral data, but voters who are indifferent are likely to split their ticket (within a single ballot) or switch their party (across two ballots). The spatial model also clarified how candidate level characteristics and move the cutpoint at which a voter splits their ticket. Therefore, an ideal setting to study this individual choice is a setting with sufficient survey data where there is large variation of candidates across districts and across time. U.S. House elections which are held every two years provide such a setting.

2.2.1 Survey Data and MRP Estimates

I use the Cooperative Congressional Election Study (CCES) in each even year from 2012 to 2018 for this study. The CCES is an online political survey with a sample size of about 50,000 to 60,000 respondents each even year.

I primarily measure the outcome of a swing voter by comparing the vote choice in two offices – the President and US House. Therefore, I effectively examine the joint distribution of vote choice in two concrete offices (rather than comparing against partisan identification) to measure ticket splitting explicitly. No two offices provide ideal coverage over years and geography, but I primarily use votes for the Presidency because it is the only nationally elected office, and the US House for its frequent elections every two years and the variation in districts and candidates. An individual respondent is considered a swing vote for the Democratic House candidate if he votes for a Republican presidential candidate in the concurrent or most recent persidential election, but votes for a Democratic candidate in the contest for US House. If the House election is uncontested, I drop the respondent from the analysis.

The CCES is a desirable dataset for several measurement reasons beyond its sample size. First, it includes indicators for the congressional district (CD) the voter is registered in, which is crucial for creating CD-level estimates. Second, its survey questions on the House vote present the full name of the specific US House candidate with their party affiliation, and measure vote choice before and after the election in two waves, which makes for a more reliable measure of House vote compared to a generic question. Third, the CCES validates each respondent's turnout by matching the personal information of each respondent to state voterfiles. Because about 20-30 percent of survey respondents misreport their turnout (Ansolabehere and Hersh 2012), I limit all my survey analysis to respondents whose turnout was validated. Of course, the measure of vote choice is still by self-report, and contains some degree of measurement error.

Even surveys as large as the CCES are prone to selection bias at smaller geographies, so I adjust CD-level estimates by Multilevel Regression Post-stratification (MRP) to reduce the mean square error of district specific samples (Warshaw and Rodden 2012). The CCES sample contains about 60 to 120 voting respondents for each congressional district each year. Both bias and variance are issues. The particular sample from the district may not be representative of the district. As I elaborate in Chapter 4, the CCES is not designed to be representative of particular districts and the weights only weight to state. And even if it were a conditionally unbiased sample, the estimator suffers from large variance. Some district samples, in fact, contain zero Trump to Democrat vote switchers. This is not surprising given that the expected size of this population is about 2.5 percent, but it is not plausible that there were actually zero voters in that entire district who were switchers.

The MRP specification used here is a standard one, with a notable addition being that I stratify on the incumbency status of each district's candidate on top of standard demographic predictors. First, I fit a logistic regression on the CCES regressing individual swing on voter demographics and district characteristics. All individual-level variables are categorical and modeled as random effects which induces partial pooling. In this way, coefficient estimates are a weighted average between the group-specific coefficient and a global estimate from pooled data (Si 2020). I model varying intercepts by district, age group and geographic division combinations, and education and geographic division combinations. Past work finds that stratifying by district-level variables improves MRP's predictive accuracy (Hanretty, Lauderdale, and Vivyan 2016). Therefore I include the district's presidential vote share, and, given my theoretical predictions, a binary variable for incumbency. I then create a MRP estimate for each district by taking the average of predicted values for a given demographic strata weighted by the joint distribution of age, gender, and education in each district supplied by each survey year's 5-year ACS. In Chapter 4, I propose a way to improve the model further to model the turnout electorate and poststratify on party registration.

2.3 Predictive Variables

Before estimating poststratified district estimates, we can first examine the estimated predictors of the outcome to characterize who swing voters are. Using the CCES sample of validated voters in contested districts, I estimate a simple logit model with standard demographic variables

```
swing ~  female + age + education + race + newsint + pid7
```

where `swing` is whether or not the CCES respondent splits their ticket or switches their vote, `newsint` is a common CCES variable that asks voters how often they follow the news, and `pid7` is a seven-part partisan self-identification variable. In this identification, voters are first asked if they identify as a Democrat, Republican, or Independent. Democrats and Republicans are further given the option of a "strong" or "not very strong" partisan, and Independents are further given the option of leaning Republican, leaning Democrat, or lean for no party.

Figure 2.1 converts the estimated coefficients to the predicted proportion of swing in the survey sample for interpretability. I estimated predicted probabilities (also called estimated marginal means) by selecting focal variables from the regression, while averaging across different levels of the other variables weighted by the empirical prevalence of those variables.

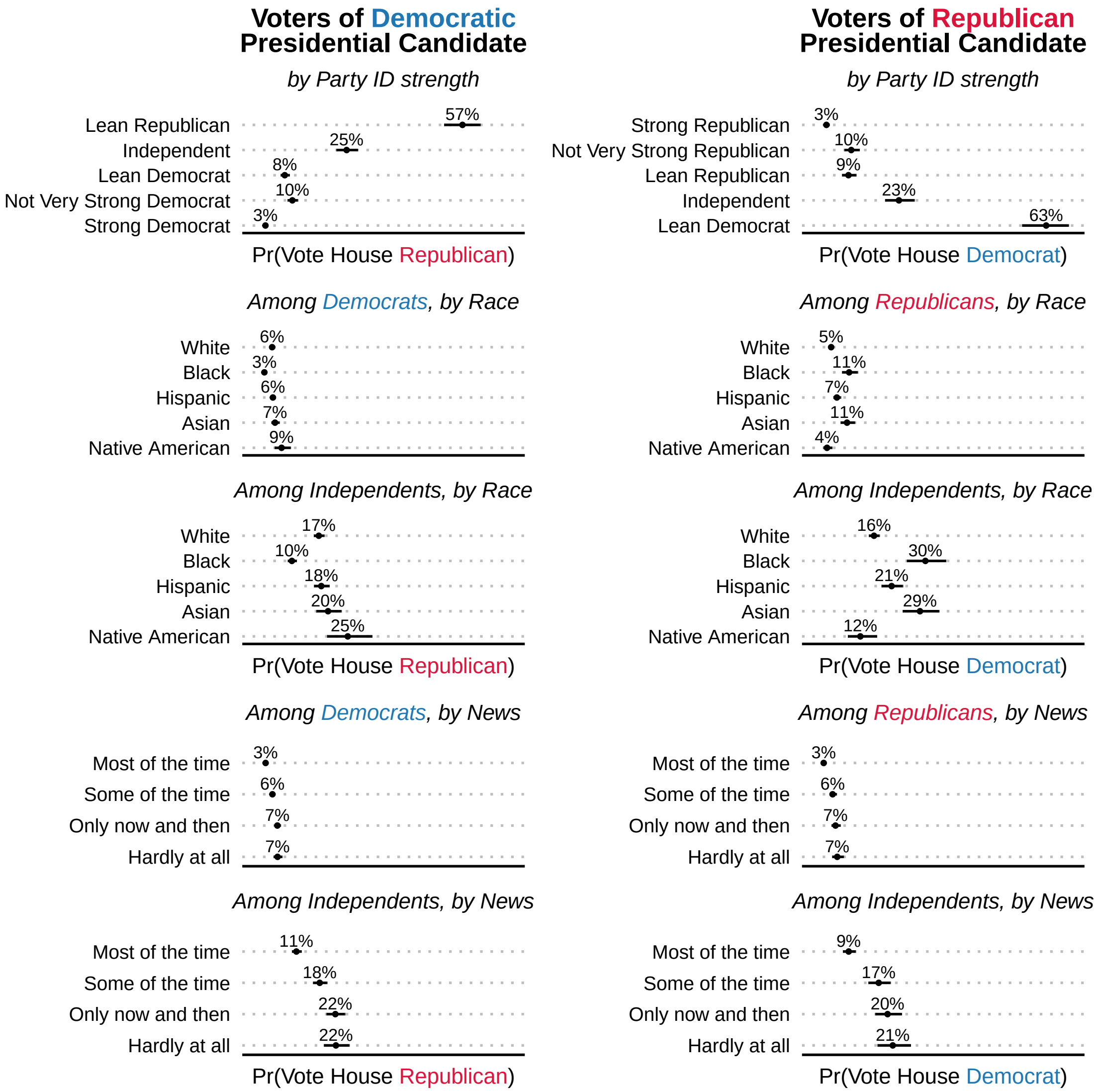


Figure 2.1: Voter Level Predictors of Split Ticket Voting between the President and U.S. House

Note: Each point is the estimated probability of ticket splitting to the House candidates, with 95 percent confidence intervals. All graphs on the left show the probability of voting for a Republican House candidate conditional on voting for a Democratic Presidential candidate and other conditions. All graphs on the right show the probability of voting for a Democratic House candidate conditional on voting for a Republican Presidential candidate.

Method: Probabilities are computed as marginal means: the focal variables in each graph (such as race) are held constant and the logit model generates predicted values. Non-focal variables are held at their levels and then a marginal mean is computed by averaging across the levels of the non-focal variables, weighted by the prevalence of the level in the data.

Data: Validated general election voters in the CCES, 2008 - 2018.

These patterns highlight four main predictors of being a swing voter. First, respondents who do not identify strongly with either party are more likely to swing, consistent with the spatial model. Second, whites are more likely to be swing voters compared to other racial groups, while Blacks are consistently straight ticket voters. This is consistent with the established finding that Blacks are more likely to be steadfast Democrats. Black voters who vote for Republican Presidential candidate are more likely to split their ticket for a Democratic House candidate, but these voters are relatively rare. Third, those who do *not* follow the news most of the time are more likely to be swing voters. Each of these differences by race and news interest constitutes a 2-5 percentage point difference, which is large enough to be decisive.

While the estimates provided here do control for variables that scholars have found to be relevant in ticket splitting, the estimates should not be taken as causal. For example, it could be that low interest in Congressional politics is correlated both with identifying as an independent, not having a college degree, and the propensity to switch parties. I focus on ruling out omitted confounders in the subsequent analyses focusing on candidate quality.

2.4 The Prevalence and Geographic Distribution of Swing Voters

To provide estimates for each Congressional district in each election of the prevalence of ticket splitters or party switchers requires additional modeling adjustments. Because there can often be only about 60 respondents in a given Congressional district and the national rate of ticket splitting can be in the low single digits, there is a decent probability that no ticket splitters are sampled into a Congressional district subset by chance alone. In such cases, no amount of weighting will move the point estimate of ticket splitting in that district from 0, even though there is likely a small per-

Table 2.1: Ticket Splitting and Party Switching at the Congressional Distrit Level

	Individual-Level			District-Level		
	Mean	SD	Obs.	Mean	SD	Obs.
Outcome: Vote Switch from						
Republican President to Democratic House	0.034	0.18	97,267	0.053	0.03	1,524
Democratic President to Republican House	0.033	0.18	97,577	0.054	0.03	1,524
District-Level Predictors						
Democratic Incumbent				0.286	0.45	1,524
Republican Incumbent				0.424	0.49	1,524
Presidential Republican vote share				0.469	0.15	1,524
Other Predictors						
Identifies as Independent on 3-point PID	0.270	0.44	109,182	0.285	0.04	1,524

Note: All district-level predictors are either population characteristics measured without error, or CCES survey estimates adjusted for small area estimation through MRP.

centage of ticket splitters in the population. To avoid this problem, survey researchers have used partial pooling regressions such as random effects to smooth out small sample idiosyncracies (Gelman and Little 1997; Ghitza and Gelman 2013). I estimate an outcome model similar to the one in the previous section, but I remove variables which cannot be post-stratified (such as news interest) and add district level variables such as past Presidential voteshare and the incumbency status of the House candidate that help partial pooling. I investigate the variability in these Congressional district level MRP models in Chapter 2.

Descriptive statistics in Table 2.1 show that about 6 percent of a general election electorate are swing voters under my definition: 3 percent who both vote for a Republican President and a Democratic House candidate, and 3 percent who vote for a Democratic President and Republican House candidate. For the average district, MRP estimates that about 10 percent of the consecutive electorate are swing voters. These numbers average across years, but each year typically has a national swing. In 2018, the national tide advantaged Democrats: 1.4 percent were Clinton voters who then switched to a Republican House candidate, and 2.6 percent were Trump voters who voted for a Democrat. Analysis of the turnout and vote by Ghitza (2019) suggest that in 2018, this vote switching was the critical piece that best explained the change in seat control. In 2014, more Obama voters defected to Republican House candidates than did Romney voters to Democratic House candidates. In addition, Ghitza (2019) identifies 2014 as an election year where turnout differentials also made a difference. The turnout differentials also advantaged Republicans, with more 2012 Obama voters not turning out to vote compared to 2012 Romney voters.

Each year's geographic distribution of swing voters is shown in Figure 2.2. These choropleth maps show both geographic and year-to-year variation in our main outcome measure of interest. In a given year, the range of swing voters is rather limited. For example, all districts had less than 5 percent of vote switching to Republicans in 2018. Table 2.2 lists the Members of Congress who represent the districts with the highest amount of swing in 2018. Some members of this list are expected – they include well-known moderates such as Dan Lipinski (IL-09, more conservative than 85% of his Democratic colleagues in the House by NOMINATE), Collin Peterson (MN-07, the third most conservative Democratic member), and Paul Cook (CA-08, more liberal than 80% of his Republican colleagues). However, some of the other members who draw large proportions of ticket splitters are at the extreme ends of their party.

The map, combined with the standard deviations presented in Table 2.1, shows that the cross-district variation in swing voters is not as high as one might think. In Table 2.1 the standard deviation in the district-level proportion of swing voters is about 0.03 for all three proxies, whereas the standard deviation for two-party vote share is five times larger. Although political observers classify districts and states as swing vs. safe districts, the difference between those types of districts is only on the

Table 2.2: Districts with Most Crossover Voters in 2018

CD	Trump to D	Place	Incumbent	Trump %
IL-03	12.7%	Southwestern Chicago (Chicago)	Lipinski, Dan (D)	40%
CA-27	12.5%	San Gabriel Valley (Pasadena)	Chu, Judy (D)	28%
CT-02	11.6%	Eastern Connecticut (Norwich)	Courtney, Joe (D)	46%
NY-26	11.3%	Greater Buffalo (Buffalo)	Higgins, Brian (D)	38%
CA-06	11.3%	Sacramento (Sacramento)	Matsui, Doris (D)	24%
FL-13	11.2%	St. Petersburg area (St. Petersburg)	Crist, Charlie (D)	46%
HI-02	11.0%	Northern Oahu and al (Hilo)	Gabbard, Tulsi (D)	30%
MI-05	10.6%	Flint, Saginaw, and (Flint)	Kildee, Dan (D)	46%
CT-03	10.2%	New Haven area (New Haven)	DeLauro, Rosa (D)	40%
AZ-01	10.2%	Northeastern Arizona (Flagstaff)	O'Halleran, Tom (D)	48%
CA-44	10.1%	South Los Angeles, i (Los Angeles)	Barragan, Nanette (D)	12%
HI-01	9.8%	Honolulu (Urban Honolulu)	Case, Ed (D)	30%
MN-07	9.7%	Western Minnesota (Moorhead)	Peterson, Collin (D)	62%
WI-03	9.4%	Southwestern Wiscons (Eau Claire)	Kind, Ron (D)	49%
IL-08	9.3%	Northwestern Chicago (Elgin)	Krishnamoorthi, Raja (D)	36%

CD	Clinton to R	Place	Incumbent	Trump %
CA-08	15.4%	Northern San Bernard (Victorville)	Cook, Paul (R)	55%
TX-04	5.7%	Northeastern Texas (Sherman)	Ratcliffe, John (R)	75%
WV-01	5.6%	Northern West Virgin (Parkersburg)	McKinley, David (R)	68%
FL-08	5.4%	Space Coast (Palm Bay)	Posey, Bill (R)	58%
WI-08	5.2%	Northeastern Wiscons (Green Bay)	Gallagher, Mike (R)	56%
MI-03	5.1%	Grand Rapids area (Grand Rapids)	Amash, Justin (I)	52%
OH-06	5.0%	Appalachian Ohio (Steubenville)	Johnson, Bill (R)	69%
IL-18	5.0%	West-central Illinoi (Peoria)	LaHood, Darin (R)	61%
TX-14	4.9%	Galveston area (Beaumont)	Weber, Randy (R)	58%
FL-02	4.9%	Florida panhandle (Tallahassee)	Dunn, Neal (R)	66%
TX-19	4.9%	Lubbock and rural We (Lubbock)	Arrington, Jodey (R)	72%
WI-01	4.7%	Southeastern Wiscons (Kenosha)	Steil, Bryan (R)	53%
OH-10	4.6%	Greater Dayton area (Dayton)	Turner, Michael (R)	51%
AR-01	4.6%	Northeastern Arkansa (Jonesboro)	Crawford, Rick (R)	65%
FL-12	4.5%	Northern Tampa subur (Palm Harbor)	Bilirakis, Gus (R)	57%

Note: Crossover estimates are from MRP models from the CCES. Place names come from Daily Kos (2019).

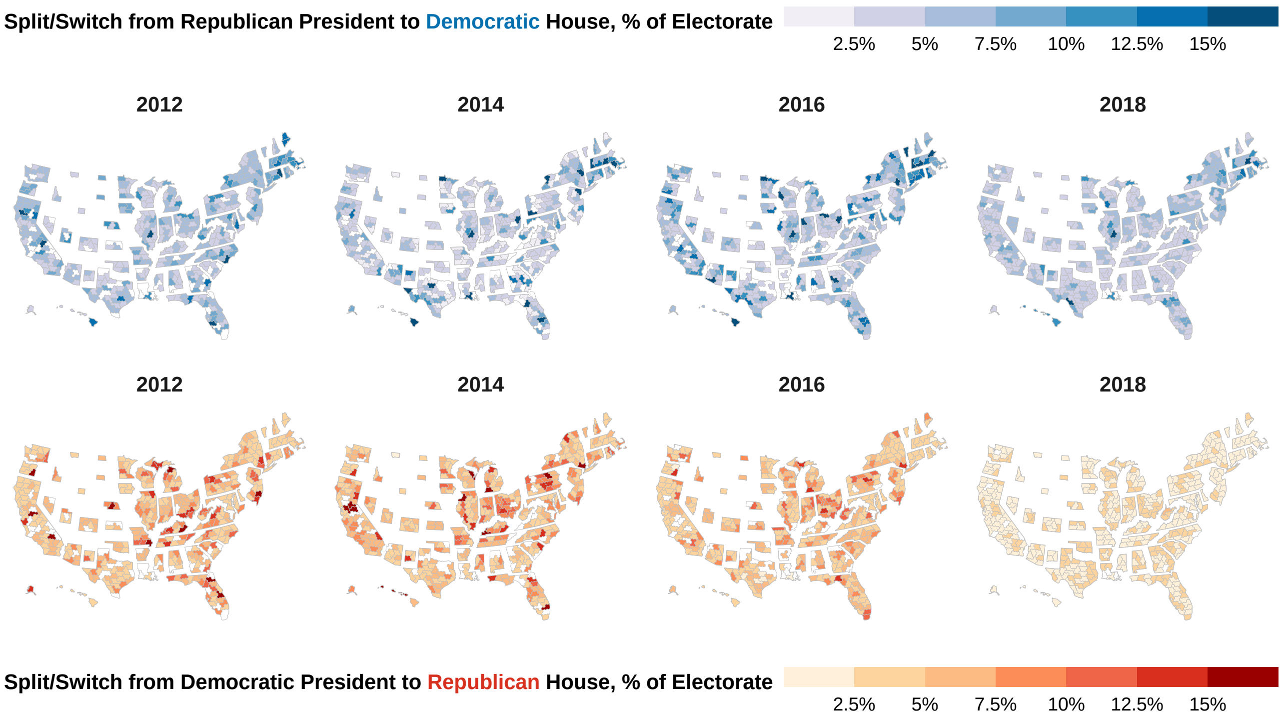


Figure 2.2: Swing Voters by Congressional District

Note: Maps colored by congressional district in each general election year. Top maps show the proportion of Republican to Democrat ticket splitters/switchers each year, while the bottom maps show the proportion of Democrat to Republican ticket splitters / switchers each year. Values are computed for 2008 and 2010 as well but not shown. Uncolored districts are uncontested, i.e. only a Republican or only a Democrat ran, and data from these districts were omitted from the MRP computation.

order of a percentage point or two.

The range of values estimated in Figure 2.2 may also strike most observers as exceedingly small. They are indeed an order of magnitude smaller than Key's estimates from the 1940s and 1950s, although Key compared switching between Presidential elections 4 years apart. My estimates are in line with the estimates by multiple other sources — for example, Jacobson (2019) finds similar numbers using a collection of polls. That the typical rate of individual-level swing is on the order of 5 to 10 percent is consistent with the narrative of increasing party loyalty and nationalization.

How do we square this small number with sizable seat swings – that for example in 2018, when we see the *lowest* level of individual-swing in the MRP estimates, the most number of congressional districts changed party control? A common explanation is differential turnout. 2018 saw a historic surge in turnout, so it could be that these new voters gave Democrats their critical votes, while 2016-2018 voters actually did not change their preference. This explanation, however, is not supported by the data in 2018. Ghitza (2019) estimates that about 90 percent of the total vote margin in 2018 was due to persuasion, not turnout.[1] They estimate that the gain in vote margin of drop-off and surge voters effectively offset each other, at least nationally. Historically, midterms have a pro-Republican turnout bias because Democratic voters stay home. In 2018, the midterm surge had a pro-Democratic lean, and so the turnout effect of each group cancelled out in the final vote margin. In contrast, their analysis finds that differential turnout was likely a decisive factor in 2014.

I argue instead that the association between individual-level swing and seat-level swing depends on whether swing voters in each district are pivotal. When a district is lopsided with partisan straight-ticket voters of a particular party, even a large bloc of swing voters that comprise 20 percent of the electorate may not be enough. But in a battleground where each party holds 47 percent of the electorate, even a small group that comprises 2 percent of the electorate is decisive. For example in 2016, Donald Trump won Michigan with a vote margin of two-tenths of a percent, won Wisconsin with five-tenths of a percent, and Pennsylvania with seven-tenths of a percent, and consequently won the electoral college. In Chapter 1, I showed that the control over Congress was similarly close in the modern era. The Presidency, the U.S. House, and

[1] To compute this number, Ghitza uses the full voter file maintained by Catalist and impute the vote choice or vote choice of each registrant in 2016 (President) and in 2018 (House). Access to the voter file allows them to subset the population into three types: Presidential drop-off voters (who vote in 2016 but stay home in 2018), the midterm surge (those who do not vote in 2016 but voted in 2018), and 2016-2018 voters.

U.S. Senate have slightly different dynamics — the U.S. House has more districts so its control hinges less on a single district, and the U.S. Senate only puts a third of its members for election — but each has had close control change rapidly.

In Section 2.6 I quantify pivotality by taking the observed vote share and computing a simple hypothetical vote share had swing voters switched backed their vote as a bloc. Under this definition, swing voters were probably decisive in 125 out of the 435 congressional districts, especially in suburban ones.

2.5 The Role of Candidate Incumbency

In a given year, the estimated prevalence of swing voters in US House districts appear similar across districts with no obvious geographic concentration. What, then, are the factors that describe the variation? With estimates by congressional district, I can test the contribution of candidate-level predictors, instead of only demographic predictors. I present three sets of analyses with increasing levels of causal identification: a predictive individual-level model, a two-way fixed effects model in a district-level panel, and finally an instrumental variables model. The goal is not to choose one explanation over the other, but rather to introduce a class of explanations that recent exploration of a single office might miss.

I estimate the contribution of incumbency in ticket splitting by an fixed effects approach. While the predictive model in the previous section suggest that the fact that a candidate is an incumbent leads to more ticket splitting, it may be the case that unmeasured confounding between the type of voters in a district and the tendency for incumbents to stay in office in that district biases the estimates. One way to account for this possible confounding is to group data at the constituency level (or district, indexed by c) and year level (indexed by t), and estimate the following two-way fixed

Table 2.3: Incumbency and Ticket Splitting

	Switch to Democratic Candidate				Switch to Republican Candidate			
Method	2FE	PM	PM	PM	2FE	PM	PM	PM
Lags Used to Match		1	2	3	1	2	3	
House Incumbent	0.028*	0.018*	0.021*	0.009	0.015*	0.018*	0.015*	0.002
	(0.003)	(0.005)	(0.008)	(0.009)	(0.003)	(0.006)	(0.004)	(0.011)
Open Seat	0.004*				0.002			
	(0.001)				(0.002)			
Republican vote share	0.018				−0.031			
	(0.018)				(0.021)			
CD Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓				✓			
Observations	1,524				1,524			

Note: Each column is a regression estimating the effect of having a Democratic (Republican) Incumbent as a US House Candidate on the proportion of voters who split their ticket for the Democratic (Republican) Candidate. 2FE is a standard two-way fixed effects method, and PM uses the PanelMatch method by Imai, Kim, and Wang. Lags indicate the number of cycles that are used to generate the matched set. Standard errors are clustered at the district level for 2FE and are computed from bootstraps in PM. $^*: p < 0.05$.

effects model:

$$S^{\text{L}}_{ct} = \beta_0 + \beta I^{\text{L}}_{ct} + \gamma_c + \gamma_t + \varepsilon_{ct} \tag{2.1}$$

where S^{L}_{ct} is the percentage of voters in the district that split their ticket for candidate L as measured by MRP, I^{L}_{ct} is a binary treatment variable for whether the Democratic (L) candidate is an incumbent, and γ_c and γ_t are constituency and time fixed effects. The least squares estimator for β identifies the one-shot effect of the Democratic candidate becoming an incumbent on splitting, controlling for time-invariant characteristics of the district.

The two-way fixed effects estimator is equivalent to a weighted average of difference-in-differences estimators that each estimate the Average Treatment Effect on the Treated (ATT) from a matched set of control units, but where each match is not optimal (Imai and Kim 2020). To improve the quality of the matches, I also estimate a matched difference-in-difference estimator that matches pre-trends based on covariates (Imai and Kim 2019).

The current dataset covers all contested congressional districts each observed in four cycles (2012-2018). When a district is uncontested, the observation is set as missing. When estimating each treated unit's match, I also include the missingness patterns as a matching criterion, as well as lagged values for presidential vote share and lagged values for the treatment indicator.

Table 2.3 shows the main coefficient of the panel regressions. Across specifications, being a House incumbent is associated with about a 2 percentage point increase in the percent of ticket splitting. This estimate is on the same order of magnitude as the predictive models estimated in the previous section. The columns labelled with 2FE show the two-way fixed effects estimate, and columns with PM show the PanelMatch method. Using sets matched on pre-treatment covariates generate similar sized estimates, except when three pre-treatment periods are used to generate matches. This is not surprising because there are only four time periods in the data and so the effective sample size for such a match is small.

2.6 Hypothetical Vote Margins

The district estimates of the proportion of a President's supporters who vote for the opposing part in the U.S. House is a useful representation because it can also be compared on the same units as actual election outcomes. This allows us to consider some hypothetical scenarios for election outcomes if vote switchers had not switched their vote. By providing district specific estimates, moreover, I avoid the uniform swing assumption used in Chapter 1.

As a concrete example, consider a congressional district where the Democrat narrowly won by 4 percentage points, i.e. the two party voteshare was 52 to 48. Suppose that 3 percent of those who voted were crossover voters in favor of the eventual winner (the Democrat), and 1 percent of those who voted were crossover voters in favor of the eventual loser (the Republican). Now consider the hypothetical where friendly crossover voters had *not* crossed over and stayed with their previous vote for the Republican party. Shifting over the mass to the other side, the two party voteshare is now 49 to 51 and the Democrat *loses* by 2 points. Next suppose that the unfriendly crossover voters did not switch as well. Then the margin is 50 to 50 and the Democrat ties. It is common knowledge in political campaigns that persuasion is worth "twice" more than mobilization because changes in vote choice will have double the effect in margin.

In Figure 2.3 I align pairwise comparisons of the actual win margin and hypothetical margin of each of the 390 contested 2018 House districts. Each comparison is depicted with an arrow. The starting point of the arrow is the win margin observed from the election result, and the end point of the arrow is the hypothetical win margin under two conditions. In other words, let $M_{0i} \in [0 + \epsilon, 1]$ be the observed win margin for district i. Then, using the survey data, we attempt to estimate the proportion $\psi_i \in [0, 1]$ of the friendly crossover voters in district i. Separately, we estimate the mass of "unfriendly" crossover voters are of mass $\tilde{\psi}_i$. Then I define the hypothetical margin, which I denote M_{1i} as the margin, still for the same candidate, as:

$$\widehat{M_{1i}} = \widehat{M_{0i}} - 2\widehat{\psi_i} \tag{2.2}$$

when the counterfactual is that only friendly cross-over voters swing back, and

$$\widehat{\widetilde{M}_{1i}} = \widehat{M_{0i}} - 2(\widehat{\psi_i} - \widehat{\tilde{\psi}_i}) \tag{2.3}$$

for the counterfactual that both types of crossover voters switch back to their 2016 vote. Figure 2.3 shows the same value of M_{0i} but draws the line to either $\widehat{M}_{1i}$ or $\widehat{\widetilde{M}}_{1i}$.

Both hypotheticals are important in their own ways. The first indicates the worst case scenario if the winner's coalition asymmetrically defected. The second indicates a partisan polarized case where no one defects. However, because elections tend to have roughly uniform swing that advantages one party over the other, it may be less plausible to imagine a symmetrical decline in both types of switches either.

The cross-over voter is *pivotal* when the arrow ends up reaches at our beyond 0. In the figure, the districts are ordered first by their rural - suburban categorization, and then by their value of M_{0i}. In each facet, I count the percentage of plotted districts whose hypothetical estimate crosses zero, and call those districts "pivotal." With the exception of completely rural and completely urban districts, the first hypothetical indicates that friendly crossover voters were pivotal in 20 to 35 percent of districts.

The first key implication from this first hypothetical is that although 1-3 percent of vote swing may at first seem small, it can still more than explain the consequential seat swings in the Congress as a whole (of 30 flips) here. The second implication is that whether or not swing voters are consequential depends not only their size in the population but also their spatial position in the district. For example, in suburban districts, win margins were sufficiently low to begin with — i.e. the districts were probably more competitive — that small amounts of cross-over voters were more likely to be decisive in those districts.

The second panel in Figure 2.3 shows a similar, if more moderate, picture. Because friendly and unfriendly crossover voters cancel out, the lengths of the arrows are always smaller than the top figure, and can go in different directions. Interestingly, Democrat candidates across the board tended to win thanks to vote switchers, whereas winners of Republican candidates tended to win *despite* vote switchers.

These figures use $\widehat{\psi}_i$ as fixed but there is of course uncertainty around these estimates. To account for this, I approximate confidence intervals around my estimate in the following way. I take the standard error around $\widehat{\psi}_i$ implied by the MRP credible intervals, and then multiply it by $\sqrt{2}$ and then by the Normal cumulative density $\Phi(0.9)$ to estimate the 80 percent (as is standard for Bayesian models) confidence interval around $2\widehat{\psi}_i$. I take the observed win margin as constant. In Figure 2.4, I drop the observed margin and simply plot $\widehat{M_{1i}}$ with those confidence intervals. If the confidence intervals cross 0, that means I fail to reject the null hypothesis that $\widehat{M_{1i}} < 0$ with $\alpha = 0.20$, implying (loosely) that swing voters were likely in that district. I naturally get larger estimates of pivotality based on these estimates.

2.7 Conclusion

A systematic examination of survey data which are then translated into quantities directly comparable with district voteshares leads to a richer picture of how the swing voter fits into nationalized electoral politics. The first set of predictive models shows that ticket splitters tend to have low news interest and have low educational attainment, but are distributed across the urban-rural divide. Through a set of panel regressions, I also showed that there is a candidate component as well as a voter component to explain whether or not someone is a swing voter. These findings accord with the spatial voting model that I used initially to justify the operationalization of the swing voter in survey data.

There are two main areas for methodological improvement in this approach. First, more covariates can be accounted for in the post-stratification and pooling to stabilize estimates. I show in Chapter 4 that modeling a synthetic population table by combining other population margins. Second, we can take further steps to decompose electoral change between a presidential and subsequent midterm year into persuasion and turnout changes. Ghitza (2019) provides one way forward to do this with voter file data and Hill, Hopkins, and Huber (2021) use precinct-level ecological data to estimate this decomposition, but the validity of both methods should be tested in other contexts.

Estimating district-specific quantities of the swing voter allows us to test more precisely the Swing Voter Paradox: Even though they comprise a small proportion of the electorate, ticket splitters and party switchers are collectively pivotal in crucial races that determine the control of Congress. The scenarios in 2018 show that this is indeed the case. Swing voters are not concentrated in sububran districts, but they were enough in competitive districts where Republicans and Democrats were roughly equally prevalent and the sliver of ticket splitters could swing the winner. It turned out in 2018 that these competitive districts were largely suburban, and were tipping point districts.

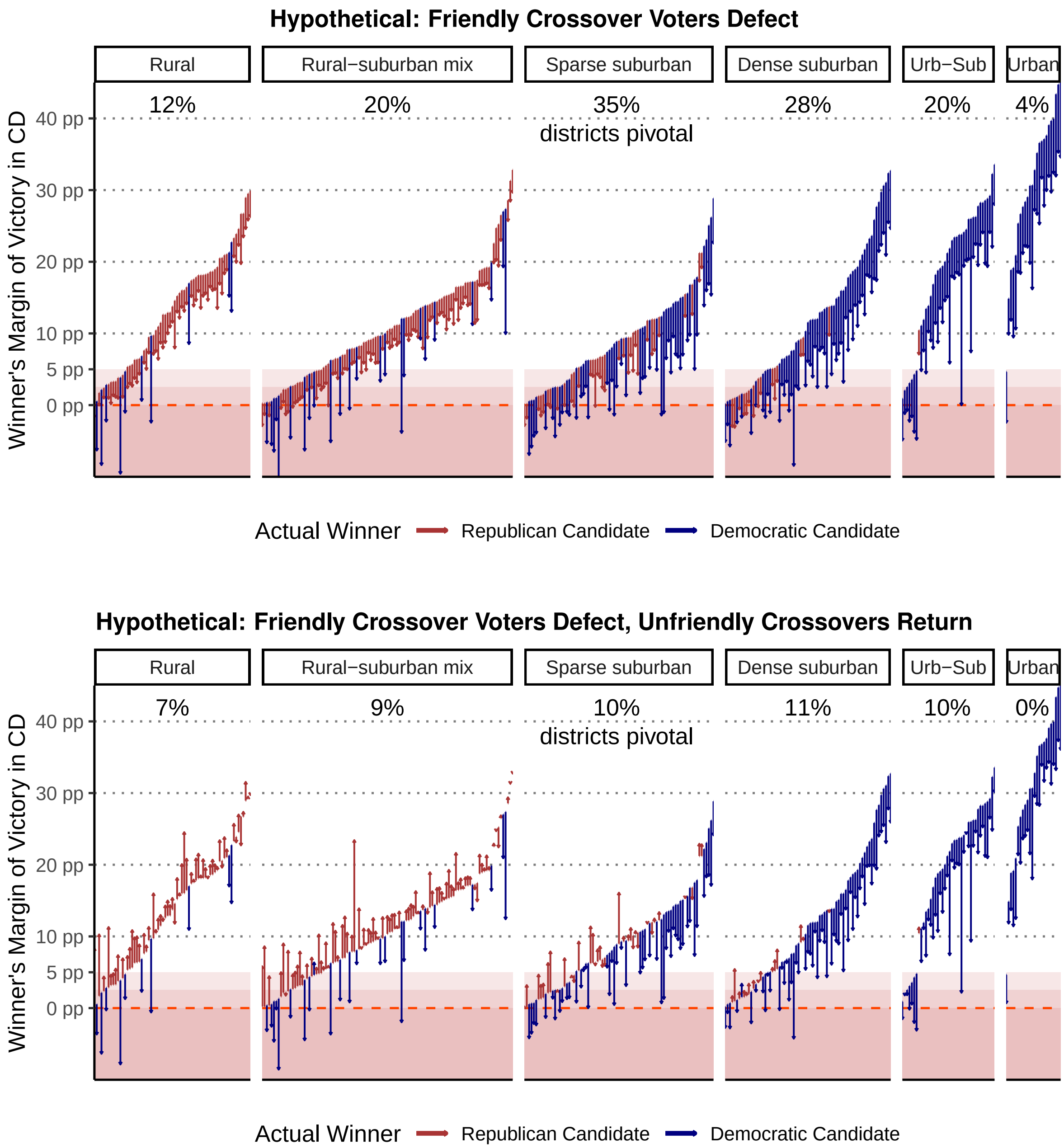


Figure 2.3: Pivotality of Vote Switchers in 2018

Note: Arrows connect actual margin to hypothetical margin when vote switchers do not switch. See text for measurement.

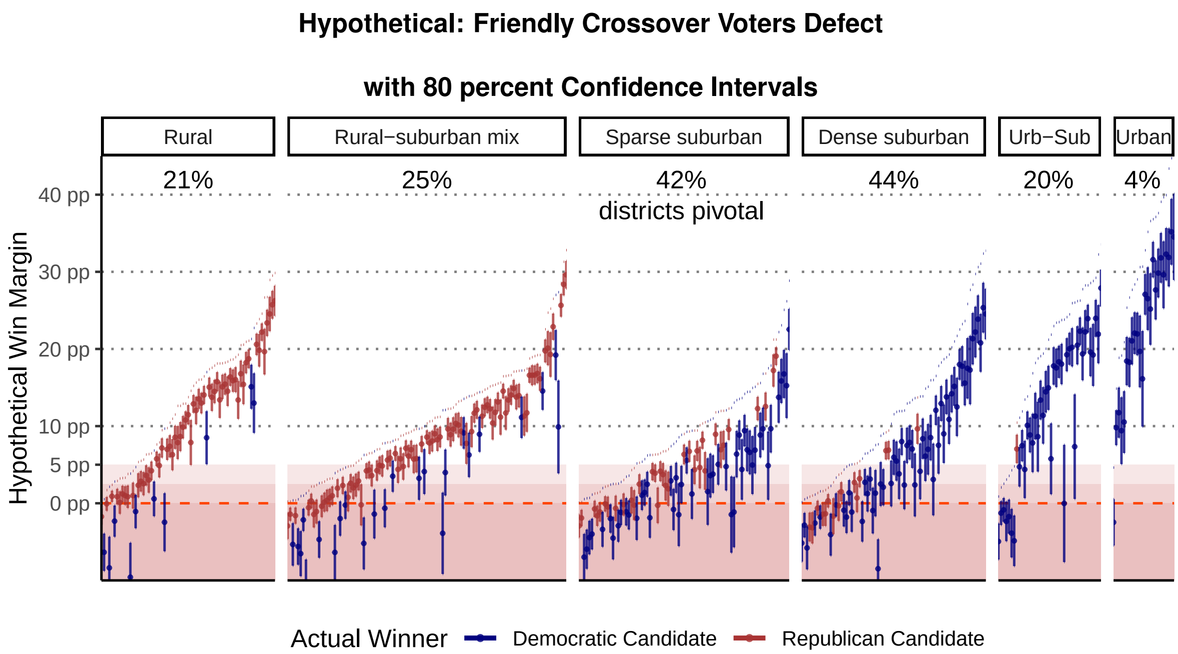


Figure 2.4: Estimates of Hypothetical Margins with Confidence Intervals

Note: Same values as the endpoints in Figure 2.3 top panel, but puts 80 percent confidence intervals around those estimates.

3 | Party Loyalty on the Long Ballot: Ticket Splitting in State and Local Elections

Abstract

Many believe that party loyalty in U.S. elections has reached heights unprecedented in the post-war era, although this finding relies on evidence from presidential, congressional, and gubernatorial elections. If party labels are a heuristic, we would expect party-line voting to be even more dominant in lower-information elections. Yet, here I show that the prevalence of ticket splitting in state and local offices is often similar to or higher than in national offices because of larger incumbency advantages and starker candidate valence differentials. Because neither surveys nor election returns have been able to reliably measure individual vote choice in downballot races, I introduce an underused source of voter data: cast vote records. I create a database from voting machines that reveals the vote choices of 6.6 million voters for all offices on the long ballot, and I design a clustering algorithm tailored to such ballot data. In contrast to ticket splitting rates of 5 to 7 percent in U.S. House races, about 15 to 20 percent of voters split their ticket in a modal Sheriff race. Even in a nationalized politics, a fraction of voters still cross party lines to vote for the more experienced candidate in state and local elections.

* I thank Duncan Buell, Jeff Lewis, Stephen Pettigrew, and Charles Stewart for their expertise in cast vote records; Carolyn Abott, Yuki Atsusaka, Peter Buisseret, Justin de Benedictis Kessner, Barry Burden, Hanno Hilbig, David Kimball, Edward Lawson, Jr., Daniel Moskowitz, Socorro Puy, Andrew Stone, Clémence Tricaud, Chris Warshaw, Soichiro Yamauchi, Hye Young You, Michael Zoorob, members of the Imai Research Group, members of the Democracy Policy Lab at Stanford for their comments; and to Steve Ansolabehere, Matt Blackwell, Kosuke Imai, and Jim Snyder for their guidance.

3.1 Introduction

The nationalization of voter behavior in recent decades is thought to have shifted the electoral landscape, changing the conventional wisdom about the U.S. electorate. More people vote for the same party's candidates in races for President and the U.S. House Representatives, or for U.S. Senator and Governor, than at any other point in the post-war era (Jacobson 2015; Sievert and McKee 2018). The implications of nationalization are of particular concern in state and local politics. Citizens in a federal system can elect state and local representatives separately from national ones, but under a thoroughly nationalized politics these constituents would vote for the nominee of their preferred party regardless of each candidate's relevant experience or the duties of the respective office. As Hopkins (2018) warns, party loyalty gone too far may turn into blind partisanship and dampen electoral accountability.

However, almost all of the evidence in recent accounts of nationalization comes from presidential, congressional, and gubernatorial contests.[1] Is ticket splitting more prevalent in races for state legislature, sheriff, or county council than they are in the evidently nationalized offices of President, Congress, or Governor? This paper constitutes one of the first analyses of individual vote choice in partisan state and local offices relative to national offices. I find that party loyalty is strong among national offices but weakens modestly and becomes more variable in down-ballot partisan races. Importantly, I show that this ticket splitting is systematic: it tends to advantage the incumbent in contested seats and the candidate with more campaign finance contributions in open-seat races.

The fact that ticket splitting exists at all during a time of nationalized politics

---

[1] The literature of ticket splitting based on these elections is extensive. In American Politics, see Campbell and Miller (1957), Beck et al. (1992), and Burden and Kimball (2002). In Comparative Politics, see Burden and Helmke (2009) and references therein.

might surprise some readers. Existing literature finds nationalization in gubernatorial elections (Sievert and McKee 2018), state legislative roll call votes (Rogers 2017) and public opinion (Tausanovitch and Warshaw 2014), which would lead us to speculate that local politics has also nationalized – local candidates in nationalized elections would take the same positions as their national party platform, and voters would cast a straight ticket, voting for the same party across national, state, and local offices. Moreover, theories of partisanship as a heuristic would predict that blind partisan voting would be *more* prevalent in low-information contests like state and local races where candidate information is harder to come by (Peterson 2017). But what these predictions fail to take into account is the possibility that other candidate-based factors like the incumbency advantage in state and local offices could be large enough to counteract straight ticket voting. In classic spatial voting models, the probability of ticket splitting is increasing in one candidate' advantage on "valence" factors: a bundle of aspects such as incumbency, positive name recognition, and effort that all voters prefer more to less. If voters can perceive *and* care about these attributes, they may vote for the candidate with a valence advantage even if she is of a different party.

Testing the relative importance of party line voting and the incumbency advantage is well-studied in Congress, but the existing literature is only suggestive when it comes to state and local politics because they rely on aggregate election returns (Trounstine 2018) or limited survey responses measured with error (Beck et al. 1992). And all of these studies examine one subset of offices on the ballot at a time. In a recent review of the literature on local elections and representation, Warshaw (2019) calls for "much more survey evidence on local elections than there currently exists" to understand vote choice in state and local elections.

This paper takes a different methodological approach to measure individual vote choice by taking cast vote records, or complete records of an individual voter's vote choice on the entire long ballot handled by election administrators. Improving upon past studies of cast vote records which study only several counties in a single year (Gerber and Lewis 2004; Bafumi et al. 2012; Hansen 2015), I reconstruct the ballot layouts from South Carolina voting machines for five general elections between 2010-2018 and combine it with candidate information found in state newspapers and campaign finance disclosures. These individual records of actual vote choice prove illuminating especially for offices where currently no survey-based measure of ticket splitting exists, including state legislatures (Rogers 2017), county councils (de Benedictis-Kessner and Warshaw 2020), and sheriffs (Farris and Holman 2017; Thompson 2019).

I present my main findings in two parts. In the first, I provide an accounting of straight ticket voting for the full range of offices. In a single contested race, around 80 to 90 percent of voters vote for the same party, and so across all the offices on the ballot 20 to 40 percent of voters cast at least cast one split ticket vote, depending on the number of contested races. With a clustering algorithm, I show that about 20 to 35 percent of voters have voting patterns that can be classified as a swing voter bloc, with a 20 to 50 percent probability of splitting their ticket. Moreover, I document that straight ticket voting is often lower in state and local offices than in congressional ones, and considerably more variable.

The second part explains these patterns by showing that measures of candidate valence qualities such as coverage in the news, fundraising, and incumbency are all positively associated with ticket splitting. With individual-level regressions of down ballot vote choice controlling for national party preference as revealed by choices at the top of the ticket, I show that a substantial number of voters defect from their national party preference when the candidate who shares a voter's national party choice is challenging an incumbent. My analysis is agnostic about which aspect of incumbency affects voter's choices the most. Clearly, candidate quality, campaign effort, and name familiarity are plausibly endogenous to each other, but further analyses suggest that it is not name familiarity or one-time campaigning alone that leads voters to support state and local candidates from an opposing party.

This paper therefore contributes previously unknown facts about vote choice in state and local races, offers an explanation that would hold even in an electoral environment where candidates have nationalized, and raises several puzzles for future research. The partisan election is a central feature of American Politics that links national politics with state and local policy. With the aid of cast vote records, I show that state and local candidates with more experience net a substantial amount of votes from opposing partisans, not unlike the Congressional candidates of a less nationalized era.

3.2 Ticket Splitting in a Nationalized Era

In the 2018 general election, Henry McMaster, the incumbent Republican Governor of South Carolina, won all 30 counties that President Donald Trump won in 2016, and flipped only one out of 16 counties won by Hilary Clinton. McMaster's and Trump's county-level two-party vote share were correlated at 0.99. This tight link between voter's choices between national and state offices is a typical example of how straight ticket rates among presidential, congressional, and gubernatorial races have increased steadily since the 1980s.

Accounts of nationalization, most comprehensively argued by Hopkins (2018), explains trends like these, where a partisan geography of a Governor's votes became almost indistinguishable from those for President or Congress. The two most well-studied dynamics in American electoral behavior are party-line voting — voting for the same party's candidates across offices — and the incumbency advantage — the support for incumbents regardless of their party (Ansolabehere and Snyder 2002). The theory of nationalization argues that as parties polarized in Congress, the platforms of state parties became more similar to national party positions, and voters began to identified with the Republican or Democratic party undifferentiated by the level of government in question. The incumbency advantage for U.S. House elections declined in tandem (Jacobson 2015). Many of the same scholars also believe that the decline of local news accelerated this trend. Moskowitz (2020) suggests that the decline in the supply of news related to state government leads to a decline in ticket splitting for state offices such as Senate and Governor.

What is not nearly as well as understood is how these patterns play out in state and local politics. Another feature of American elections is that voters directly elect many important local legislative and executive offices, often with ballots where these candidates run on partisan labels. Some partial evidence suggests state and local offices may be different. To take the same 2016 election in South Carolina as an example, voters collectively elected representatives in 546 partisan races. The results in these races differed widely by office: Democratic candidates won only 1 out of the 7 congressional seats, but won 50 percent of countywide sheriff races and 49 percent of partisan county council seats.

Does the discrepancy between the Democratic seat share in these offices imply that some voters are voting for Republicans in federal elections but splitting their ticket for Democrats in local races? Ultimately, election returns are inconclusive because they are aggregate measures. For example, suppose hypothetically that South Carolina's 46 counties were composed of 23 purely Republican counties and 23 purely Democratic counties, and congressional districts were gerrymandered to crack the Democratic vote. That electorate could have produced the aforementioned seat shares without a single voter splitting her ticket for candidates of a different party on the same ballot.

3.2.1 Limitations to Existing Studies

To find how often individual voters split their ticket and why, existing studies almost exclusively rely on two types of data: Aggregated election returns and survey samples measuring self-reported vote choice. The use of election returns dates at least back to V.O. Key's chapter in his treatise of state politics, "The Lottery of the Long Ballot" (Key 1963, ch.7) and continues to be used in recent work (Trounstine 2018). While comparisons of election returns in different districts can provide a sense of the directionality of an "incumbency effect," for example, they do not reveal individual voting patterns and can severely underestimate the prevalence of ticket splitting. As I showed in Figure 1, the difference in voteshares is often a lower bound for the total number of ticket splitters that exist. In the past few decades, social scientists have developed ecological inference methods to estimate the actual amount of ticket splitting from office-level aggregate data. These methods estimate the joint voting probabilities that best comports with aggregated returns subject to modeling assumptions (King 1997; Wakefield 2004; Greiner and Quinn 2009). Burden and Kimball (2002), for example, were one of the first to apply ecological inference techniques to election returns and estimate the degree of ticket splitting between votes for U.S. House and President. As developed as these methods are, they are inherently model-based inferences. Their output may be biased and the opportunity to quantify the direction of that bias is rare.

Surveys circumvent the aggregation problem by sampling enough individuals and prompting them to self-report their vote choice. They serve as the central piece of evidence in documenting the rise of straight ticket voting uncovering its mechanisms (Jacobson 2015; Abramowitz and Webster 2016; Davis and Mason 2016). But when studying state and local elections, surveys become prohibitive for measuring any type of vote choice, much less ticket splitting, due to two limitations. The first is measurement error through misreporting. Surveys require voters to either recognize the names of the candidates they will vote for in advance of the election, or recall which candidates they voted for several days after. With a ballot that can range from 10 to 20 names in a general election, this process can be fraught with error. For example, in the 2014 Cooperative Congressional Election Study (CCES), 89 percent of the 453 respondents in South Carolina reported voting for a candidate in state senate even though there was no state senate race in South Carolina that year (only 6 percent reported that “there was no race for this office”). The second limitation is simply that very few surveys poll on state and local races. Beck et al. (1992) is the only published study, to my knowledge, that polled statewide offices such as Attorney General by name, and Arceneaux (2006) reports that there were only three cities in the November 2002 election for which the U.S. Senate, Governor, and Mayor were on the same ballot. Moreover, sample sizes for these two studies were limited to around 500 and 1,200, respectively.

In this paper I propose to analyze cast vote records, a dataset that is free of these measurement problems. Before describing the data, however, I first outline possible explanations for why voters would split their ticket.

3.2.2 Potential Explanations for Why Voters Split their Ticket

Analysts typically interpret the lack of ticket splitting as a measure of the nationalization and polarization of voter preferences, but the literature on ticket splitting has long shown that voters split their ticket for many other reasons. I will ultimately focus on a valence advantage explanation, which provides some of the clearest theoretical predictions.

A spatial voting framework bears out the logic of existing explanations. In a canonical spatial model, citizens choose the candidate whose policy position is closest to them on an ideological spectrum. Voting on the long ballot is akin to a citizen in these models making a series of choices between candidates of two parties. If all Republican candidates for these choices held identical policy positions and all Democratic candidates held another set of identical positions, a voter would cast a straight ticket vote. This setting mimics a completely nationalized politics: co-partisan candidates for local and national office run on identical platforms and voters vote accordingly for a party slate.

There are at least three classes of explanations for why a voter might split their ticket. First, voters could cross party lines in races where the candidates are less polarized. This is the simplest explanation because we still presume a single issue dimension. A second explanation for ticket splitting posits that state and local politics is contested over different issue dimensions. For example, Oliver, Ha, and Callen (2012) document how local politics revolves around contestations over land use, economic development, and other issues specific to that locality. And recent survey evidence shows that voter's preferences over those policy debates often do not align with their partisanship (Jensen et al. 2019). If local elections feature candidates with differing views on the environment, for example, while candidates for national office all take the same position on the environment, environmentally conscious voters would vote for the same party within national offices but then defect from that party allegiance in local races to vote for the pro-environment candidate (Besley and Coate 2008).

The third reason voters may split their ticket, the one I focus on the most in this paper, is that voters care about valence and one candidate has a valence advantage. Valence is an attribute that all voters prefer more of to less, such as candidate compentence, effort, or experience for the job. The main insight from spatial models with valence is that ticket splitting for a candidate would be increasing in that candidates' valence advantage relative to the candidate's spatial distance. I show these results formally in the Introduction of this dissertation (equation 3). Furthermore in this setup, more certainty around a candidate's policy position effectively constitutes as a valence advantage as well. In this sense, candiate visibility, or salience, is captured within the concept of valence as well.

Testing a model of valence taps into a long literature on the incumbency advantage. This literature has established that (1) Congressional incumbents have always been advantaged by their status at least since the 19th century (Carson, Sievert, and Williamson 2019), (2) when these incumbents first won their seats in an open race, they tended to have more relevant experience than their opponent (Hirano and Snyder 2019), and (3) such candidate quality differences between the incumbent and the challenger drives fluctuations in the advantage (Cox and Katz 1996) although parsing apart the effect of quality, scare-off, and advantages that accrue from governing is challenging (Eggers 2017; Ashworth, Bueno de Mesquita, and Friedenberg 2019). Using incumbency as a measure of positive valence is a natural choice in state and local politics: regression discontinuity designs have found that the effect of being an incumbent on staying in office the next election to be substantial in city council and mayoral elections, as large as 20 to 30 percentage points (Trounstine 2011; de Benedictis-Kessner 2017; Warshaw 2019). Existing empirical work on split ticket voting in higher profile offices also lends support for the model that the higher quality candidate nets more by voters splitting their ticket. Beck et al. (1992) showed that highly visible candidates draw more split ticket voting, and Burden and Kimball (2002) showed that congressional candidates with larger campaign expenditures appear to compel more voters to split their ticket.

Information is crucial in these valence-based accounts. A candidate's valence advantage cannot factor into voter's decisions unless that information reaches voters before they vote, for example through campaigns and press coverage. In all but three

U.S. states, ballots do not contain anything other than the candidate's name and party.[2] That is why theories of information processing may lead one to predict *more* straight ticket voting in low-salience elections (Darr, Hitt, and Dunaway 2018; Moskowitz 2020) . Peterson (2017) showed systematically that the lack of candidate-specific information increases the likelihood that voters vote straight ticket

In this paper, I primarily test the valence hypothesis for theoretical clarity, as well as to show how ticket splitting could occur even when candidates have, as conventional wisdom goes, nationalized. Of course, multiple issue dimensions may be at work as well. But spatial models with multiple issue dimensions are notoriously intractable, so their theoretical predictions are less clear. And while moderation may be a factor, valence is an explanation that is theoretically plausible even if all candidates of the same party are polarized, as an account of nationalization would stipulate.[3]

3.3 Data, Methods, and Case

Despite the centrality of straight ticket voting to the discussion of nationalization and the incumbency advantage, past work has struggled to measure this individual level behavior in the offices that stand to change the most drastically if it were to nationalize. What is needed is an approach that drills down each individual's long ballot, observing the entire series of a voter's choices. Fortunately, cast vote records do precisely that.[4] I first show how these records compare to traditional data sources. I then introduce the cast vote records in South Carolina as the main dataset in this paper.

---

[2] The exceptions, as of 2018, are California, Georgia, and Massachusetts, based on the samples acquired by Ballotpedia. https://perma.cc/8ADA-B5YT.

[3] Testing the moderation hypothesis is also complicated by the fact that candidate positioning is likely endogenous to their valence advantage (Groseclose 2007).

[4] Cast vote records are also referred to as ballot image logs.

3.3.1 Cast Vote Records

Cast vote records are complete readouts from voting machines. Ballots in the U.S. are either paper-based or electronic (referred to as DRE, or Direct-Recording Electronic). DRE machines record each vote as it is cast through a touchscreen. Ballot anonymity ensures that these records are de-linked with the identity of the voter from the outset. While precinct election officials only report candidates' vote totals, information about each ballot is technically retrievable from the DRE software. Jurisdictions sometimes choose to release these records as public record for post-election audits.

Table 3.1 summarizes how cast vote records differ in their measurement of electoral behavior from aggregate election returns, surveys, and voter files. None of the three traditional alternatives simultaneously provide data at the individual level *and* measure vote choice without sampling error, but cast vote records do. They are not without their limitations — For example, demographic information from voter files cannot be linked to individual votes to protect ballot secrecy.[5] Yet, once they are formatted and linked to information about candidates, they are uniquely conducive to study vote choice across the long ballot.

The use of cast vote records in academic research is not new, but the few existing studies using cast vote records have not attempted to describe or explain the patterns of vote choice in state and local (as opposed to statewide or congressional) contests. Past contributions to measurement are exemplified for instance by Park, Hanmer, and Biggers (2014), who use the records as a ground truth for ecological inference estimates in federal offices. Election researchers use it to study specific elections such as ranked choice voting (Alvarez, Hall, and Levin 2018). The past studies most similar to

[5] This rules out analyses of racial voting, an important feature of politics in South Carolina, at least at the individual level. Votes by the same person across separate elections are not linkable either.

Table 3.1: Cast Vote Records Compared to Traditional Electoral Datasets

	Cast Vote Records	Election Returns	Surveys	Voter Files
Individual level?	✓		✓	✓
Vote choice observed?	✓	✓	✓+ error	
Covers all offices on the ballot?	✓	✓		
Personally identifiable information?			Imperfect	✓
Contains precinct identifier?	✓	✓		✓

Note: Each column lists the properties of a type of major election dataset.

this paper also use cast vote records to describe voting patterns at a level of granularity that surveys could not achieve. Gerber and Lewis (2004) standardized ballot images from Los Angeles county in the 1992 general election and estimated voter's ideal points from their choices in statewide ballot referendums. Later Herron and Lewis (2007) standardized ballot images from ten Florida counties from the 2000 presidential election to estimate the partisanship of Ralph Nader voters based on their votes in down ballot partisan contests. Both studies use cast vote records primarily to estimate latent preferences, while this paper describes and models the full set of votes that would underly such a summary measure.

3.3.2 Processing South Carolina Cast Vote Records

In this study I use records from 6.6 million voters across five general elections in the same state, the largest collection of cast vote records to date. Counties vary widely in how they administer elections but South Carolina offers a rare opportunity to study the long ballot because it runs a centralized and transparent election administration system. From 2005 to 2018, all counties in South Carolina used the iVotronic DRE voting machine manufactured by Election Systems and Software (ES & S). Sample images of the iVotronic appear in Appendix A.1.

Figure 3.1 shows images displayed in the ES & S iVotronic. All South Carolina machines use the same make of machine, and have similar displays. Offices are usually placed in the order of federal, state, county, local non-partisan races, and referendums. Within each contest, candidates are always ordered by party. For example in this ballot, the Democratic candidate is placed before the Republican candidate.

The State Election Commission has made the cast vote records from these machines public from the 2010 general election onwards as part of their post-election audit. In 2010, an unexpected upset in the state's Democratic primary for U.S. Senate led Buell et al. (2011) and Bafumi et al. (2012) to analyze the state's cast vote records through a public records request.

The iVotronic cast vote records are not immediately useful because they are logs of the choices each voter cast, and the set of choices a voter chose *from* must be inferred by the analyst. I therefore reconstruct the ballot from the cast votes in the following steps. I first standardize each county's log output into a tabular form with identifiers for each voter, precinct, and contest. I then merge in the party affiliation of each chosen candidate using official filing records, and infer which contests were up for election in each precinct-ballot style combination from the set of votes. Details of these steps are left to Appendix A.1. As a result of this procedure, I am able to capture how each election day voter voted on the full extent of their ballot.

The final dataset covers 59 elections from 2010 overseen by the state, but among these I study all election day voters in the general elections held between 2010 and 2018 (See Appendix A.2 for a summary table). These are the elections that featured a long ballot, asking each voter to make choices for on average 18 offices in a midterm year and 11 offices in a presidential year. Partisan contests, i.e. contests where candidate's party affiliation is shown on the ballot, comprise about two thirds of these contests.

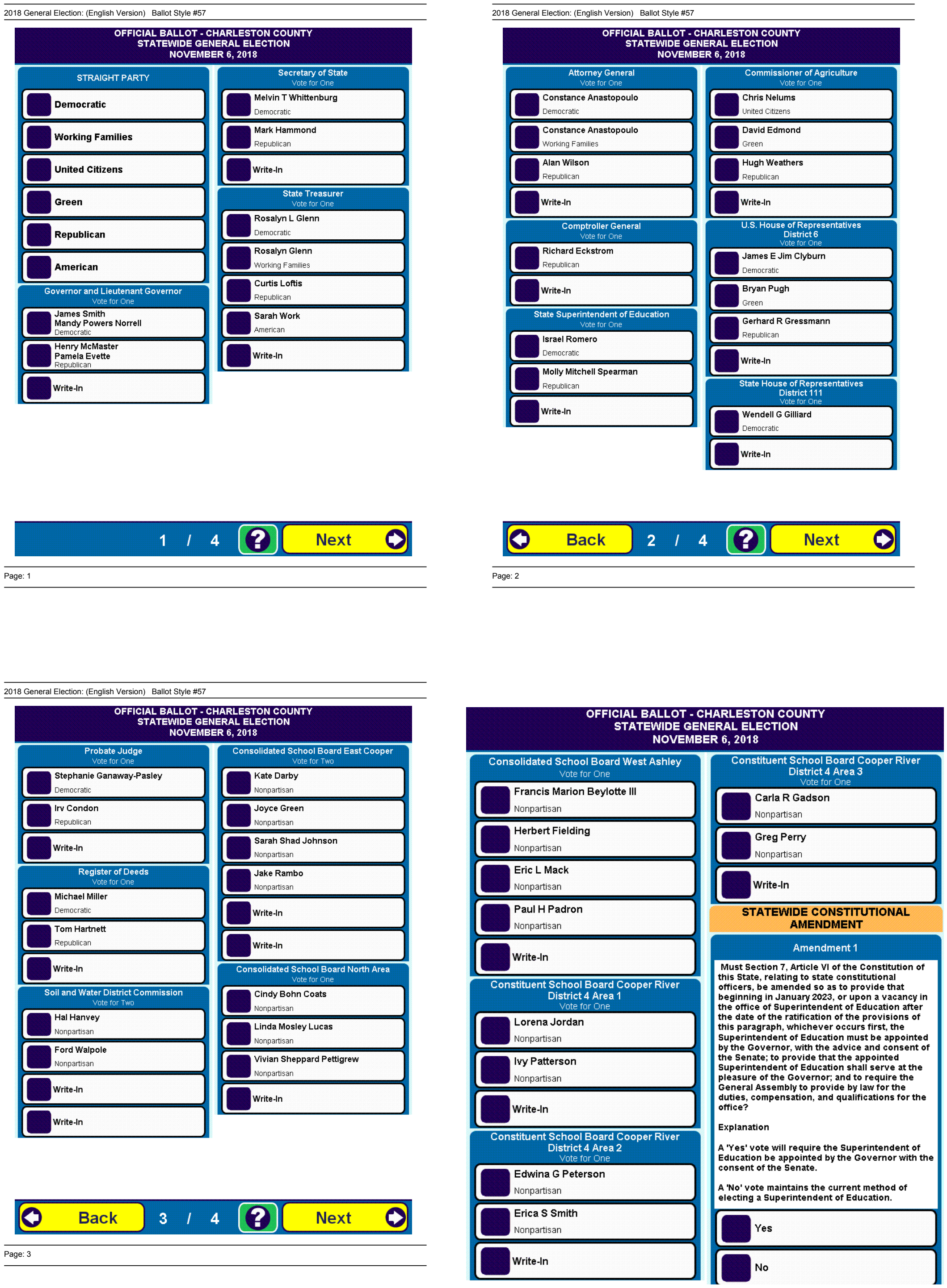


Figure 3.1: Samples of the iVotronic Touch Screen

Note: Figures show screen images as they are displayed on the iVotronic. This example comes from a particularly short ballot style in Charleston County in 2018, in which all contests fit in four screens.

3.3.3 State and Local Elections and Politics in South Carolina

South Carolina is comparable to other states in the number of elected offices on each voter's general election ballot. The state has 7 congressional districts, 124 state house districts, 46 state senate districts, and 46 counties each with a county council often elected through single member districts. Statewide, Attorneys General, Secretaries of State, agricultural commissioners, and superintendents are elected in conjunction with the Governor's race in midterm years. Countywide elections include the partisan offices of sheriff, county clerk, treasurer, and probate court judges.

Existing research suggests that state legislatures, which in South Carolina deliberate on issues including education spending, environmental regulation, and abortion, to be as polarized as Congress.[6] However, other offices tend to focus on administrative matters (Oliver, Ha, and Callen 2012). County councils are legislative bodies that often discuss transportation infrastructure, public facilities, and sales taxes. Sheriffs are the chief law enforcement officer and manage county jails, auditors calculate millage rates, treasurers collects taxes and oversees the disbursement to other jurisdictions, the clerk of court manages court dockets and manages the collection of fines and fees, and coroners perform independent investigations of deaths. In the judicial branch, circuit solicitors (known as district attorneys in other states) serve as the chief prosecutor of state government, and probate court judges have jurisdiction over civil cases such as estate inheritance. Despite their administrative functions, all of these offices are directly elected through partisan elections in general elections. Almost all candidates register for the Republican or Democratic party and win a party primary to be elected.

---

[6] According to legislator ideology estimates from Shor and McCarty (2011), during 1996 to 2009, the spatial gap between the median Democrat and median Republican in the South Carolina State House was about as large of the spatial gap between members of Congress. Updated data from Shor (2018) during 2010-2016 shows that the gap in South Carolina has grown about 20 percent.

While South Carolina is a solidly Republican state in national elections, Democrats win seats at considerable rates on the long ballot. For example, just over a half of all countywide executive offices elected on a partisan ballot in 2016-2018 were Democrats (Appendix A.2 tabulates the result of nearly 10,000 partisan contests in the state over the past four decades). The Republican party has controlled the governorship and the state legislature since 2003, and every Republican presidential candidate since 1964, except Gerald Ford, has carried the state. But alongside these Republican victories, the same South Carolina general election electorate has voted in a considerable number of Democrats in state and local offices, well after the critical elections associated with realignment and nationalization.

The available survey evidence suggests that straight ticket voting in South Carolina is comparable to the national average. Among the 4,512 respondents in the CCES between 2010 and 2018 voting for major party candidates in the state, 93 percent voted for the same party between the Presidency and the U.S. House, 91 percent between the U.S. Senate and the Governor, and 92 percent between the U.S. House and Governor. All three numbers are within one percentage of their respective national average (n = 318,346).

3.3.4 Additional Candidate Attributes

After measuring the prevalence of ticket splitting across the long ballot, I then combine this information with information about the candidates in each race. Through web campaign filing reports and old versions of county websites, I mark the incumbency status of each candidate in my dataset. Other than incumbency, systematic information about both winning and losing candidates in local elections is sparse. I collect additional data from two sources — media coverage and campaign finances, which both measure other aspects of valence.

I further collect candidate data as measures of valence. Media coverage proxies for name recognition and the amount of campaign contributions a candidate raises proxies for candidate effort and candidate quality (Prat, Puglisi, and Snyder 2010). For media coverage, I search a newspaper database that has ample coverage of state and local newspapers, and count the number of articles in South Carolina's 86 state newspapers that mention their full name during the length of the term for that office. In total, this search accounts for 356,209 article hits across 764 candidate-election combinations. A detailed description of this search is left to Appendix A.1. I also record the amount of dollars each candidate has raised during the election cycle, from campaign finance data collected from the Federal Election Commission for federal offices and the State Ethics Commission for state and local offices. The procedure and potential sources of measurement error are again documented in Appendix A.1.

3.4 Party Loyalty on the Long Ballot

A descriptive analysis of the pattern of votes can start to rule out several hypotheses. If candidates and voters were thoroughly nationalized, straight ticket rates should be equally high in every office. And if voters chose candidates based on party and valence but information about a candidate's valence attributes was harder to come by in state and local races, straight ticket rates should be higher in state and local offices. I present three sets of analyses: straight ticket voting rates at the voter level, the office level, and finally an analysis of the principal dimensions of vote choice.

3.4.1 Voter-Level Straight Ticket Voting

Throughout this analysis, I refer to *straight-ticket voting* as the action of choosing candidates of the same party for *all* partisan offices under consideration. There are two subtleties to this operationalization. First, uncontested races do not offer voters a real choice to either vote straight or split ticket, and therefore I will limit my analysis to contested races. Throughout I will use *contested* to mean that the contest features both a Democratic and Republican candidate. For example, if a contest features a Republican candidate, a Green party candidate, and a Libertarian candidate, I still count that as an uncontested race. In the discussion, I consider the implication for this restriction when generalizing to voter behavior in other districts.

Second, South Carolina is one of the few states in which voters have an option to explicitly cast a straight ticket. A voter can either click through the entire touch-screen ballot, or he can select the "Straight Ticket Party Option" that appears as the first question on every ballot (See Appendix A.1 for an example). To avoid confusion of terms, I refer to this latter option as using the *party lever*, a slightly dated phrase originating from the era when voters pulled a physical lever on a voting machine to the same effect.[7] Pulling the lever for a particular party auto-fills the voter's ballot to select that party's candidates for every applicable contest. These selections are reversible case-by-case before the ballot is cast, and in my dataset I find slightly below 3 percent of voters who use the Republican or Democrat party lever later switch their vote in a contested race.

The Appendix Table A.4 shows the prevalence of straight ticket voting in my entire dataset. The number of contested races on a voter's ballot can range between 1 to 12. I compute the proportion of straight ticket voters among those contested choices only, and show the proportion as well as the general distribution of party loyalty. In the modal case of a ballot with 5 contested races, 77 percent of voters are straight ticket voters. The proportion drops to the 60s among voters who happen to face a ballot with more contested races. These numbers arguably inflate the proportion of full

[7] States have gradually discontinued the party lever. In 2018, only Alabama, Indiana (except for at-large races), Kentucky, Oklahoma, Pennsylvania, South Carolina, Texas (until 2019), and Utah used the party lever.

straight ticket voting because it includes those who pulled the party lever and likely gave less consideration to each office. Among the half of the electorate that opted out of the party lever, the prevalence of straight ticket voters is about 10 to 30 percentage points less than the full sample.

3.4.2 Party Loyalty by Office

Are defections from a straight party ticket more prevalent in some offices than others? As a first-order description, Figure 3.2 sets the office at the top of each election's ballot as the reference category and shows the overall rate of split ticketting by office. In red are the rates of voting among "Republican" voters: those who voted for Romney, Trump, or the gubernatorial candidates Haley or McMaster, depending on the year. In blue are "Democratic" voters who voted for Obama, Clinton, or the gubernatorial candidates Sheheen or Smith. For example, the figure shows that among voters who voted for a Republican President or Governor candidate in a congressional district where the House race was contested, only 4 percent of them split their ticket, or voted for the Democrat.

Because Figure 3.2 does not make within-person comparisons, I use a tailored clustering algorithm to summarize the data into interpretable prototypes of voting patterns while still leveraging the full distribution of voting patterns that the cast vote records reveal. Clustering is an attractive approach several reasons. Like ideal point estimation, clustering efficiently incorporates data in which the same individual makes multiple choices, and it can do so even when there is no information about the individual other than their choices (as in cast vote records).[8] In contrast, simple comparison of ticket splitting rates by office will treat votes for each office separately without

8 Ideal points have also been used to analyze voting matrices likely this, but it imposes a spatial model of vote choice that may be may be less appropriate for voters' preferences than it is for legislators' rollcall votes (Broockman 2016). Moreover, ideal point methods often use hundreds of votes and lack convergence properties with fewer votes, which is the setting here.

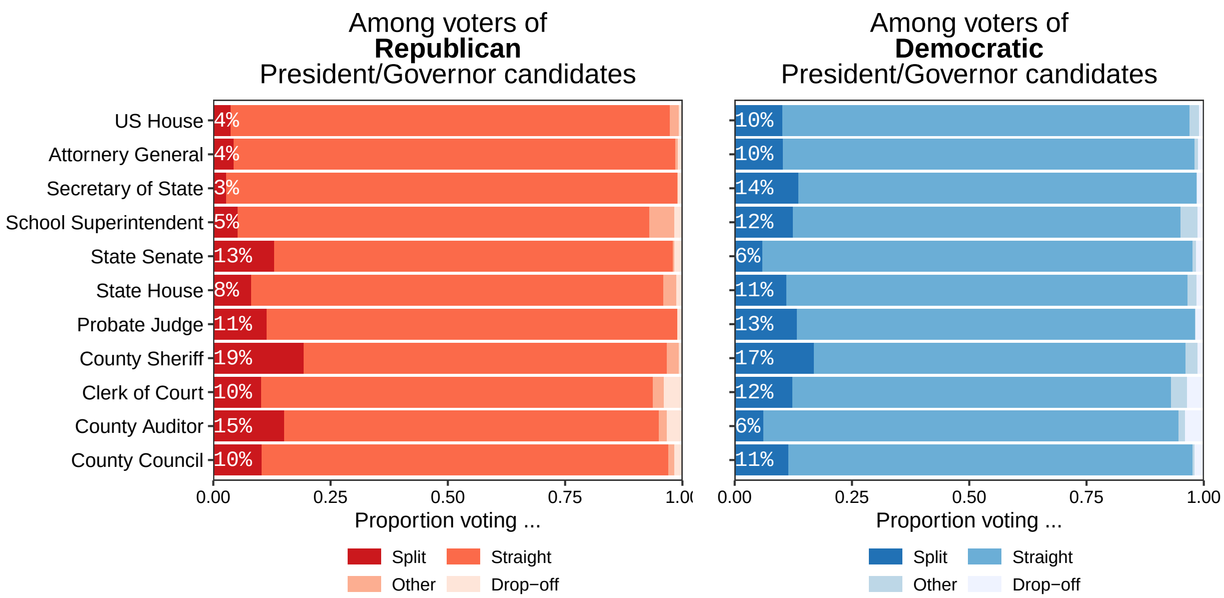


Figure 3.2: Straight Ticket, Split Ticket, Third Party, and Abstention Voting

Note: Each bar marks the proportion of votes for a party relative to a voter's vote for President or Governor. Red bars are proportions among voters who voted for a Republican presidential or gubernatorial candidate. Blue bars are proportions among those voting for the Democrat. Proportions are only computed among contested races in South Carolina general elections, 2010 - 2018. The sample sizes for each election year - office combination are shown in Appendix A.3.

leveraging the multidimensional nature of the data, and compares a different set of voters depending on the office.

In this clustering method, the user picks the number of clusters to divide the voters into, and a fast Expectation Maximization (EM) algorithm identifies the set of cluster assignments that best fits the data. Formally, I posit that each voter i belongs to one of K clusters, but that cluster membership Z_i is unobserved. Instead we only observe a vector of J choices $\mathbf{Y}_i = [Y_{i1}, ..., Y_{iJ}]$ for each voter: a straight ticket, split ticket, or abstain for each of the J offices on the long ballot. In its simplest form, a clustering algorithm uses only the matrix $\mathbf{Y}$ and a simple model of vote choice to estimate two quantities: The overall prevalence of cluster k:

$$\pi_k = \Pr(Z_i = k), \text{ where } \sum_{k=1}^{K} \pi_k = 1,$$

and the probability that a member of cluster k votes for choice ℓ in office j:

$$\mu_{kj\ell} = \Pr(Y_{ij} = \ell \mid Z_i = k), \text{ where } \sum_{\ell=1}^{L} \mu_{kj\ell} = 1.$$

The model of vote choice underlying this representation is that a vote in one office is independent of each other, *within* each cluster, i.e.,

$$\Pr(\mathbf{Y}_i, |Z_i = k, \pi) = \prod_{j=1}^{J} \prod_{\ell=1}^{L} (\mu_{kj\ell})^{\mathbf{1}(Y_{ij}=\ell)}.$$

This also serves as the identification assumption to estimated the parameters in the clustering algorithm.

The algorithm is designed to be tailored to three features of the ballot data, with formal derivations in Chapters 5 and C. First, to account for abstentions and third party votes, outcomes are allowed to be unordered categorical variables. Second, unlike a canonical clustering model, I account for the fact that uncontested races provide voters with a limited pool of candidates, the method incorporates these varying choice sets by an independence of irrelevant alternatives assumption. Third, to handle over a million votes, the algorithm uses a C++ backend to perform internal calculations quickly.

Figure 3.3 shows the point estimates of the cluster prevalence π_k and the characteristics of each cluster $\mu_{kj\ell}$ estimated with four clusters, in the two presidential races available in the data. Each vote is recoded relative to the same voter's presidential vote, and, as in Figure 3.2, subsetted to voters who voted for the Democratic and Republican Presidential candidates.

Both the choice of the number of clusters (K) and the substantive interpretation of each cluster is determined by the user. Although this leaves room for some ambiguity when implementing the clustering algorithm, one does not need to commit to the view that there exists a single correct number of clusters in the data. One cluster can often be divided into two slightly more homogeneous clusters. To provide some guidance, I cluster the same data with values of K between 2 and 10, and compute the BIC fit statistic. I ultimately choose to present results with $K = 4$ given that is where the fit statistics start to level off in 2016 (Appendix A.3). The BIC statistic uses the observed log likelihood that the EM algorithm tries to maximizes, penalized by the number of parameters it is asked to estimate. I then provide a label for each of the clusters according to the values of the estimated values of the vote choice parameters μ.

In 2016, a bare majority of both Republican and Democratic voters (as inferred from their presidential vote choice) are solid partisans in their votes, because they vote solidly for the same party up and down the ticket. In 2012, only about 40 percent of the electorate is classified as solid partisans. Even in 2016, this group is not large enough a group to deliver a election-winning majority for a particular candidate, as Romney and Trump comprised about 55 percent of the state's electorate.

The second largest cluster of voters vote for the same party in congressional races but are more likely to split their ticket in state elections. This pattern is particularly noticeable among Republican voters, where for example 5 percent of Trump voters in cluster 2 split their ticket in the U.S. House but 15 to 50 percent of them split their ticket for the Democrat in their vote for the state legislature, sheriff, and county council. This cluster comprises about 35 percent of both Trump and Clinton voters, which makes them large enough a group to be pivotal even in a statewide race. The third and fourth largest cluster of voters vary in their voting patterns by year and party.

Many of these groups primarily abstain after voting for President, while a the fourth cluster among Clinton voters appears to be solid Republicans who only broke away from their party preference in the race for President.

Finally, the clustering differentiates between ticket splitting for particular offices. Among 2012 Obama voters, for example, two out of its four clusters had substantial probabilities of splitting their ticket, but while the second largest cluster was most likely to split in the vote for State House, the third largest cluster was more likely to split in the office of Sheriff.

Therefore, despite one line of reasoning that would predict straight ticket voting to be more prevalent in down-ballot races where candidate specific information is scarce, election day voters tend to defect from their national party loyalty as much as, if not more than, national congressional races. These cast vote records reveal new patterns of voting behavior that have been not possible to measure in existing surveys and election returns.

To summarize, both Figures 3.2 and 3.3 exhibit the same general pattern. First, it is not the case that voters vote more straight ticket in state and county offices than they do for Congress. Most straight ticket rates in state legislative and county executive offices are lower than those for Congress, and especially so for sheriff and county council races. For example, while 94 percent of top of the ticket Republicans voted for a Republican congressional candidate, only 77 percent of them in contested sheriff elections voted for a sheriff. Accordingly, ticket splitting is more prevalent for the majority of these state and county offices. Roll-off is also higher further down the ballot: typically around 1 percent in congressional elections and 2 to 4 percent in down bal-

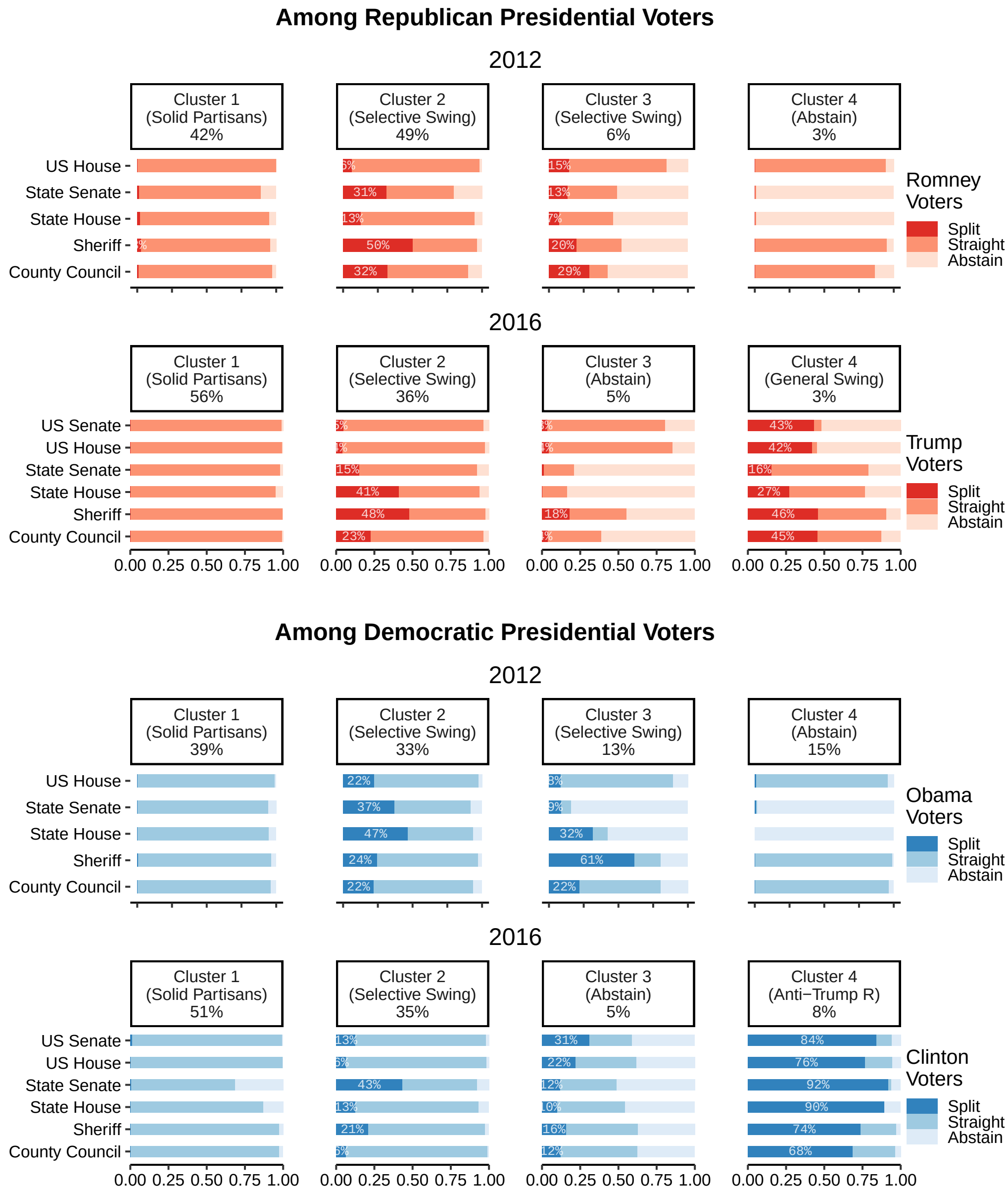


Figure 3.3: Rates of Ticket Splitting in 2012 and 2016 Votes

Note: Figures show estimates from the clustering method. Each row of facets represents one population of voting data, and each facet within each row represents the estimated cluster. Clusters are numbered by their estimated size in the population (parameter π_k), as shown in the facet label. Within each cluster, the algorithm estimates the estimated probability that a voter in that cluster votes a certain way in a given office (parameter $\mu_{kj\ell}$). Vote choices are recoded so that they are relative to the voter's Presidential vote choice. Therefore, this figure analyzes South Carolina voters who voted for a major party Presidential candidate.

lot races do not cast a vote.[9] Yet the statewide executive offices of Attorney General and Secretary of State are notable exceptions. The straight ticket voting rate in these offices are as high as those in congressional elections.

Figure 3.4 next highlights how this individual-level variation manifests at the level of electoral districts. This level of analysis is important because that is where elections are won or lost. Put another way, the prevalence of ticket splitters may not be relevant if they are so thinly dispersed that they are not pivotal anywhere. The first panel of the Figure displays comparisons involving the office at the top of the ticket. Each successive distribution shows that a district's straight-ticket voting rates are slightly lower and clearly more variable further down the ticket. Rates of straight ticket voting between the top of the ticket and Congress, the state legislature, and county councils all have a mode at around 90 percent, but each distribution has successively fatter tails: Straight-ticket rates involving sheriff are sometimes as low as 50 percent. The second panel compares state offices with each other, instead of pegging each comparison to the President or Governor vote. Between these races, the rates are even more variable, ranging from 25 to 100 percent.

3.5 Incumbency and Split-Ticket Voting

The analyses so far show clear variation in the proportion of straight ticket voting. What, then, explains that variation? The past work on local politics and ticket splitting suggest that incumbency is a natural factor to inspect. Incumbency is the primary indicator that proxies for, or at least correlates with, the whole bundle of these valence attributes. Some component attributes like name familiarity and campaign

[9] This rate is smaller compared to those reported in other studies of roll-off, which show roll-off to be about 5 - 10 percent. However, most of these other studies examine non-partisan elections or ballot measures. Additionally, the values in Figure 3.2 take voters who have already voted for a major party at the top of the ticket as its denominator, and in South Carolina the presence of the party lever likely decreases roll-off in partisan contests.

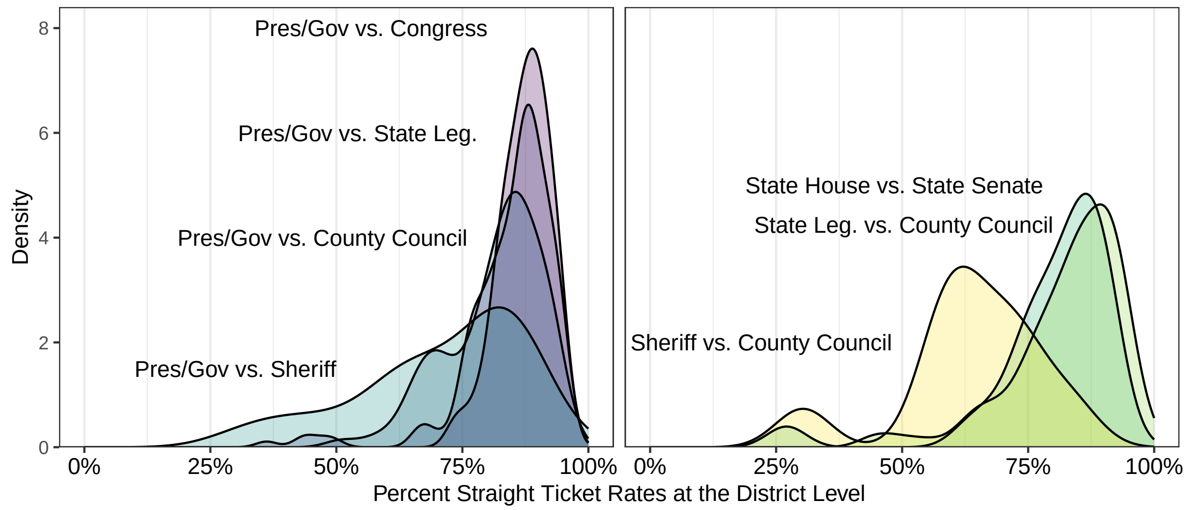


Figure 3.4: Straight Ticket Voting at the District Level

Note: Each density curve summarizes the distribution of straight-ticket rates at the district level for a given class of pairs of offices. The types of offices are labeled adjacently to each density curve. The plots separate out pairwise straight ticket rates that involve a President or Governor's vote (left) and those that do not.

salience can be measured by the newspaper coverage and campaign finance reports.

3.5.1 Difference in Means

Access to the full ballots allow some straightforward calculations to estimate how much ticket splitting is explained by incumbency. I first count the fraction of split ticket votes that were cast towards the incumbent. I subset the data to six offices with a sufficient number of contested contests in enough districts, and further subset to contests which featured an incumbent running for re-election against a major party challenger. After these restrictions, we are left with 566,232 split ticket votes (as defined in Figure 3.2) cast by 495,138 voters. In each of these split ticket choices where the voter could choose between the incumbent or the challenger, a clear majority of 69 percent voted for the incumbent.

Open-seat contests serve as an additional useful comparison because incumbency is not at play. In Table 3.2, I show the proportion of the straight ticket voting rates

Table 3.2: Straight Ticket Rates by Incumbency

	Contests with an incumbent			Open contests		
	Voters for whom same-party candidate is ...					
	(i)	(ii)		(iii)		
	The Incumbent	The Challenger		No Incumbent		
Office	Vote same-party	Vote same-party	Diff. (i) - (ii)	Vote same-party	Diff. (i) - (iii)	Voters
U.S. House	0.94	0.87	0.073	0.87	0.074	5,793k
State Senate	0.93	0.80	0.128	0.88	0.052	541k
State House	0.93	0.81	0.124	0.82	0.108	1,695k
Probate Judge	0.94	0.78	0.158	0.86	0.080	494k
Sheriff	0.92	0.66	0.260	0.77	0.150	319k
County Council	0.90	0.82	0.082	0.83	0.072	470k

Note: Proportions show straight party voting rates between six offices and the top of the ticket (President or Governor) by candidate incumbency. The difference between (i) and (ii) indicate the difference in straight ticket voting associated with incumbency (as opposed to being a challenger). The difference between (i) and (iii) indicate the difference between an incumbent and a race with no incumbency on the ballot. n indicates number of voters in 1000s.

separated by the presence of an incumbent and the party affiliation of the incumbent. If voters value the qualities associated with incumbency, we would expect to see the most same-party votes when (i) the incumbent is of the same party as the voter's top of the ticket choice, and fewer same-party votes when that (ii) entails voting against the incumbent. Finally, the rate of straight ticket voting when (iii) there is no incumbent should be lower than case (i) but higher than case (ii).

Consistent with those expectations, in all of the six offices covered in Table 3.2, straight ticket voting is highest when doing so coincides with voting for the incumbent. Among contested U.S. House races, 94 percent of voters whose party choice at the top of the ticket happened to align with the party affiliation of their U.S. House incumbent voted for that incumbent (column (i)). But when they did not align (column (ii)), only 87 percent of these voters voted straight ticket, indicating a split ticket to vote for the incumbent. The rate in open-seats where no incumbent exits (column (iii)) tends to fall in the middle of the two values for all offices. If voters did not value the qualities associated with incumbency, all three proportions for each office would have been the same. Instead, we see sizable differences.

3.5.2 Contributions of valence controlling for partisanship

An individual-level regression allows for a more controlled comparison, modeling vote choice after matching individuals of similar revealed preferences in national offices. For each voter i making a choice for race j on their ballot, ticket splitting can be modeled from a linear probability model of the form

$$\mathrm{Split}_{ij} = \alpha + f(\mathrm{D}_i) + \gamma_0 \mathrm{V}_{d[ij]} + \beta D_i \mathrm{V}_{d[ij]} + \varepsilon_{i,d[ij]}. \tag{3.1}$$

Here the binary outcome Split_{ij} is 1 if individual i votes for the same party in office j as they did for at the top of the ticket. The control D_i measures the voter's general party preference towards the Democrat by the proportion of times a voter chose the Democrat in the top of the ticket contests available: President, Governor, U.S. Senate, and the party lever. The function f allows for this measure of partisan preference to vary flexibly; here I use a second-order polynomial. The model interacts this with valence, denoted by $\mathrm{V}_{d[ij]}$, which I code so that positive values indicate the valence advantage of the Republican candidate, indexed by the district of office j in which voter i resides ($d[ij]$). Therefore I cluster standard errors by assuming errors $\varepsilon_{i,d[ij]}$ are potentially correlated across individuals within a district.

Three variables operationalize the valence advantage. The first is an indicator for whether or not the Republican candidate in a (closed) contest is an incumbent. To measure other aspects of valence that can be defined in open-seat races, I also use the Republican advantage in newspaper coverage and in campaign contributions. As de-

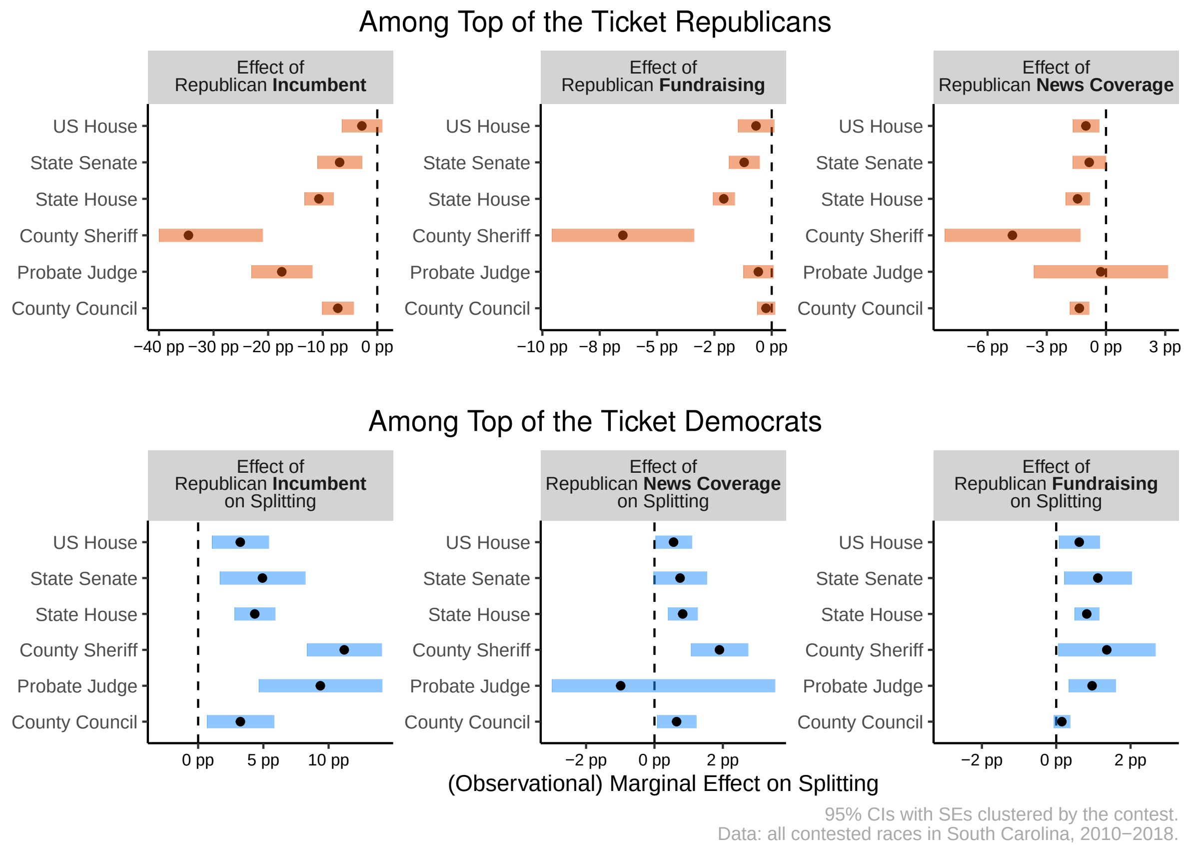


Figure 3.5: Contribution of Valence Attributes to Ticket Splitting

Note: Each facet shows marginal effects (on the probability scale) from a separate regression, each following equation 3.1. Confidence intervals indicate 95 percent interval using standard errors clustered by race. All regressions control for individual-level party preference as measured by available votes for President, U.S. Senate, Governor, and the party lever.

scribed in Appendix A.1, I take the ratio of the Republican candidate's metric over the Democrat's so that positive values indicate a Republican advantage, and take the natural log of the ratio to model diminishing returns of newspaper coverage and campaign contributions. Both measures may be endogenous to incumbency, so I take each measure of valence one by one.

Figure 3.5 presents the marginal effect estimate for each of the three measures of valence. Because the effect of a Republican valence advantage likely moves moderate and extreme voters to varying extents, I control for each individaul's voteshare at the top of the tickeet (where no candidate variation exists) and estimate marginal effects at each endpoint of the distribution of proportions, where most of the data lies. If voters are complete straight-ticket voters, their straight ticket voting would be completely explained by their top of the ticket choices and all other coefficients would be indistinguishable from zero. Conversely if voters could perceive and value candidate's valence even conditional on their national party preference, the valence advantage of the Republican should attract ticket splitting from Democrats, even after controlling for each voter's national partisanship.

The left panels of Figure 3.5 consistently shows that incumbents draw more split ticket votes. Each point represents the difference in the probability of deviating from the top of the ticket when the Republican is the incumbent, compared to when the Democrat is the incumbent. The top panel with red confidence intervals shows effects among Republican voters (i.e., voters who voted for all Republicans for statewide races). The negative estimates, interpreted causally, suggests that Republican voters *increase* their support of the Republican downballot candidate when that candidate is an incumbent. The bottom panel with blue confidence intervals shows the effects of a Republican valence advantage among *Democratic* voters. Positive values here indicate that Democratic voters for whom the Republican candidate in a particular office is an incumbent are more likely to split their ticket for that Republican, compared to if the Republican was not an incumbent (i.e., an open race or a race with a Democratic incumbent). Because these leverage cross-district comparisons, the standard errors are clustered by the district and are wide. However, the point estimates are large enough to distinguish almost all estimates for incumbency with zero.

The coefficients for newspaper coverage and fundraising advantage in the middle and right rows of Figure 3.5 show similar patterns of ticket splitting. With the exception of the effect of newspaper coverage on probate judge vote choice, when the

Republican candidate has more newspaper coverage or more contributions than the Democratic candidate, they tend to attract more votes from Democratic leaning voters. Because the outcome variables are binary, the Republican advantage is the log of ratios, and the coefficient estimates in a given panel are roughly symmetrical around 0, the converse holds as well. A one-unit increase on a party's log ratio advantage measure is associated with about a 2 to 3 percentage point increase in ticket splitting in the opposing party.

These estimated effects of incumbency in state and local races persist for several offices even after controlling for newspaper coverage and the campaign fundraising, with multivariate regression results presented in Appendix A.3. These findings suggest that incumbency is not merely a proxy for name familiarity and news coverage. Although the multivariate regressions cannot pinpoint the specific mechanism at play, it suggests that voters both can perceive and care about the range of factors that originates from experience on the job and other reputational advantages.

In summary, these analyses show that majority of ticket splitting is cast in favor of the incumbent in state and local offices, even though the ballots do not include incumbency or any other information about the candidate. The deviations from straight ticket voting in the previous section are not arbitrary, but systematically benefit candidates who have relevant experience as incumbents, are better known, and raise more campaign funds from voters.

3.6 Generalizability

One limitation of these findings is that they examine contested races in a single state. In Appendix A.4, I analyze a smaller set of cast vote records in two states – Maryland and Florida, and find similar patterns of higher ticket splitting in state and local offices. Other than that, several considerations suggest that the main implications here are generalizable to most contexts in contemporary American politics.

South Carolina state legislative seats are one of the least contested in the country. For example, according to *Ballotpedia*, only 30 percent of state legislative districts were contested in the 2012 general election, putting the state's competitiveness index only ahead of Georgia and Massachusetts. But it is reasonable to expect that split ticket rates would be *higher* if parties contested more districts. Districts with no challenger tend to be those where the disadvantaged party's chance of victory is slim to begin with (Rogers 2015). Therefore, voters who value the correlates of incumbency should be even more likely to cross party lines if a disadvantaged party were to enter a lopsided race.

When extending to other states, the findings here suggest that ticket splitting would be less prevalent in down ballot races where candidate-specific information is sparse, two-party competition is high, and the incumbency advantage is weak. One might worry that South Carolina is an outlier in this regard: An uncompetitive state consisting of lopsided districts. But as Fraga and Hersh (2018) showed using congressional, statewide, and state legislative elections, it is rare for any single voter to reside in that sort of enclave. South Carolina is no exception. The long ballot and the high degree of district overlap in the U.S. electoral system all but ensure that most voters' long ballots feature competitive contests as well as uncompetitive ones.

Another concern when predicting patterns in other states is South Carolina's history as a Southern State, where Democrats such as Strom Thurmond switched to the Republican party in a massive realignment in the 1960s and 1970s (Mickey 2015; Key 1948, also documented in Table A.2). It is likely that some of the pattern here is driven by older voters who, like in the election of Dwight Eisenhower, voted for Republican national candidates but Democratic candidates in state and local candidates. On the other hand, realignment is a common feature of a two party system and many other U.S. states outside the South experienced realignments, if not to the same degree. Moreover, the logic of the incumbency advantage and valence does not rely on such massive realignments.

I now finally turn to the more speculative question of whether the dynamics of split ticket voting in state and local races documented here will eventually disappear in an era of increasing nationalization. Most studies describe nationalization as largely a top-down process (Abramowitz and Saunders 1998; Aldrich 2000). The electoral trends over the past 40 years also document a consistent trend towards Republican dominance in the state. But this realignment did not impact all levels of elections at once, nor did it progress at the same rate (Appendix A.2). Within the set of elections collected in my dataset, I do find an uptick in the rates of straight ticket voting overtime (Appendix A.3), but how these rates would change beyond the sample depends on the future party alignment in national politics, something beyond the scope of this paper. My results do suggest that the incumbency advantage is still a significant force in state and local elections and may delay the tides of nationalization that has swept congressional and gubernatorial elections.

3.7 Conclusion

The picture of the electorate that emerges from these analyses is one that votes largely along party lines, but still with important variations across offices and candidate's incumbency. A new dataset that provides an unprecedented view into voters choices showed that about seven out of ten voters in South Carolina vote a complete ticket, and in any given office, about eight to nine out of ten voters vote the same party as the President or Governor. Ticket splitting is especially prevalent in sheriff contests, and state and local contests are more varied in their level of split-ticket voting.

Should we consider the statistic that 80 percent of Trump (Clinton) voters voted for a Republican (Democratic) county sheriff candidate to be a large or small number? On the one hand, as I have suggested, this is smaller than what we would expect from a fully nationalized politics. A district in which 20 percent of voters is up for grabs is by most measures a volatile one. On the other hand, there are also good grounds to interpret party loyalty of such degree as too high. The traditional view of local politics has been that it is void of partisanship altogether, with recent work updating that view (Tausanovitch and Warshaw 2014; Bucchianeri 2020). Another study of close elections between Democratic and Republican sheriffs finds that party control does not cause changes in how sheriffs implement policy (Thompson 2019), suggesting that partisan splits in voter's supports for that office is disconnected from the policy preferences of the candidate.

Regardless of one's interpretations, the findings presented here would not have been obvious without empirical investigation. Extending the nationalization literature into state and local politics, one might have predicted that the rate of straight ticket voting to be equally high for all pairs of offices. And extending theories of partisanship and political communication, one might have predicted that, if anything, straight ticket voting rates would be *higher* in down ballot races than in national ones because voters would have less candidate specific information to inform their choices for county council than they would for the U.S. Senate.

The first part of the empirical findings do not lend support for these predictions, at least for the average voter and the average election. The second part starts to reveal why. Incumbents systematically netted more votes from party defection than challengers or contestants in an open race (Table 3.2), even after controlling for a voter's own party allegiance (Figure 3.5). This pattern is consistent with a model of elections with nationalized partisan candidates in state and local offices with different levels of experience or quality. In other words, even if we assume that nationalization has so thoroughly polarized *candidates* such that even candidates for state and local offices follow the national ideological platforms of their respective parties (Hopkins 2018; Shor and McCarty 2011), some voters know and care enough about non-ideological aspects of the candidates such that they split their ticket.

By constructing the first database of cast vote records spanning an entire state across multiple elections, this study has overcome some common challenges researchers face in studying vote choice for state and local office. Cast vote records will prove valuable for understanding electoral behavior more widely. In addition to ticket splitting, they allow researchers to study vote choice in party primaries, ballot measures, and elections for non-partisan offices such as school board elections. Existing studies of these three types of elections are limited by the same sort of measurement problems that surveys and election returns have for studying ticket splitting, and therefore can benefit from wider use of cast vote records.

The findings of this paper are not without their limitations. Cast vote records reveal how people vote in state and local offices, but they reveal much less about the demographic characteristics of those voters. And partly because of this limitation of survey or demographic evidence, it becomes difficult to disentangle potential mechanisms underlying the findings, i.e., a valence advantage, candidate moderation, or multidimensional voting. Future research that combines cast vote records with precinct-level data could help distinguish more carefully the process through which voters form their preferences for state and local offices. This paper does, on the other hand, establish some baseline expectations for state and local elections in a nationalized politics. On election day, U.S. voters must make a series of choices with limited information beyond party labels. But after an accumulation of campaign outreach, media coverage, and information acquired through everyday observation, a considerable number of voters deviate from a complete straight ticket vote.

4 | Target Estimation for Weighting to Small Areas: A Validation and Open Source Workflow

Abstract

Estimating population target distributions from incomplete data is a significant practical barrier to survey weighting, especially when researchers are interested in making inferences on small subgroups. Yet recent survey research tends to focus on methods to fit complex outcome models, instead of methods to expand the set of poststratfication variables. Does poststratifying to additional variables in fact improve estimates? If so, which variables matter, and how should researchers estimate joint distributions of variables beyond what is publicly available? I answer these questions by proposing a method to expand publicly available three-way tables into six-way joint tables including turnout, simultaneously calibrated to known marginals instead of fixing post-estimation. I find that poststratifying on standard demographics does not improve MRP estimates of vote share at the congressional district level. This may be because the survey I use already adjusts for these population targets in its initial sample matching. The proposed target estimation improves estimates by about half a percentage point (from a root mean square error of 8.0 to 7.7 points), and poststratifyng on party registration improves them further by about 1 to 2 percentage points (to 5.9 points).

* I acknowledge the support of NSF Grant 1926424 and thank Steve Ansolabehere, Andrew Gelman, Yair Ghitza, Lauren Kennedy, Jonathan Robinson, and especially Soichiro Yamauchi for numerous discussions on the findings related to this chapter. I thank Douglas Rivers, Eddie Mertz, and Brandon Bertelsen for sharing the summary statistics from YouGov's database to enable extensions.

Surveys continue to be the main way through which scholars study electoral behavior. As survey samples have become increasingly larger, social scientists have turned to estimating quantities at particular subgroups of the entire data (Broockman and Skovron 2018; Kalla and Porter 2020; Hertel-Fernandez, Mildenberger, and Stokes 2019). At the same time, the political survey community has also become more conscious about selection bias and unrepresentativeness in these data. Accurate subgroup estimates of voter behavior at the state and legislative district level are crucial for studies of electoral politics like the one I explore in this dissertation.

As ubiquitous as the use of pollster's survey weighting is, however, the construction of weights is still an open discussion in social science research for which little guidance exists. The observation that "survey weighting is a mess... the construction of weighting itself is an uncondified process" (Gelman 2007) still rings true. The situation has undoubtedly improved, with pollsters documenting their own complex weighting process (Ansolabehere and Rivers 2013) and review texts that connect weighting methods in a single framework (Caughey et al. 2020). However, many of the methods are still out of reach for applied researchers who want to adjust existing weights to their own subgroup of interests.

Moreover, the suitability of a set of weights is not a black-and-white issue. Because much of the validity of an estimated weight depends on the quality of imperfect data, an empirical accounting of how well a set of survey weights adjust a sample to various geographies is therefore necessary for applied researchers.

Methods for reweighting, and the estimation of synthetic population targets that are required to enable to such a weighting, has wide applications to modern survey research. For example, it is crucial for improving Multilevel Regression and Poststratification (MRP) models as well as non-MRP estimates. MRP combines the traditional study of shrinkage and partial pooling that is mostly concerned with variance reduction with standard infrastructure for poststratification weighting that is mostly concerned with bias reduction (Gelman and Little 1997). Much of the recent research on MRP has exclusively focused on the former: improving the model that induces partial pooling in the outcome. It has held constant the poststratification table constant, often with off-the-shelf Census datasets that do not include the variables pollsters typically use. Improving the estimation procedure for population targets for subnational geographic units can benefit both traditional weighting and any MRP model.

Here I conduct a validation and outline an open-source method that encompasses all these aspects through a concrete example, the Cooperative Congressional Election Study (CCES). I discuss the sampling and small area problem in the CCES for congressional districts, where the survey sample for each district is only around 50 respondents. Specifically, I propose a workflow to construct a poststratification target that approximates the joint distribution of six standard variables: age group, sex, education, race, turnout, and congressional district (which are nested in states). A multinomial logit model with simultaneous calibration properties implemented by Yamauchi and Kuriwaki (2021) allows this joint estimation, which is more scalable and has better theoretical guarantees than existing attempts for using synthetic distributions in MRP (Leemann and Wasserfallen 2017; Ghitza and Steitz 2020).

I then show how such a reweighting improves the estimates of vote choice at the congressional district level in the 2016 CCES, while off-the-shelf poststratification with only a few demographic variables does not noticeably improve the aggregate error of estimates relative to a simple raw average. Partial pooling alone, which precedes the postratificaiton step in MRP, also does not improve the overall accuracy of the estimates. The proposed workflow is open-source and draws from datasets that can be downloaded by a user-friendly interface (Kuriwaki 2021a). In an extension, I show how non-public data such as party registration statistics in voterfiles can be used to further

Table 4.1: Sample Sizes in Modern Surveys

	CCES Sample		Pew Sample		Population	
	n	$1/\sqrt{n}$	n	$1/\sqrt{n}$	N	Trump
United States	64,600	0.3pp	2,583	2.0pp	323 million	49%
California	6,021	1.2pp	259	6.2pp	39 million	32%
California's 48th Congressional District	95	10.2pp	Not available		0.7 million	46%

Note: The CCES is the 2016 Pre-election Survey. The Pew Sample is the October 2016 Political Survey. n indicates the sample size and $1/\sqrt{n}$ is a rough estimate of the standard error around a proportion from a simple random sample of size n. Weighting will often lead to larger standard error. This table illustrates that the CCES has direct measures of congressional districts, but still suffers from a large standard error.

reduce nonresponse bias.

4.1 The Rise of Online Surveys and Calibration Weighting

As the typical sample size of data have grown larger through technological innovation, one might think that survey researchers are no longer befuddled by small sample problems. But this is not so for two main reasons. Even with large datasets, scholars have turned to estimating population quantities at smaller and smaller subnational geographies. Table 4.1 compares the sample sizes of two common datasets at the national level, state level, and sub-state congressional district level. Even with a survey like the CCES which is an order of magnitude larger than the typical national poll, there are fewer than a hundred observations from a given congressional district (which represents more than half a million people). Second, as sample sizes have become larger, response rates have also plummeted, raising the danger that the survey samples we do collect are less representative (Meng 2018).

This work contributes to recent literatures in political science and survey statistics that has arised to keep up with the technical realities of polling. Three bodies of work are particularly relevant. Applied examinations of calibration weighting, recent statistical innovations in estimating calibration weights, and the existing small area estimation literature in political science which has largely focused on MRP.

An overview of weighting methods appears in Caughey et al. (2020). They identify the problem of target estimation as a challenging task, for which "how best to approach the problem is still an open question and a subject of ongoing research." I provide such an extension in estimating synthetic population data for a turnout electorate. I also focus on the issue of small area estimation, a topic Caughey et al. only discuss in passing and leave for further research. A concrete description of poststratification in the CCES is given in Ansolabehere and Rivers (2013). However, their benchmarks to election results and benchmarks stop at the state level, where survey samples are large (about 1000 respondents) and the weighting specifically target demographic distributions at that level. In this chapter, I investigate smaller areas of geography that the pre-computed weights are not adjusted to.

Target estimation and calibration weighting is a broad field, featuring classic studies that have enabled now standard tools such as rake weighting (Deming and Stephan 1940). But statistical methods in this area are continuously evolving, seeking to improve the stability of estimated weights and adding more calibration constraints to an approximation of the propensity score model. Contrary to the canonical model of inverse probability weighting typically associated with survey weighting, "survey weights are not in general equal to inverse probabilities of selection" (Gelman 2007). Instead, the population distribution that the weights target needs to be estimated itself, through a series of statistical imputation methods (Caughey et al. 2020). Because many of these constraints are not observed in practice, there is room for improved modeling (Ben-Michael, Feller, and Rothstein 2020; Zubizarreta 2015; Imai and Ratkovic 2014). This chapter draws from the insights that have recently emerged in this statistical research, summarized most recently by Chattopadhyay, Hase, and Zubizarreta (2020). The central idea is that the calibration estimation is an approximation to the true propensity model, and bias-variance trade-offs exist in choosing an optimal set of weights.

Finally, MRP is an increasingly common method for survey inference at small subgroups, especially in political science (Lax and Phillips 2009; Warshaw and Rodden 2012; Buttice and Highton 2013). While MRP is a general procedure that covers many of the practical issues in subgroup analysis, it is important to remember it is essentially "a modification of the conventional poststratification estimator" (Caughey et al. 2020, p.70) and its main innovation, the Multilevel Regression, does not directly address concerns for nonresponse bias. As I clarify in the next section, the multilevel regression stage of MRP uses a shrinkage method to deal with the high variance of small samples, but the identification assumption to validate this step is distinct from that of representativeness. Put another way, if the poststratification stage of MRP is biased or insufficient, so will MRP. MRP is also a data-intensive method. Practically all of the numerous studies that validate MRP or improve with machine learning methods innovate on the regression model and does not vary the post-stratification dataset. Even those that do (Leemann and Wasserfallen 2017) propose fairly simple methods for extending target areas, either assuming away the ecological inference problem or applying iterated proportional fitting.

4.2 Methodological Foundations of Calibration Weighting

The fundamental problem in survey inference as well as the general idea in most survey adjustment methods is shown in Figure 4.1. I will refer back to the diagram as a unifying framework for target estimation and MRP. For now, I denote individuals i in a large finite population of size N, with binary outcome Y_i and covariates $\mathbf{X}_i$. None

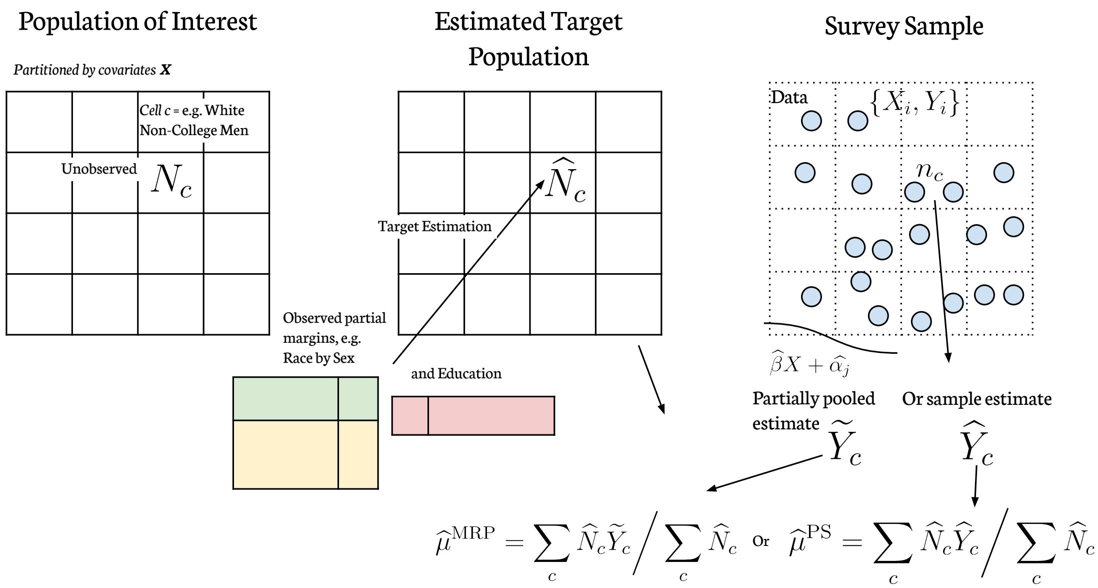


Figure 4.1: Overview of Target Estimation and Survey Reweighting

Note: The diagram represents the key steps in survey adjustment for unrepresentativeness and small samples. Observed data are shaded. The first panel shows the population of interest, partitioned into strata of covariates $\boldsymbol{X}$, indexed by cells c. Not only is the outcome Y unobserved, but the population distribution of the strata (counts N_c) is also unknown. Partial marginal distributions are observed through Census records and voter files. *Target Estimation* refers to the process of generating estimates $\widehat{N}_c$ through marginals (perhaps assisted by other data). Survey data observes with observed $\{X_i, Y_i\}$ for every survey respondent i. Cell averages are $\widehat{Y}_c = \frac{1}{n_c}\sum_{i:C_i=c} Y_i$ but some cells may be too small and have no observations. An outcome model that, for example, partially pools data from outside cells, can be fit to produce $\widetilde{Y}_c$. The post-stratified estimator is simply the survey cell average re-weighted by the estimated population cell size. Replacing the cell average $\widetilde{Y}_c$ gives the MRP estimator.

of the data in this population is observed, and a sample is drawn through an unobserved selection mechanism to make inferences. The binary variable for selection, S_i, is 1 if the individual ends up in the sample survey, and 0 otherwise. $\mathcal{S}$ denotes the set $\{i : S_i = 1\}$. I use n to denote the sample size $\sum_{i=1}^{N} S_i$.

Suppose that the primary subgroup of interest is a subnational geography, such as congressional district, which I denote with the random variable $A_i \in \{1, ..., J\}$. The subgroups need not be geographic but can easily be demographic subgroups or the interaction of the two. Geographic subgroups are of common interest for political science research and a subgroup where election results can be validated from official election results.

We denote the area of interest by j, and the quantity of interest as the population average in each area μ_j is therefore represented as

$$\mu_j = \frac{1}{N_j} \sum_{i=1}^{N} \mathbf{1}\left(A_i = j\right) Y_i,$$

where $N_j = \sum_{i=1}^{N} \mathbf{1}\left(A_i = j\right)$ is the population size for area j.

To introduce the calibration (or post-stratification) weights that are now the norm in online surveys, it is useful to begin with the classic inverse probability model. Without complete random sample the sample average is no longer an unbiased estimator of the population, but if the selection probability for every individual is known, then an inverse probability weighting renders the estimator unbiased. This is the core of most survey weighting approaches as well as the power of propensity score weighting in causal inference (Dehejia and Wahba 1999). The standard correction weight, w_i, then is proportional the inverse of the selection probability π_i:

$$w_i \propto \frac{1}{\Pr(S_i = 1 \mid \mathbf{X}_i, A_i)}. \tag{4.1}$$

In practice, the weight would be normalized by multiplying by a constant so that $\boldsymbol{w}$ is mean 1.

A common question in practice is if a weight for a national survey estimating a national population is valid when applied to a subset of the survey to estimate that subgroup population. In the ideal situation where the propensity score is known, the answer is yes. To see this, we can first see how the weighted proportion of the entire sample is consistent for the population mean. Because Y_i is constant in the finite population setting and π_i is constant if observed,

$$\begin{aligned}
\mathbb{E}\left\{\frac{1}{n}\sum_{i\in\mathcal{S}} w_i Y_i\right\} &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{N} \mathbf{1}\left(S_i = 1\right) w_i Y_i\right\} \\
&= \frac{1}{n}\sum_{i=1}^{N} \mathbb{E}\left\{\mathbb{E}\left\{\mathbf{1}\left(S_i = 1\right) \mid \mathbf{X}_i, A_i\right\}\right\} w_i Y_i \qquad (4.2) \\
&= \frac{1}{n}\sum_{i=1}^{N} \Pr(S_i = 1 \mid \mathbf{X}_i, A_i) w_i Y_i
\end{aligned}$$

after which $\Pr(S_i = 1 \mid \mathbf{X}_i, A_i)w_i$ will cancel and generate μ. The constant rescaling of the weights combine with the denominator $1/n$ is designed to adjust to the correct scaling. While a ratio estimator like this are not unbiased in general, it is asymptotically unbiased. In practice the statistical bias is often small and with more data the estimator converges. Now if we target the subgroup quantity in similar fashion, we simply keep the conditioning of A so that

$$\begin{aligned}
\mathbb{E}\left\{\frac{1}{n_j}\sum_{i\in\mathcal{S}} \mathbf{1}\left(A_i = j\right) w_i Y_i\right\} &= \mathbb{E}\left\{\frac{1}{n_j}\sum_{i=1}^{N} \mathbf{1}\left(S_i = 1\right)\mathbf{1}\left(A_i = j\right) w_i Y_i\right\} \\
&= \frac{1}{n_j}\sum_{i=1}^{N} \Pr(S_i = 1|\mathbf{X}_i, A_i)\mathbf{1}\left(A_i = j\right) w_i Y_i
\end{aligned} \qquad (4.3)$$

will also equal μ_j. Here $n_j = \sum_{i=1}^{N} \mathbf{1}\left(A_i = j, S_i = j\right)$ is the survey sample size for area j.

However, modeling the selection probability is difficult. In the causal inference setting, the propensity scores estimated through, for example, a logit regression do not guarantee balance in the particular sample (Imai and Ratkovic 2014). In the survey setting, the population of $S_i = 0$ is unobserved so running such a regression is impossible.

That is why any survey weights computed after a survey is run are estimated through calibration methods (Zubizarreta 2015; Chattopadhyay, Hase, and Zubizarreta 2020). Calibration methods in one form or another compute a vector of weights that meet a balancing constraint the user defines comparing the target population and sample. Constraints can handle joint distributions through a distinct metric (Hainmueller 2012), but in surveys the weighted only feasible constraint are moment conditions for observable covariates. That is, we find a vector of weights such that $\frac{1}{n}\sum_{i\in\mathcal{S}} w_i X_i = \frac{1}{N}\sum_i X$ where the value of the right hand side is observed in the Census and other larger datasets. Because $\mathbf{X}$ covers multiple categorical covariates such as age group, education, and race, it is convenient to index all the poissle joint combinations of each level of the covariates and denote them as *cells*. Specifically, C_i is a deterministic function of the covariate vector of respondent i and returns a number in $\{1, ..., C\}$.

Post-stratification weighting, rake weighting, iterated proportional fitting weighting falls in this broad umbrella of calibration methods because they follow this pattern as well (Caughey et al. 2020). Because online surveys through river samples generate cannot create design weights, the bulk of weighting that researchers encounter and can model fall under some sort of calibration weight.

Calibration weights implicitly model the propensity score, but inputs to calibration are almost always insufficient in the survey setting. Population moments that serve as the balancing conditions may not be observable for the covariates that are important in the propensity score. Despite the theoretical elegance of calibrating a survey to population constraints, in practice most population constraints are measured with error, measured from several years ago, or measured from yet another survey. Conditions for a given geography may be even harder to obtain. For example, to calibrate the survey to the population distribution of religion, the CCES uses the national breakdown of religion reported from Pew's religion survey. However, Pew does not report breakdowns of religion by state, so the CCES cannot generate calibration weights that apply those constraints. These issues do not even touch on the estimation error due to functional form assumptions.

Once we frame survey weighting as an causal inference problem for observational data (Kuriwaki and Yamauchi 2021), the conditions that a calibration weight must satisfy to make the resulting area estimates unbiased is clear. The same qualifications to the validity of propensity scores apply here. Applied to each area separately:

$$Y_{ij} \perp\!\!\!\perp S_{ij} \mid w_{ij}, \quad \text{for all } j \in \{1, ..., J\} \tag{4.4}$$

$$0 < \Pr(S_{ij} = 1 \mid w_{ij}) < 1, \quad \text{for all values of } \mathbf{X}. \tag{4.5}$$

In other words, selection must be conditionally independent from the outcome after weights are incorporated, and there must be covariate overlap in the sample and the target population. If the weights are calibrated on all the covariates to that are inducing correlation between the outcome and selection, then they will render selection as good as random and return the estimator to a simple random sample.

What are the practical set of issues a researcher faces when weighting a political survey like the CCES to examine representativeness in geographic subgroups? Political surveys for vote choice offer a rare opportunity to assess the representativeness of a survey because elections eventually reveal a population quantity for one of the main outcomes researchers are interested in. How well self-reported vote choice for the office of President can lines up with actual election results at each geographic unit is the central exercise of the subsequent results. A challenge, however, is that the population quantity of interest measures the proportion of outcome among the subset of the population of vote that *turned out to vote*, which is systematically different from the general adult population from the Census. The next section provides an overview of these complex, partially overlapping coverage of key variables in existing data.

4.3 Existing Data

This chapter tests the empirical performance of finer post-stratification using common, open-source data and methods. First, I fix the data to the 2016 Cooperative Congressional Election Study (CCES). The CCES is a good case to examine how far estimates are from ground truth values. I then outline a new approach to creating poststratification with auxiliary data that can be replicated from open-source datasets and APIs. The CCES is the basis of hundreds of articles, and one uniquely positioned for subgroup estimation due to its large size.[1] The CCES is also representative of other modern surveys which are run from online samples with post-stratification, and so the methods here are instructive for other surveys as well.

The existing poststratification weights in the 2016 CCES are constructed by a multistage process (Ansolabehere, Schaffner, and Luks 2017, p.16):

> "The matched cases and the frame were combined and the combined cases were balanced on multiple moment conditions. The moment conditions included age, gender, education, race, voter registration, ideology, baseline party ID, born again status, political interest, plus their interactions. The resultant weights were then post-stratified by age, gender, education, race, and voter registration status, as needed. Additionally, for the common content, the weights were post-stratified across states and statewide political races. Weights larger than 15 in the common content were trimmed and the final weights normalized to equal sample size."

Note that the 2016 CCES does balance on statewide political races in one stage of the process, so can be considered as being calibrated to the state level (as well as the national level) but not at the CD level.

[1] See `https://perma.cc/5P6U-REC9` for a list of publications that use the CCES.

It will be important for later discussion of results that the CCES computes weights to a sample that has already been matched to a target distribution. This sort of pruning of respondents based on the pollster's sampling frame is often a crucial first step in online opt-in panels (Rivers 2007). The 2016 CCES combined data from multiple panel provides (the bulk provided by YouGov itself) based on a `[age x gender x race x education x state]` stratification, and matched to a sampling frame that YouGov modeled from various Census sources with a distance metric incorporating the following variables: `[gender, age, education, employment, ideology, party ID, religion, and voter registration]`. Target values for religion, party, and ideology were taken from a 2007 Pew Survey, and registration was taken from the 2008 CPS. Ultimately, the target values drawn with stratification by `age x race x gender x education x and voter registration` (Ansolabehere, Schaffner, and Luks 2017, p.14–15).

As I have previewed in the previous section, modeling electoral outcomes at small areas therefore requires a complex process of target estimation, i.e., estimating what target distribution the survey should weight to. Figure 4.1 shows the general problem of estimating cell sizes N_c from combining partial margins. Specifically, Table 4.2 summarizes the variable availability across three common datasets used for election modeling in the CCES and other large resource-heavy datasets: The American Community Survey (ACS), Current Population Survey (CPS), and voter files supplied by commercial vendors such as Catalist, TargetSmart, and L2 (Hersh 2015). One takeaway from the table is that no single dataset covers all the standard variables often required for weighting. The ACS provides annual counts of age, sex, education, race, and citizenship. But being a Census dataset, the ACS does not survey party identification, vote choice, or and turnout. Voterfiles are population censuses of at least the set of voters that turned out to vote in a given election, updated constantly. However,

Table 4.2: Imperfect Population Data

	ACS		CPS		Voterfiles	
Population Distributions	State	CD	State	CD	State	CD
Age	✓	✓	✓		✓	✓
Sex	✓	✓	✓		Most states	
Education	✓	✓	✓			
Race	✓	✓	✓		6 Southern States	
Turnout			✓*		✓	✓
Party Registration					31 States	

Note: A ✓indicates the variable is recorded and available at the state level or congressional district level (CD). Data shown are the American Community Survey (ACS), Current Population Survey (CPS), and commercial voter files. * Unlike the voterfile, the CPS records self-reported turnout which is often an overestimate of actual turnout. All information refers to 2016.

The six southern states that record race on the voterfile are North Carolina, South Carolina, Georgia, Florida, Alabama, and Louisiana.

The states that record party on the voter file are Alaska, Arizona, California, Colorado, Connecticut, Delaware, Florida, Iowa, Idaho, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, North Carolina, Nebraska, New Hampshire, New Jersey, New Mexico, Nevada, New York, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Dakota, Utah, West Virginia, and Wyoming.

the Secretaries of State that maintain voter rolls does not collect information on education. Racial identification is collected as part of the voter registration places in only six Southern states, and party registration is required only in about 30. Voterfile vendors therefore use survey and commercial data to merge or impute these variables, which may introduce additional error.

A practical resource for weighting surveys to small areas must take account of these data limitations. In the next section, I outline a workflow that estimates a reasonable poststratification table with existing, publicly available data. Such a poststratification table is equally valuable to be used directly used for weighting, or combined with a partial pooling model that imputes the outcome in each cell as in MRP.

4.4 Synthetic Estimation of Poststratification Targets

I propose the following procedure to construct a poststratification target that approximates the joint distribution of age group, sex, education, race, turnout, and Congressional District (which are nested in states). We start with the following datasets:

- The CCES survey data, which includes the outcome Y, and all covariates $\mathbf{X}$ and A which will be drawn in from other population datasets.
- The ACS estimates of population sizes at the congressional district level. At this level of geography, the ACS does not give a full joint distribution. We must rely on two separate tables. One that records the population counts of `[age x sex x race x CD]` and another that records the population counts of `[age x sex x educ x CD]`).

Therefore, the main challenge is that the ACS only gives a three-way distribution of demographics while poststratification requires a single, fully joint table, and the ACS includes no data on party or turnout.

(1) Fit a multinomial logit bmlogit (Yamauchi and Kuriwaki 2021) predicting four categories of education using race, age, and sex with the CCES data, with the balancing constraint that within each CD, the estimated marginal proportions of education match the education margins reported in the separate ACS table.

(2) Use the predicted values of the model in (1) to predict on the ACS table for `[age x sex x race x CD]`, expanding it into a five-way table of `[age x sex x race x education x CD]`.

(3) Fit a logit model (again with bmlogit) predicting a binary indicator for turnout using the CCES data, where we use the indicator for voterfile match supplied

by the Catalist (included in the public CCES dataset). The population constraint is given by the turnout rate among the voting age population, which can be computed from the ratio of total votes cast to the ACS estimates of the Voting Age Population. The process can falter when at least one set of survey data has at least one cell with zero observations, so here I use a simple specification of: `turnout = race * age + female + educ`

(4) Use the predicted values of the model in (3) to generate a six-way table of `[age x sex x race x education x turnout x CD]`. Subset to the population table cells for which `turnout = 1` to obtain an estimate of the joint distribution of demographics in the turnout electorate at each congressional district

(5) (optional) If party registration data is available and in the states where party registration is available, repeat the same process where the outcome in the multinomial regression is whether the voter in the CCES is a registered as a Democrat, a Republican, or anything else.

(6) Poststratify the survey estimates of the outcome to the resulting synthetic table. If sample sizes for the resulting cells are too small, fit a regression model for the outcome, such as a multilevel model as in MRP.

Here, the balancing multinomial logit is a powerful population constraint. While a regular multinomial logit can fit the same sort of predictive model as in Kastellec et al. (2015), it is likely to simply propagate any bias due to unrepresentativeness into the resulting estimates. Yamauchi and Kuriwaki (2021) implements software to estimate the multinomial regressions as a constrained optimization problem, where an additional constraint that the marginal distribution of the estimated outcome must match a user-supplied population constraint. Users can set a tolerance value to control the degree to which the constraint is enforced relative to the best fitting model in the microdata.

Estimation of population targets is a rich literature of its own, and the approach I propose here is simple relative to other approaches that use proprietary data or software. For example, the CCES itself uses a sampling frame constructed by YouGov that also relies largely on the ACS (Ansolabehere, Schaffner, and Luks 2017). To this table, YouGov adds turnout estimates from the CPS and religion from Pew. However, this sampling frame is proprietary to YouGov and it is only calibrated to the state level. Ghitza and Steitz (2020) use the state-level ACS microdata to estimate onto individual census-tracts, while correcting for representativeness through a type of rake weighting. In contrast, an attractive feature of the proposed model is that it imposes a balancing constraint simultaneously with parameter estimation, and uses publicly available data and summary statistics.

Resulting estimates do not come for free. To overcome the small sample problem, the outcome model partially pools observations from multiple CDs and uses those parameter estimates to predict the outcomes in a single CD. However, this requires the assumption that the demographic predictors and the CD random intercept is sufficient to model the variations in the relationship between the outcome and predictors across the multiple districts (Si 2020). This becomes a classic bias-variance tradeoff, where pooling across districts induces bias but subsetting to specific districts in fitting the model suffers from large variance or even demographic strata with 0 observations.

Another limitation of poststratification, including MRP, is that it cannot balance on important variables if its population distribution is unknown. This is an important omission for political surveys because partisanship is clearly heavily predictive of vote choice but partisan self-identification, the most commonly used measure of partisanship in surveys, is only measured in surveys themselves. Two other related measures of partisanship are vote choice from the past election and party registration where it is available. Incorporating lagged vote would require specific adjustments for voters who did not participate in the previous election. Incorporating party registration is a promising approach, especially for the CCES that includes Catalist's matched voter registration for every respondent. In this chapter, I use summary statistics of party registration in the 2016 election provided by YouGov and show that its incorporation indeed improves estimates. However, it is unclear to extrapolate this calibration to states where party registration is not recorded. The limitation is shared by virtually all methods for poststratification and is a topic of future work.

4.5 Empirical Assessment: Existing Weights

I test these strategies on the problem of measuring Donald Trump's vote share as a proportion of the two-party vote in the 2016 election. Because congressional districts cut across election reporting administrative units in complex ways, some care is needed to compute the ground truth all subsequent estimates will be compared against. I use values computed by Daily Kos (Daily Kos 2021).

The CCES is a survey of voting age adults, while the population of interest is those who voted. When estimating the outcome of interest, therefore, I subset the CCES to respondents who meet all three of the following criteria:

1. Those who responded to the post-election wave (82 percent),
2. Those who self-reported voting for either Donald Trump or Hilary Clinton after the election (76 percent of those who took the post-election), and
3. Those who matched to Catalist's voterfile as having cast a ballot for the 2016 General Election (56 percent)

This leaves a total of $n = 28,462$ respondents, or 44 percent of the available 2016 CCES. When fitting multinomial models to construct the population target, I use all 64,000 respondents to match the coverage of the ACS.

I first assess how standard weighting that CCES includes applies at the subgroup level. Figure 4.2 compares these standard weighted estimates with population variables. I plot the raw proportions and weighted proportions side by side, and compute standard errors by the standard formula

$$\mathrm{SE}_j = \sqrt{\widehat{Y}_j(1-\widehat{Y}_j)/n_j^{\mathrm{eff}}}$$

where $\widehat{Y}_j$ is the proportion estimator (either the simple average or the weighted average) for the area of interest and n_j^{eff} is the effective sample size. For the unweighted case the effective sample size is equal to the sample size ($n_j^{\mathrm{eff}} = n_j$), but for the weighted proportion it is computed with the Kish design effect correction:

$$n_j^{\mathrm{eff}} = \left(\sum_{i \in \mathcal{S}} \mathbf{1}\left(A_i = j\right) w_i\right)^2 \Bigg/ \sum_{i \in \mathcal{S}} \left(\mathbf{1}\left(A_i = j\right) w_i\right)^2. \tag{4.6}$$

This effective sample size can be rewritten as a function of the inverse of the sample variance of weights. It decreases as the weights get more variable.

The state level estimates in Panel (A) show the power of weighting. While 23 states have 95 percent confidence intervals do that include the actual result without weights, all but one (California) of the weighted state estimates include the actual election result. The root mean squared error (RMSE) of the set of estimates improves nearly three-fold. This is of course not surprising given that the weights were calibrated to statewide election returns. As for the national popular vote, the weights give an estimate of 49.5 percent while the raw average gives an estimate of 45.8 percent (Trump's two-party popular vote was 48.9 percent).

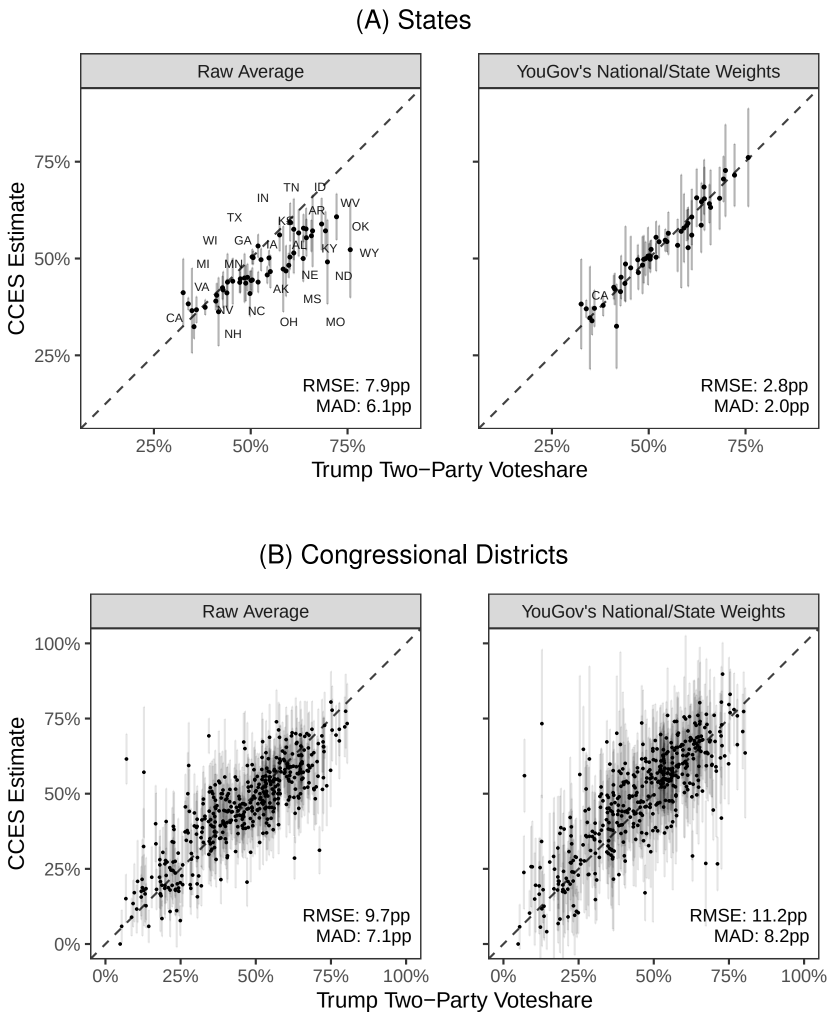


Figure 4.2: Survey Accuracy at State and Congressional Districts

Note: Bars show 90 percent confidence intervals. Labeled states are those where the confidence intervals does *not* include the actual vote share. RMSE indicates Root Mean Squared Error, MAD indicates Mean Absolute Difference.

In the Congressional District level estimates of Figure 4.2 Panel (B), we see the national weights and state weights failing to reduce aggregate error. The average CD has a raw estimate that is off by 7 percentage points compared to the 6 points at the state level. But while weighting dramatically improves the range of estimates at the state level, applying the same weights to the CD level does not improve but instead worsens the average deviation from the district vote share. More CDs have 90 percent confidence intervals using the weighted estimates include the true value than do the confidence intervals using the unweighted estimates (80 percent as opposed to 76 percent). But this is likely because the standard errors of the estimates increased from 6 percentage points to 8 percentage points due to weighting (equation 4.6). The RMSE and average deviation, which does not take into account the standard around each estimate, gets worse after weighting.

The finding in Figure 4.2 is not surprising in the sense that the CCES weights were not designed to match to the Congressional district level, whereas they were calibrated to the state level. It is almost rather impressive that the aggregate error is limited to around 10 percentage points when district has an effective sample size of about $n_j^{\text{eff}} \approx 40$ in the weighted case and no explicit adjustment is made to weight to the turnout electorate. In any case, we see the theoretical results in equation 4.3 not appearing to hold in this data. Understanding calibration methods as an approximation to the propensity score likely explains why. It suggests that the balancing constraints that were used to construct the weights were not sufficient to render the selection ignorable for every district. The next question is whether we can create poststratification tables to design better weights.

4.6 Empirical Application: Proposed Poststratification

There is no obvious way to visualize the result of a six-way cross-tabulation, but Figure 4.3 is one representation of the values resulting from the target estimation procedure. The procedure produces cell counts $\widehat{N}_{cj}$ for area j, where $c \in \{1, ..., C\}$ indexes the multi-way table of categorical demographic variables. In this instant, c indexes the combination of age group (5 levels), sex (2 levels), education (4 levels), race (4 levels), and turnout (2 levels), so $C = 320$. Groupings were determined to match the levels of the ACS variables, and grouped together so that at least every state had one CCES observation of that level. The annotated point on the figure shows, for example, that each poststratification cell is around 0 to 2 percent of the estimated electorate. The estimated size is of course a function of the size of the group in the population. One interesting comparison is the proportions across the turnout and non-voting groups. Some CDs have relatively high levels of Hispanic representation, while in other CDs Hispanics comprise a relatively large group of the non-voting electorate. Non-citizens are included in the non-voting (voting age) electorate, which may explain these high numbers in Texas.

We cannot validate each of these estimates of the population quantities, given that it estimates a joint distribution of variables none of the population datasets can provide (Table 4.2). The method guarantees, however, that these estimates of the joint distribution match all population marginals. And instead of assuming that the distribution of covariates are independent and taking the product of marginals, I use individual survey data to assist in learning the joint distribution.

Weighting the outcome to this target population requires survey sample estimates of the outcome for each of the $C \cdot J$ cells. For each cell cj, denote the average of the outcome in the cell as $\widehat{Y}_{cj}$. When cells are too fine such that $n_{cj} = 0$ for some cells, we

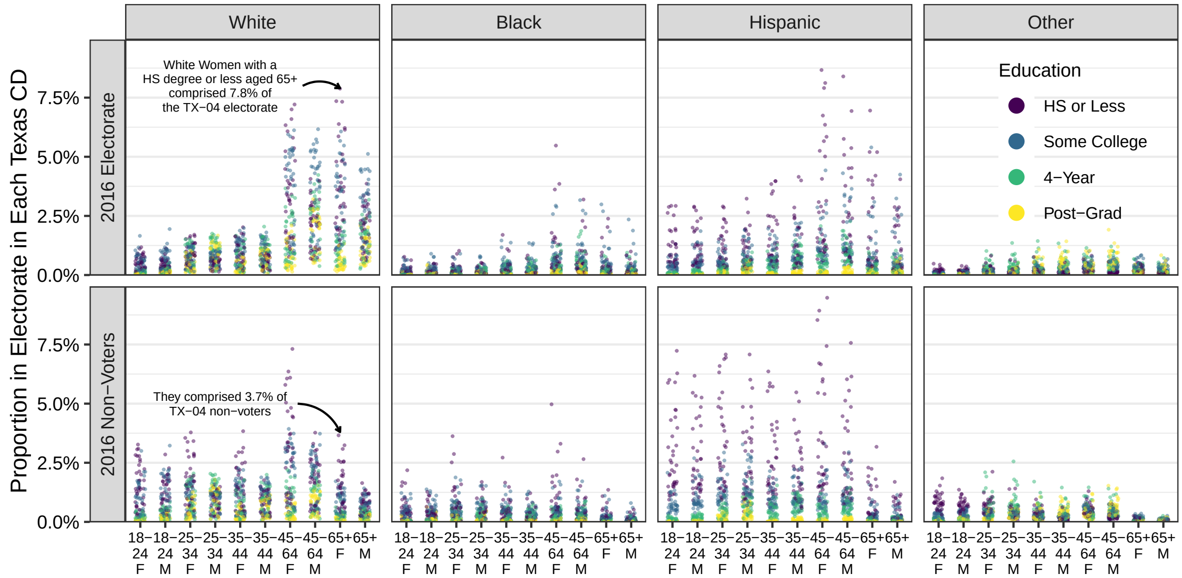


Figure 4.3: Population Cell Size Estimates from Target Estimation

Note: Each point is a CD - demographic - turnout cell. The vertical axis shows the size of that cell in the CD electorate, as estimated from the proposed procedure. Demographic values are aligned by age and sex on the horizontal axis, race and turnout in small multiples, and education in color. Points are jittered horizontally to show variation, but are not jittered vertically.

model a outcome regression with a shrinkage property to estimate these values, such as $\widetilde{Y}_{cj}$. As Figure 4.1 shows, the post-stratification estimator and the MRP estimator is simply the sum of these estimates reweighted to the estimated size of the population:

$$\begin{aligned}\widehat{\mu}_j^{\text{PS}} &= \sum_{c=1}^{C} \widehat{N}_{cj}\widehat{Y}_{cj} \Bigg/ \sum_{c=1}^{C} \widehat{N}_{cj} \\ \widehat{\mu}_j^{\text{MRP}} &= \sum_{c=1}^{C} \widehat{N}_{cj}\widetilde{Y}_{cj} \Bigg/ \sum_{c=1}^{C} \widehat{N}_{cj}\end{aligned} \tag{4.7}$$

where $\widehat{\mu}_j^{\text{PS}}$ denotes the post-stratification estimator for area j and $\widehat{\mu}_j^{\text{MRP}}$ denotes the MRP estimator. As this model shows, the only difference between a MRP estimator and a poststratification weighted estimator is whether a modeled estimate of the outcome in each poststratification cell used instead of the raw average. Since the estimation of the population target is the main focus of this chapter, I use a common implementation of the outcome model and document the details in the Appendix B.

Figure 4.4 show these partially pooled and poststratified estimates. All models are MRP estimates but the key comparison these three specifications allow is that of the post-stratification rather than the multilevel regression.

As a minimal baseline, the first model in Figure 4.4 attempts to isolate the "partial pooling" and outcome modeling aspect of MRP by using a model with no demographic covariates but only random intercepts for state and congressional district. The second model poststratifies on demographics but only those readily available in a standard ACS table. It is labelled Off the Shelf because it requires no extra modeling on the target estimation. Finally, the third model implements the steps (1) through (6) in the proposed workflow, skipping (5) which is left to the next section. If the synthetic target estimation in this workflow was sufficiently accurate in terms of the correlate with the outcome and selection, poststratifying the outcome to this target should improve the accuracy of estimates.

That is, all three estimates in the Figure use an outcome model to generate estimates Y_c that are then poststratified to the respective tables. The predictors in the outcome model correspond 1:1 to the poststratification scheme. Specifically, in R notation, I fit:

1. No Post-stratification: `trump ~ (1|st/cd)`, post-stratified to CD population counts.

2. Off the Shelf: `trump ~ female + age + education + (1|st/cd)`, post-stratified to ACS 3-way table of adult population.

3. Synthetic Population: `trump ~ female + age + education + race + (1|st/cd)`, post-stratified to synthetic 4-way table adjusted for turnout

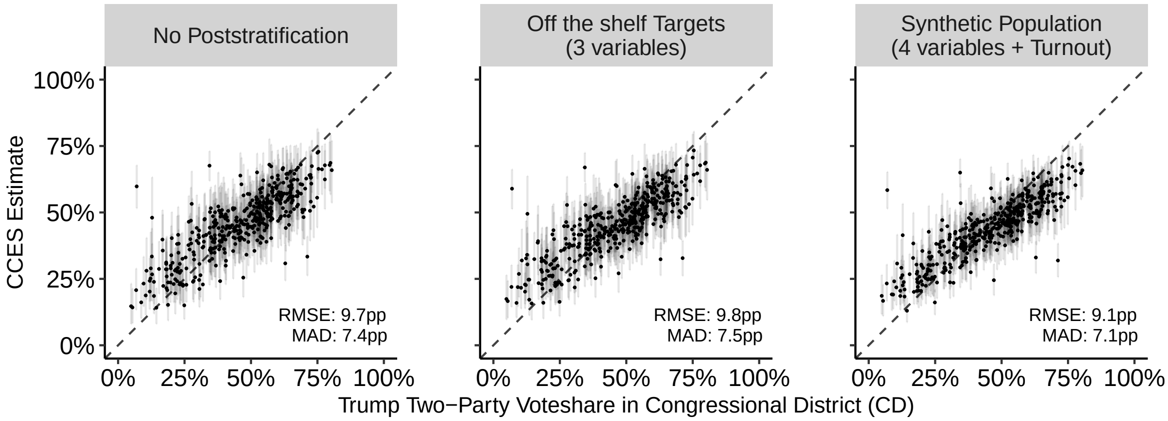


Figure 4.4: Accuracy of Poststratified Estimates

Note: Estimates of Trump vote share using the same survey dataset as Figure 4.2 (B). The off the shelf model uses a standard post-stratification estimate uses a simple poststratification table with three demographic variables directly provided by the ACS. The baseline model uses a synthetic population dataset for the turnout electorate constructed from the following model. Error bars show 90 percent credible intervals from 2,000 MCMC samples.

Compared to the direct, small-sample estimators of Figure 4.2, the smoothed estimators in Figure 4.4 feature tighter credible intervals and modest reductions in the discrepancy between the true vote share. The first estimates show that simply partially pooling the survey data by congressional district does not lead to a improvement in the aggregate error in this case. However, poststratifying on a three-way demographic table appears to have no clear reduction in aggregate error. The estimates improve only in the final panel, when a four-way table is modeled so that surveys can be re-weighted to the joint distribution of race, education, age, and sex by Congressional district, within an estimated electorate instead of the all voting age adults.

A common extension we consider only at the end of this chapter is to add a area-level continuous variable, such as prior vote share in the district, in the outcome model. This covariate has been shown to improve the overall accuracy of predicting electoral outcomes, perhaps more so than individual demographic variables (Hanretty, Lauderdale, and Vivyan 2016). While all our models here likely benefit from this addition, we do not show results for this here because these covariates do not contribute to post-stratification. This can be gleaned for the fact that when district level vote share is simply included in the model, the joint distribution of vote with the other demographic variables is not known. The addition of the aggregate predictor improves the outcome model and partial pooling, that is the estimates of $\widetilde{Y}_{cj}$, but not the post-stratification (Kuriwaki and Yamauchi 2021). As previously noted, growing literature that tests various machine learning models in this aspect of MRP already exist (Bisbee 2019; Goplerud et al. 2018; Ornstein 2020), while the variation in poststratification targets has been relatively unexplored.

4.7 Extensions by Modeling Party Registration

Extending the synthetic table to include party registration is a simple repetition of the modeling procedure, but may require statistics that are not readily available. To further test the idea that modeling relevant covariates in the poststratification can improve the accuracy of estimates, I used currently non-public data to complete optional step (5) to add one more dimension to the table. YouGov maintains a curated database of the voterfile used for their own weighting, and provided a subset of their table that breaks out party registration statistics in each general election electorate by congressional district. I use these aggregate statistics provided by YouGov. That said, some secretaries of states do produce these statistics publicly and some voterfile vendors provide their statistics publicly (TargetSmart 2021) . This allows me to poststratify the CCES data to a table of `[age x sex x race x education x party registration x turnout x CD]`.

Only certain states record party registration on their voterfile (Table 4.2), so for this application I chose the following seven party registration states that cover a variety of regions and population sizes: Arizona, Florida, Iowa, Maine, North Carolina,

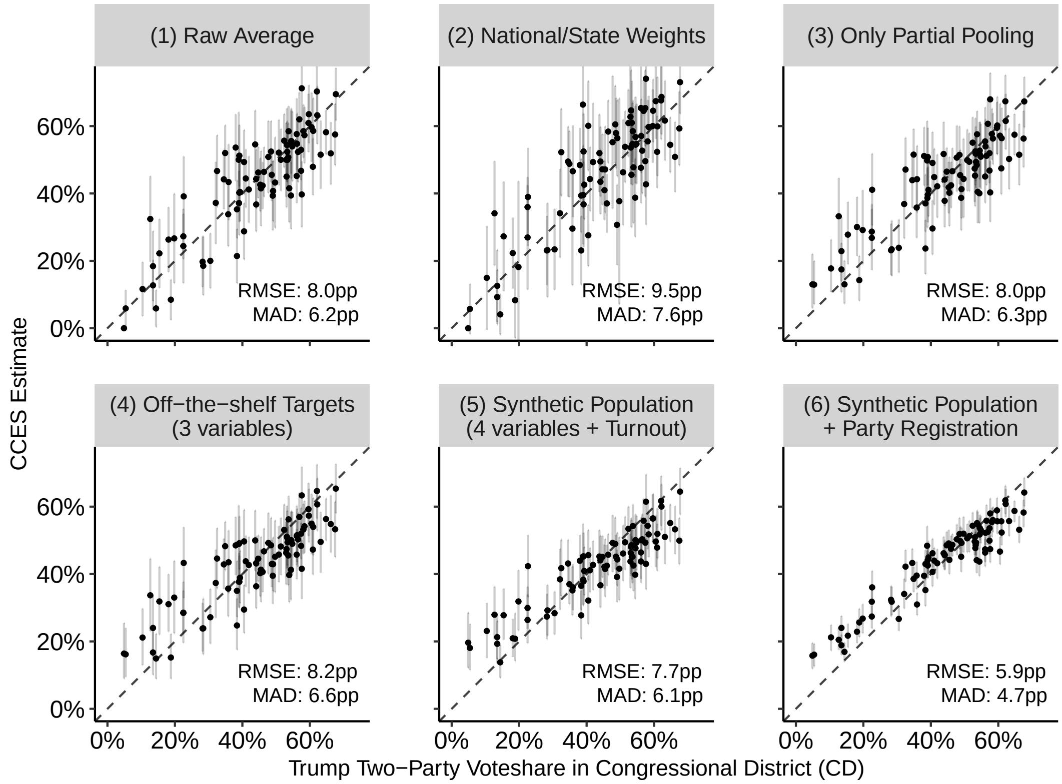


Figure 4.5: Benefits of Additionally Modeling Poststratification Targets

Note: A comparison of modeled estimates using the same survey data from seven party registration states: Arizona, Florida, Iowa, Maine, North Carolina, New York, Oregon. The final model (6) uses a synthetic population target that also includes party registration breakdowns jointly with other demographic variables. Error bar show 90 percent credible intervals from 2,000 MCMC samples.

New York, and Oregon. In order to provide a valid comparison of methods, I recompute the standard MRP and direct estimates in those same states so the underlying data is held constant.

Figure 4.5 compares the full range of relevant models using this data, ranging from simple direct estimators to MRP estimates weighted to a high-dimensional poststratification table. The extensively modeled final model achieves a root mean square error of 6 percentage points, compared to the weighted direct estimates of 9.5 percentage points. The comparisons of the intermediate models are instructive as well.

Models (1) and (2) repeat the finding from Figure 4.2 that using a weights calibrated with coarser constraints may not help and even hurt direct subgroup estimates. Model (3) - (6) are all MRP models following the fnding with all states. Model (3) applies minimal partial pooling without any demographic poststratification, as in the first model of Figure 4.4. We see that the estimates of (3) on do not differ on aggregate from the raw averages. This is one hint that much of the improvement due to MRP is the final poststratification stage rather than the first outcome modeling stage.

Models (4) - (6) vary the underlying target populations in increasing complexity. Model (4) is a simple baseline, which, as in Figure 4.4, only the ACS table measuring `[age x sex x education x turnout x CD]` is used. There is no apparent improvement in the aggregate error with this simple MRP. We only start to see improvements in model (5), which uses the proposed workflow of this chapter and creates a synthetic table of four demographic variables *and* models turnout, both through our balancing multinomial logit model. This improves the root mean square error from the raw average but only by a tenth of a percentage point or so.

The most noticeable improvement comes from model (6) which finally incorporates the party registration breakdown in the electorate. This model, again, ensures that the weighted proportion of registered Democrats and registered Republicans in the survey sample match those reported by the voterfile, for each *congressional district*. The aggregate error decreases by about 2 percentage points compared to the raw average or the partially pooled estimators. It decreases by another percentage and a half, to 4.5 percentage points, after aggregated vote share is included as an aggregate, continuous variable in the outcome model. The strength of the party registration variable is reasonable given that the outcome of interest is voting for a Republican candidate.

4.8 Modeling Aggregate Covariates

A final extension I consider is the inclusion of aggregate predictors in the estimation of the partially pooled estimates $\widetilde{Y}_c$. Although this is not the focus of the methodological innovations in this chapter because it is related neither to post-stratification or partial pooling (in the random effect sense), this sort of predictor has been shown to make a notable improvement in MRP estimates (Hanretty, Lauderdale, and Vivyan 2016) so I consider how the inclusion of these variables on top of existing work change final estimates.

A natural predictor for a district's 2016 voteshare is the district's Republican Presidential voteshare in 2012, which we might denote as μ_j^{2012}. The common setting in election modeling is that we cannot post-stratify on individual prior vote because that distribution joint with other demographics is unknown, but we *can* use the district-level vote share in informing the estimates $\widetilde{Y}_c$. This leads to a somewhat unnatural regression where individual level 2016 vote is regressed on the voteshare of the district where the voter resides, with no clustering of standard errors. Nevertheless, the point estimate on aggregate vote is highly significant and changes the post-stratified estimates.

Figure 4.6 updates models (3) - (6) after adding a spline for the voteshare in the outcome model, and leaves the poststratification table (or lack thereof) the same as in Figure 4.5. Also, to provide some context on how much the 2012 prior vote is predictive of the 2016 outcome, I show a simple comparison of the two variables in the first panel of the figure.

There are improvements across the board, with even simple outcome modeling nearing the accuracy of the most complex model. The marginal benefits of modeling different poststratification tables appear almost to have been wiped out by the large

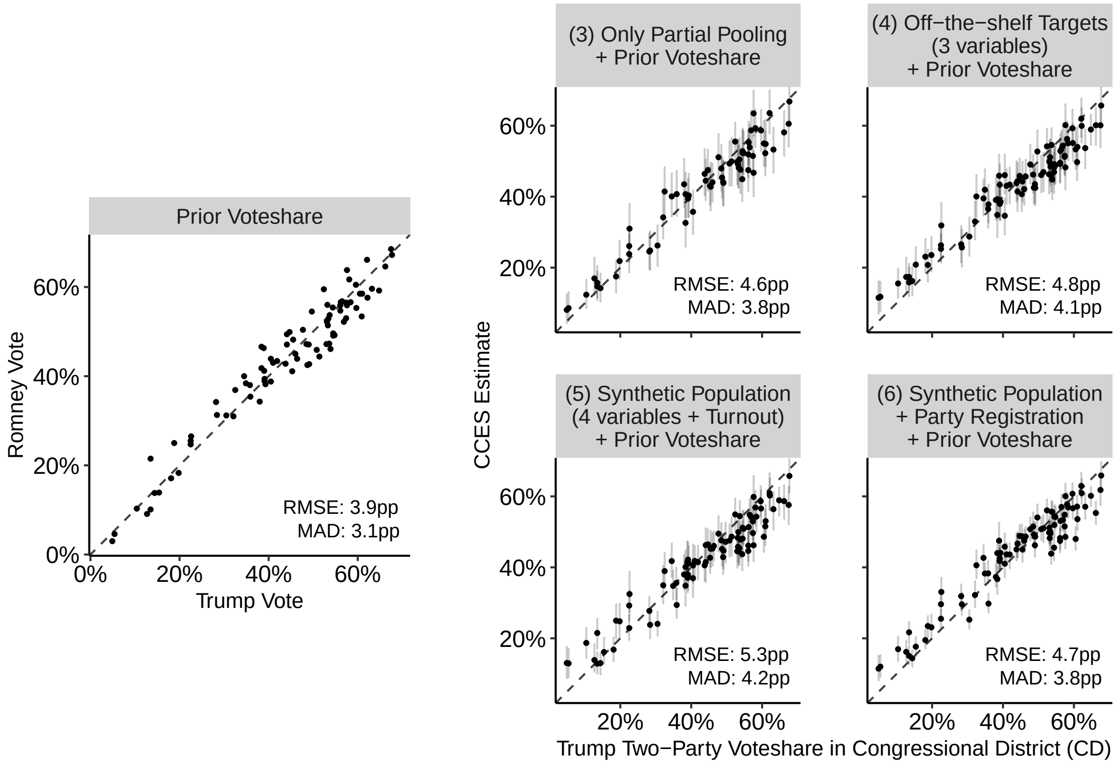


Figure 4.6: Benefits of Additionally Smoothing by District Vote share

Note: Models (3) - (6) are run the same way as in Figure 4.5 but all now additionally include a spline of the district's 2012 Romney vote share in the outcome model. District lines are held fixed, and Presidential voteshare are provided by Daily Kos in all cases.

coefficient estimated on the prior district voteshare. What is also interesting that none of the four MRP estimates do better than simply using the 2012 vote as a predictor (which has a RMSE of 3.9 percentage points, the best yet in this set of states).

This result adds some caveats for how much the extension of the target population is practically useful for survey weighting. Given the relative simplicity of including Romney's voteshare in the outcome regression, the target estimation may simply not be worth it. However, the limitation of such a strategy is that this example may be only a rare case where there is an especially good predictor of the target of interest. When estimating support for other issues, the 2012 Romney vote may not be appropriate. And again, the extent that this predictor dwarfs the gains in target estimation is a broader question for survey modeling. Because the voteshare is not quantitative and therefore treated as a fixed effect in the outcome model, the added value comes neither from post-stratification per se or partial pooling.

4.9 Conclusion

This chapter proposed a framework to improve the target estimation for geographic subgroups, a central component of survey adjustment that has recently been neglected in small area estimation. Using a novel multinomial regression model that simultaneous imposes a calibration constraint (Yamauchi and Kuriwaki 2021), I generated a synthetic population that combines disparate population datasets into a single population distribution and apply a turnout adjustment. Walking through a single, well-known example, I show the limitation of supplied weights or off the shelf MRP models when estimating the electoral outcomes at a subnational level of geography such as the congressional district. We improve estimates when additional variables, such as turnout and party registration are modeled through calibration methods.

A more extensive comparison would have tested more complex outcome models and made use of various other district level variables. Incorporating more precise estimates of the outcome ($\tilde{Y}_c$) or the intermediate steps in the target estimation would probably help and almost certainly not hurt the current estimates (Rentsch, Schaffner, and Gross 2019). The focus of this chapter has been instead on varying the complexity of the target estimation that is used for both MRP and post-stratification weighting. While expanding target populations have been as data intensive as other methods, I presented an open-source workflow with estimation occurring simultaneously with calibration.

Although the workflow does not lead directly to post-stratification weights due to the limited sample in the survey data, the general framework for extending population tables can be used for weighting, for example the proprietary state weights provided by YouGov. Future extensions can also combine such target estimates with complex outcome models.

5 | A Clustering Approach for Characterizing Ticket Splitting in Multiple Offices

Abstract

Large-scale ballot and survey data hold the potential to uncover the prevalence of swing voters and strong partisans in the electorate. However, existing approaches either employ exploratory analyses that fail to fully leverage the information available in high-dimensional data, or impose a one-dimensional spatial voting model. I derive a clustering algorithm which better captures the probabilistic way in which theories of political behavior conceptualize the swing voter. Building from the canonical finite mixture model, I tailor the model to vote data, for example by allowing uncontested races. I apply this algorithm to actual ballots in the Florida 2000 election and a multi-state survey in 2018. In Palm Beach County, I find that up to 60 percent of voters were straight ticket voters; in the 2018 survey, even higher. The remaining groups of the electorate were likely to cross the party line and split their ticket, but not monolithically: swing voters were more likely to swing for state and local candidates and popular incumbents.

* I thank Kosuke Imai and Soichiro Yamauchi for their guidance and help on this chapter. I also thank Marc Meredith for heplful comments.

5.1 Introduction

Finding and labeling voting blocs are ubiquitous in election analysis. Theories of political behavior, especially those explaining electoral change, cluster voters into interpretable prototypes and assign them labels such as core and periphery, standpatters and floating voters (Campbell 1960; Hill and Kriesi 2001; Key 1966; Smidt 2017). Almost instinctively, political consultants, journalists, and election observers latch on to labels such as the "soccer mom" or the "white working class" to construct narratives about voting behavior, even if the label may not have a uniform definition or may not be the best statistical predictor (Carroll 1999; Cohn 2019a; Carnes and Lupu 2020). In particular, a recurring voting bloc in modern accounts of the US electorate is the "swing voter" – a pivotal (and perhaps dwindling) group of voters who are indifferent between either party to a first approximation and are therefore considered persuadable.

But existing approaches to this grouping exercise in political science are either based exclusively on pre-defined groupings, or on a series of comparisons between votes in pairs of offices. The former risks not fully leveraging the information contained in the data, and the latter simply becomes intractable with high-dimensional large-N datasets with an exceeding number of possible voting patterns.

In this chapter I offer an alternative framework: a clustering algorithm that summarizes complex individual-level voting data to interpretable blocs using a probabilistic model. I focus on the specific case of identifying types of voting patterns that include core (party base) and swing, measuring their prevalence, and characterizing the office-specific voting patterns. I derive and implement a fast algorithm tailored to these data structures common in studies of elections in which a single voter votes on multiple offices for federal, state, and local office, each contested by nominees of a major party with varying backgrounds. I then apply this to two datasets: 300,000 actual ballots from Palm Beach County in the 2000 general election, and survey data from ten states from the Cooperative Congressional Election Study in 2018.

This paper enhances our understanding of swing voters – how prevalent they are in the current electorate, how likely they are to cross party lines, and what types of voters swing. It has several specific theoretical, substantive, and methodological contributions. Theoretically, following a long tradition of research (Burden and Kimball 2002; Beck et al. 1992), I infer these latent characteristics from patterns in ticket splitting. But instead of arguing deterministically that only ticket splitters are swing voters, I provide a probabilistic definition of the swing voter bloc. Further, these past studies of ticket splitting analyze one pair of offices at a time. I incorporate information from the joint set of votes on multiple offices in identifying the clusters.

Using voting data from a range of offices and elections, I generally find that a clear majority of the electorate can be classified into the partisan base, but another sizable bloc is what we would reasonably label swing. The size and patterns of this bloc varies systematically. In Palm Beach County, swing voters were more likely to split their ticket in downballot offices rather than high-information Congressional races. In survey data from ten states, I find that up to 8 in 10 of the midterm electorate in 2018 were straight party voters, but popular Governors and US Senators appear to create swing blocs of their own by drawing support from out-partisans.

Methodologically, I show how and when model-based clustering is a powerful tool for empirical discovery and calibration of theories and narratives in studies of political behavior. Despite the natural connections between the parameters in a discrete choice clustering model and the structure of the U.S. ballot, little work has applied this model to voting data,[1] perhaps due to concerns about interpretability and lack of substantive theory. My proposed approach has three methodological features on this point. First, by using a clustering algorithm as opposed to the more standard regression approach, I can properly leverage the information that is contained by the *same* voter making repeated vote choice decisions between Republicans, Democrats, and abstention (for example) in multiple offices. Second, it embraces the principle of unsupervised learning more so than ideal point models. This entails targeting the parameters of a simple model that best fit the data, instead of modeling the behavior of known or presumed voting blocs. Third, my statistical approach is grounded on a probabilistic model of political behavior (Ahlquist and Breunig 2012), instead of simply grouping observations that are close on a particular distance metric (Müllner 2013). Analysts still pick the number of clusters to estimate, and must use substantive prior knowledge to guide the interpretation of each cluster. In summary, the main virtue of the clustering approach is that it is a principled framework to leverage the information in high-dimensional voting data.

In the remainder of this paper, I describe the clustering approach and what it can reveal about swing voters in American Politics. In the models and methods section, I set up the model and show how I derive an EM algorithm to measure the parameters in an open-source program, clusterCVR (Kuriwaki 2021b). In the process, I highlight the assumptions and implications of this statistical approach. I then estimate the same parameters and use a visualization that highlights the estimated parameters in a more interpretable fashion. In particular, I find in my two applications that swing voters form a clear voter type, even though the office-specific patterns of how they swing varies by the type of office and experience of the candidate.

[1] An exception is a working paper by Dubin and Gerber (1992), who analyze ballot propositions, and Hill and Kriesi (2001) who apply the finite mixture model to longitudinal public opinion data to test Converse's black-and-white model.

5.2 Models and Methods

Two quantities of interest are key in analyzing blocs of core vs. swing. First, what is the proportion of each bloc in the electorate? If the swing voter bloc is too small, they may not be pivotal and thus of less interest for campaigns (Grimmer and Marble 2019). Second, what are the latent voting patterns of each cluster? In other words, do most swing voters split their ticket for *all* downballot races, or do their votes change depending on the particular candidate running in each office?

Clustering analysis — and model-based clustering of categorical outcomes in particular — is a well suited to estimate these quantities of interest guided by a simple model of vote choice. I use the term "clustering analysis" interchangeably with labels such as finite mixture models (McLachlan, Lee, and Rathnayake 2019) and latent classification analysis (Linzer and Lewis 2011). The literature also uses variants of the word "categorical" to mean the same thing: multinomial, discrete, qualitative, or polytomous. In the context of this paper, these all refer to the same data structure, i.e. data that is drawn from a fixed number of unordered categories. Political scientists have employed clustering on various types of high-dimensional data such as political institutions, text data, and treatment effects (Estevez-Abe, Iversen, and Soskice 2010; Grimmer and King 2011; Shiraito 2016; Sewell et al. 2016; for a review see Ahlquist and Breunig 2012) to divide them into a few meaningful prototypes, or test different data generation mechanisms (Imai and Tingley 2012). Clustering is even more widely used in fields such as psychology and marketing (Fiske et al. 2002; Wedel and Kamakura 2000). But they are still less common than regression based methods in political science, so I start with the logic of the basic model that I implement in an open-source R package, clusterCVR, and finally respond to commonly recognized limitations of clustering as a method to analyze political behavior.

5.2.1 Main Logic of the Clustering Model

Instead of defining clusters of voters by predictors of vote choice such as race and education, this paper starts with the case in which we only observe the outcome of votes. Let $\mathbf{Y}$ be the $N \times J$ vote matrix of voters $i \in \{1, ..., N\}$ voting in offices $j \in \{1, ..., J\}$. Each vote Y_{ij} takes on a discrete, unordered categorical value $\ell \in \{0, ..., L\}$. In this paper, we focus on vote data where we have coded each vote as

$$\ell \in \{\text{straight ticket, split ticket, third party, undervote}\},$$

by recoding vote choices based on a reference category, such as vote choice at the top of the ticket (as in my ballot application) or partisan identification (as in my survey example). This recoding makes blocs of partisan attachment clearly visible, but for other research questions one could use outcomes {undervote, Republican, Democrat, third party} just as easily. Our data is also high-dimensional, in the sense that for each voter, we observe J choices on different partisan contests. Instead of analyzing each office separately, we wish to leverage information from across offices to infer a voter's latent voting pattern.

Common tools for clustering and dimension reduction such as k-means, PCA, or binary classification cannot be used for such vote choice data because they all require continuous or binary outcomes. Imposing an ordinal scale on vote choice data or discretizing it will likely mask interesting patterns and nuances (Goplerud 2019).

Alternatively, one might analyze vote choice data by showing two-way and even three-way tabulations, computing the proportion of voters that exactly matches a particular voting pattern. However, this quickly becomes intractable because the combinatorics of vote choice on the US long ballot. A typical general election ballot in the US can contain a dozen or so offices, with each voter often making one of at least three choices (Republican, Democrat, and abstain) in each. Third party candidates and the reality that in some offices in some districts are not contested by a major party further increase the number of considerations. Focusing on a pair of offices (e.g. the President and US House, as in Burden and Kimball (2002)) effectively discards valuable information, while enumerating each potential voting pattern (Beck et al. 1992) reduces interpretability.

To address these issues, the clustering approach assumes that each voter i belongs to one of K "clusters", or latent groupings. We denote this membership as a random variable, Z_i, and index clusters by $k \in \{1, ..., K\}$. Importantly, although different individuals may belong to different clusters, there is no differentiation of clusters *within an individual* even across different offices.

It then posits the following model of vote choice that incorporates our two key parameters of interest. The prevalence of cluster k in the population by π_k, where $\boldsymbol{\pi}$ is a K-length proportion that sums to 1 ($\sum_{k=1}^{K} \pi_k = 1$), so that

$$Z_i \sim \text{Categorical}(\boldsymbol{\pi}). \tag{5.1}$$

Next, to characterize each cluster, $\mu_{jk\ell}$ represents the latent propensity for any member of cluster k to vote for a particular option ℓ in office j, so for a given cluster and given office, $\sum_{\ell'=0}^{L} \mu_{jk\ell'} = 1$.

$$\Pr(Y_{ij} = \ell \mid Z_i = k) = \mu_{kj\ell}. \tag{5.2}$$

For example, a political campaign may be interested in the size of the swing voter bloc and how likely that bloc is to split their ticket for a particular candidate in office j. In this case, if we had estimated two clusters, and set aside the first cluster for staunch partisans and the rest for potential swing voters, we would want to know the quantity π_2 (the size of the bloc) and $\mu_{2,j,\text{split}}$.

This modeling choice maps to a theoretical notion that ticket splitting is a probabilistic function of being a swing voter. This stands in contrast to existing approaches, which is more deterministic. Although splitting one's ticket may be a sufficient indicator of being a swing voter, it is not a necessary condition because a swing voter that is indifferent to either party should have roughly *equal* probability of choosing one candidate over the other (Larcinese, Snyder, and Testa 2013).

5.2.2 Estimation Strategy

Our goal is to estimate the unobserved parameters $\boldsymbol{\pi}, \boldsymbol{\mu}$ that is most consistent with the data that we *do* observe, i.e. the vote choice matrix $\mathbf{Y}$. To do so, we must assume the full data generation model as a function of the data and parameters. Once we assume that the probability of a particular vote is independent across offices within the same cluster, we can express the likelihood as a product of J factors:

$$\Pr\left(\mathbf{Y}_i \mid Z_i = k, \boldsymbol{\pi}\right) = \prod_{j=1}^{J} \text{Categorical}(Y_{ij} | \boldsymbol{\mu}_k) = \prod_{j=1}^{J} \prod_{\ell=0}^{L} \mu_{kj\ell}^{\mathbf{1}(Y_{ij}=\ell)} \tag{5.3}$$

which is similar to a standard multinomial regression except that we observe J data points for each voter instead of one, and that we actually do not observe the conditioning variable Z_i. The independence assumption may at first seem unrealistic: a voter's propensity to vote for a Democrat in one office is surely dependent with his propensity to vote for a Democrat in the next. Note that we assume independence only within a cluster. In other words, this model allows for the dependence across offices by averaging over clusters.

I derive an Expectation Maximization (EM) algorithm to quickly estimate the parameters (details left to the Appendix). Because clusters are latent, traditional Maximum Likelihood Estimation (MLE) proves intractable. EM is a well-established iterative procedure that is guaranteed to find a (at least local) MLE. In political science, EM has been successfully employed in ideal point estimation (Imai, Lo, and Olmsted 2016). The statistical insight of EM is that even though we do not know the cluster assignment Z_i, we can temporarily replace it with posterior expectation $\mathbb{E}[Z_i = k \mid \mathbf{Y}_i]$ (the E-step), estimate the MLE of the parameters assuming those are values (the M-step), and cycle through the same steps. The observed likelihood is re-computed at each step as a function of the parameter estimates, and we stop the model once it stops increasing by more than a pre-set threshold.

The EM approach to cluster modeling is also employed in the poLCA (polytomous variable Latent Class Analysis) package by Linzer and Lewis (2011), but the model and package I have derived has several additional features. As I show below, I allow for varying choice sets to model uncontested races instead of listwise deleting voters facing an uncontested race. By incorporating more extensive C++ backends to the computation and fast algorithms for intermediate steps (Yamauchi 2021), my package is also faster by about a factor of ten than existing software in large datasets (See the Appendix). Finally, I visualize the estimated parameters in a more intuitive and interpretable format.

It is well known that the EM algorithm can get "stuck" in local (as opposed to global) maxima depending on the starting values. Following Linzer and Lewis (2011), I initialize parameters with a simple k-means clustering on binarized data, and run 10 versions of the same model but with different random seeds for the initialization. Then, I pick the model that has the highest log likelihood.

How does the clustering approach differ from ideal point estimation, which is another common way political scientists have summarized large voting data? Ideal point estimation (Lewis 2001; Gerber and Lewis 2004) often reduces to solving a very similar likelihood function as clustering, but instead of setting the probability μ as the quantity of interest, they posit that decisions are made according to a one-dimensional spatial voting model and estimates voting preferences on a continuum, rather than as separate blocs. The clustering approach can be thought of as trading away a parsimonious one-dimensional model for a more flexible approach to classify individuals that does not rely on a spatial model of vote choice. Users can choose the numbers of clusters to estimate, and the clusters are not restricted to be placed on a particular coordinate space.

5.2.3 Additional Features of the Clustering Model

So far, the clustering model discussed here follows the canonical model of finite mixtures for categorical outcomes. Several additional features are relevant for analyzing vote data.

Respondent Level Covariates It is natural to posit that certain demographic groups are systematically more likely to be in particular clusters. Clustering models can incorporate such auxiliary data about the respondents in a straightforward manner by modeling the cluster assignment as a function of covariates. Suppose we have an indicator for whether every voter is an ideological moderate. The spatial voting model would predict that this indicator to positively correlate with assignment into a cluster that tends to have high rates of ticket splitting.

Respondent-level covariates like these can be incorporated into the EM algorithm's M-step by regressing the expectation of cluster assignment on covariates in what is essentially a weighted multinomial logit model. Formally, we replace 5.1 with

$$\pi_{ik} = \frac{\exp\left(\boldsymbol{X}_i^\top \boldsymbol{\gamma}_k\right)}{\sum_{k'=1}^{K} \exp\left(\boldsymbol{X}_i^\top \boldsymbol{\gamma}_{k'}\right)}, \tag{5.4}$$

where $\boldsymbol{X}$ is a $N \times P$ matrix and $\boldsymbol{\gamma}_k$ are $P + 1$ coefficients and an intercept. As this shows, this requires we let $\boldsymbol{\pi}$ be a matrix with N rows. When summarizing the data in subsequent analyses, I take the average mixing proportion for each cluster as an aggregate measure. A model with and without such covariates should produce roughly similar cluster assignments because we still infer these from the votes. But incorporating covariates can help stabilize the algorithm, and the values of the coefficients $\boldsymbol{\gamma}$ provide useful substantive information for interpreting cluster membership. The ECM algorithm implemented by Yamauchi (2021) makes this step fast enough to be repeated at each step in the main EM loop.

Varying Choice Sets Many elections for state and local offices are uncontested, which means that a voter still makes a choice, but from a limited menu of options. These different settings require modelling varying choice sets (Yamamoto 2014). While existing discrete clustering models rule out this possibility and therefore require analysts to drop data that includes varying choice sets, I model these separately, with an independence of irrelevant alternatives (IIA) assumption to share information and parameter values across observations.

The added complication is that the vote choice probability must now be modeled as a function of data that varies by the choice set. Let $\mathcal{Y}_{ij}$ denote the set of values that are available to voter i in option j. Such information would be clear from the candidate filings in that district, and so are directly observed. We then posit that the choice probability is generated from a ratio that is relative to the available choices for a given voter, as in a standard multinomial logit. Formally, we parameterize equation 5.2 as

$$\mu_{kj\ell} = \Pr\left(Y_{ij} = \ell \mid Z_i = k\right) = \frac{\exp\left(\psi_{kj\ell}\right)}{\sum_{\ell' \in \mathcal{Y}_{ij}} \exp\left(\psi_{kj\ell'}\right)} \tag{5.5}$$

where ψ_ℓ is a scalar that represents the intensity of preference for option $\ell \in \{1, 2\}$ relative to $\ell = 0$ (abstention). To identify the MLE, for which no closed-form equation exists, I use an optimization of the likelihood with varying choice sets. The functions and derivations are provided in Appendix C.

The IIA allows us to estimate the same parameter across voters in slightly different electoral contexts, but it can be a substantial assumption to add. The multinomial probit does not explicitly make this assumption, but replaces with a distributional assumption on the errors. In general, the validity of IIA is difficult to test because each respondent's rank ordered preferences are not observed in the cases relevant here. Yamamoto (2014) shows how one can model the varying choice sets through a choice-set specific intercept and relax this assumption. Although the method in this chapter does not conduct this modeling, there is a parallel in the approach. Yamamoto's method estimates separate effects for each choice set, whereas this method partitions voters into clusters and estimates separate choice probabilities within each cluster.

5.2.4 Limitations and Issues of Interpretation

Before moving to empirical analyses, the implications and limitations of this methodological approach are worth some comment. While clustering algorithms are widely used in fields including computer science, marketing, and psychology, this is not the case in political science or economics. It is important to consider why.

First, the type of data this method can handle are restricted to datasets where (i) the outcome measurement is categorical and (ii) come from the roughly same choice set across offices.[2] For example, it would not be possible to analyze a set of variables that includes vote choice and numerical responses. Similarly, the model also cannot handle ballot data where some variables are partisan offices (Republican, Democrat)

[2] While I allow for varying choice sets, a better name for this commonly used term is a "limited choice set," i.e. one of the options being missing.

while others are nonpartisan offices or referendums (Yes, No), unless the analyst is willing to assume that voting "Yes" on a particular referendum represents the same underlying event as voting for a Republican or Democrat. For such cases of mixed ballots that violate (ii), an ideal point model will be more appropriate because it targets a single dimensional preference estimate and maps different votes to a single space. For cases that violate (i), one must turn to other clustering methods like k-means for datasets with only continuous outcomes, and more involved models for a mix of continuous and categorical outcomes.

On a more important theoretical point, clustering algorithms almost always require the user to pre-determine the *number* of clusters K to model from the data. Therefore, one might worry that substantive findings from data may change wildly by the number of clusters. There do exist methods to pick the optimal number of clusters based on measures of model fit (Fraley and Raftery 1998) — essentially functions of the observed likelihood attempting to account for overfitting.

But we do not need to believe that there is one "correct" number of clusters the analyst has to identify in order for clustering analyses to be useful. As Broockman (2016, p. 207) argues, voter's preferences are likely formed by hundreds of small issue "dimensions", even though each one may not incrementally improve model fit. Whether one models 2 clusters or 3 clusters from the data is not a claim about the analyst believing that 2 or 3 dimensions are enough to explain voting behavior. Instead, this method can be thought of as a principled way to summarize information and characterize prototypical voting patterns *given* the user's chosen level of granularity. Substantive theory, rather than only a statistical information criterion, should guide the choice of the number of clusters.

A related concern about unsupervised learning methods is that interpretation of each cluster is arbitrary. Examining the correlation of covariates with estimated cluster assignment is a useful way to uncover some interpretation. But the analyst must also bring some of their own substantive knowledge for this clustering algorithm to be useful. Indeed, model output should not be interpreted as anything more than as a summary of the data based on a simple probabilistic model of vote choice. In the same way that there is rarely a single "correct" number of clusters, it is actually reasonable to pick the number of clusters so that it reveals clusters whose estimated parameters $\boldsymbol{\mu}$ match the theoretical quantity of interest. In my applications, the main quantities of interest are the size and voting patterns of swing voters, which the model parameters directly target.

Of course, one must start somewhere. One reasonable initial choice is $K = 2$, which is the simplest case and also has parallels to many theories like the black-white model or core-and-periphery. Or, one can start by setting the number of clusters to the number of response options there are. This allows the data to cluster into homogeneous response-specific clusters, if that is the underlying pattern. In the context of the core vs. swing model, one might posit that voters can be partitioned into swing voters who split their ticket *regardless of the office*, abstention voters who undervote regardless of the office, and so on. This is a useful null hypothesis that I test on the data, and ultimately reject.

5.3 Application to Cast Vote Records

The clustering algorithm is well suited to glean patterns from large datasets of anonymous ballots. I first illustrate the insights from the clustering approach by analyzing ballots from the Florida 2000 election, which were originally analyzed in an ideal point framework.

Both political scientists and election administrators use ballot data to understand voter behavior and ballot design, but the high-dimensional and large-N nature of these datasets makes analysis challenging. For example, early work by Jeff Lewis and colleagues have analyzed ballots from Florida, South Carolina and Los Angeles country to study the political preference of Ralph Nader voters, election anomalies, and the multidimensional structure of voter's preferences over national and local issues (Herron and Lewis 2007; Bafumi et al. 2012; Gerber and Lewis 2004). In Chapter 3 I used this data to ticket splitting behavior across the long ballot, including in offices where surveys cannot poll, and Morse (2021) collects ballots from the Florida 2018 election to study voter behavior for criminal justice and voting rights policies. Finally, cast vote records have become integral for transparent election administration, in particular for the implementation of risk limiting audits which require that state secretaries of state can readily sample from the population of ballots (Stark 2008; McCarthy et al. 2018; for a review see Kuriwaki 2020).

Herron and Lewis (2007) coded the ballot punch cards from over 3 million ballots from ten large counties in Florida, guided by the question: Did Ralph Nader spoil Gore's victory in Florida — and thus the presidency — by drawing leftist voters who would have supported voted for Gore in the absence of a third party candidate? Surveys data are too sparse and unreliable to estimate the preferences of a subpopulation like Nader voters, so the authors turn to cast vote records to analyze the down-ballot voting patterns and estimate ideal points for each voting pattern. By imposing a spatial model with candidate valence on the data, the authors are able to predict whether Nader voters were more proximate to Gore than for Bush.

In this application, I use the author's replication data but apply the clustering algorithm instead of an ideal point model to answer the main substantive question of this paper: what proportion of voters could be considered swing voters, and how likely were they to split their ticket? I leverage the joint distribution of votes available in the long ballot and analyze one large county — Palm Beach County — that had many partisan offices on the 2000 ballot. This subset still contains 300,000 votes, more than any survey of the entire state. I use almost all the partisan races on the ballot: both chambers of Congress, two statewide cabinet positions (Education Commissioner and Treasurer), the public defender for the 15th circuit court of Florida, and three county-wide executive offices (Clerk of the Court, Sheriff, and Tax collector).

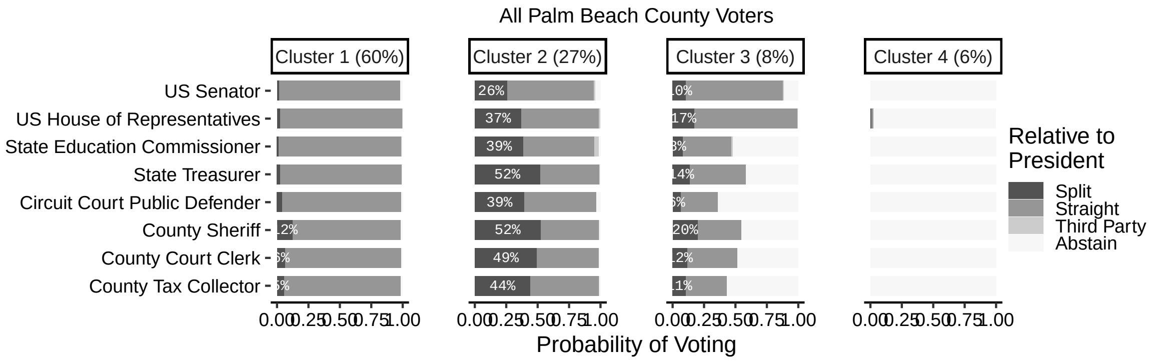
I use voter's choice in the Presidential race as the reference category for coding whether a voter votes for a straight or split ticket. Because this outcome is not well defined for voters who chose a third party candidate in the Presidential race, I drop these from consideration so that the choice set to is {straight ticket vote, split ticket vote, third party vote, or undervote (abstention)}.[3] I cannot determine the congressional district of those with an undervote in that office, so I further subset to the vast majority of voters who cast a vote in one of the four congressional districts in the county.

Figure 5.1 panel (A) shows the estimated parameters for the size of each cluster ($\boldsymbol{\pi}$) and the latent voting probabilities $\boldsymbol{\mu}$, faceted by the 4 clusters that partition the electorate. Each of the three panels represents one run of the clustering algorithm. The length of each segment represents the vote choice quantities $\boldsymbol{\mu}$ for a particular office and cluster of voters. I initially chose four clusters because there are four possible response categories, as reasoned in the previous section. Because clusters have no inherent ordering in the statistical model, throughout this chapter I number the clusters by size the estimates of $\boldsymbol{\pi}$: the cluster estimated to capture the most respondents is called cluster 1, the second largest as cluster 2, and so on. I do not use covariate data such as congressional district in this current application, but I conduct separate analysis by the Presidential vote choice covariate in panel (B) and (C).

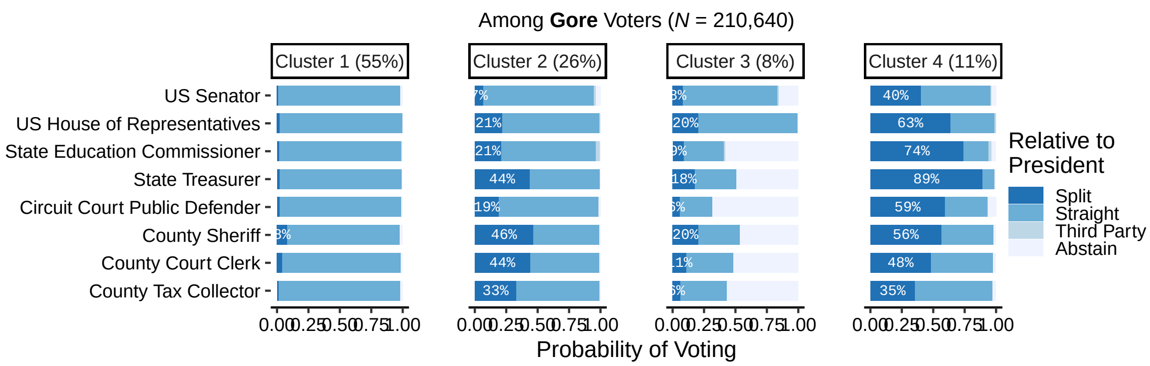
The estimated parameters offer a straightforward summary of a vast amount of

[3] In this county, none of the offices I mention were uncontested. I analyze the case of uncontested House races in the next application.

(A) All Voters for Gore or Bush



(B) Among Gore Voters (N = 210,640)



(C) Among Bush Voters (N = 124,500)

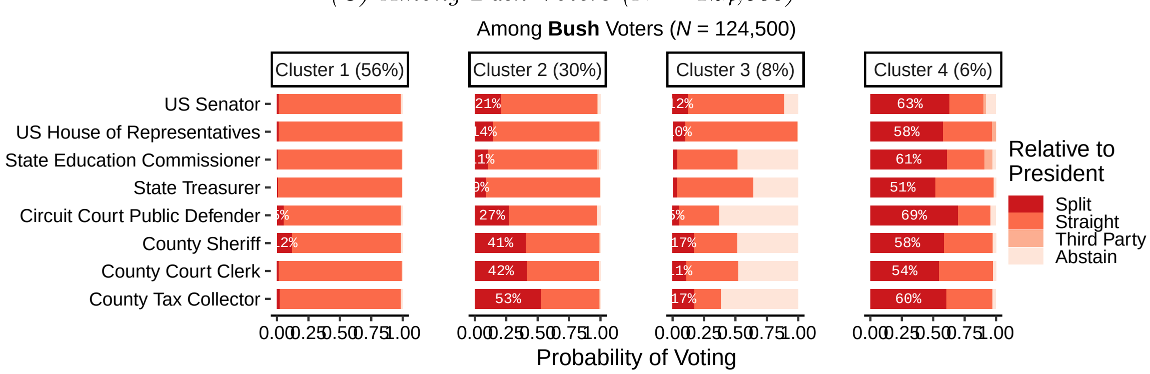


Figure 5.1: Clusters of Voting Profiles in Palm Beach County Florida, 2000

Note: Plots Estimates from the Clustering Algorithm, where one panel is from one model. The length of each bar indicates the estimated value of $\mu_{kj\ell}$, or the probability voters in cluster k votes for option $\ell \in \{\text{abstain, straight, split, other}\}$ in office j.

high dimensional data. First we see that about 55 to 60 percent of Bush and Gore voters are straight ticket voters (cluster 1), who vote for the same party's candidates (Democrats in Panel (B) and Republicans in Panel (C)). Next, about 26 to 30 percent of voters have a decent probability of crossing the party line. Interestingly, this group in cluster 2 is more likely to split further down the ballot, i.e., in state and local offices. This is consistent with the argument in Chapter 3 that although partisan cues may be relatively stronger in low information environments (Peterson 2017), there is also a strong incumbency advantage and valence differential between candidates in state and local elections.

Next, cluster 3 represents about 8 percent of the population and are mostly characterized by undervotes.[4] There is a notable office-specific pattern here, too. These voters appear to abstain in state and local offices, but still vote straight ticket for Congress. Cluster 4 also has high levels of ticket splitting but is the smallest group among Bush voters. In fact, when both of the subgroups are stacked together in Panel (A), an all-undervote cluster arises as the fourth cluster.

The original question in Herron and Lewis (2007) was to study how similar preferences of Ralph Nader voters were to Gore voters, which we can also re-assess with the current clustering approach. In Figure 5.1, I had dropped Nader voters entirely from the analysis because our outcome was coded with reference to the two-party vote in the Presidential race, but in Figure 5.2 I recode the outcomes with reference to the *US Senate* race between Nelson (D) and McColumn (R), and then run separate clustering algorithms by Presidential vote choice among Bush, Gore, and Nader voters.[5] Nader

---

[4] Among Gore voters, I flip the numbering between cluster 3 and 4 so that each clustering's voting patterns are similar to their counterparts in the other panels. This violates the rule for numbering clusters by estimated size, but here the sizes are similar enough that I opt for the gains in comparability.

[5] Therefore, one note of implication here is that now the voters in each panel are a mix of Nelson and McColumn voters.

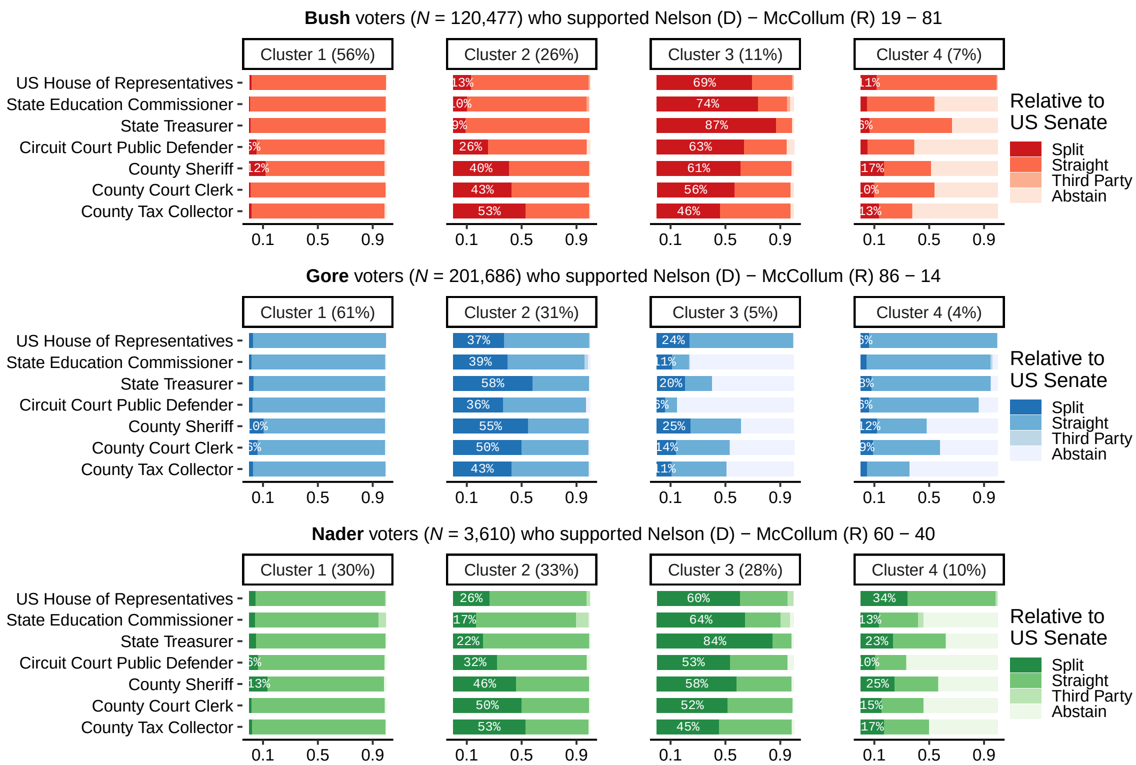


Figure 5.2: Swing Voters by Presidential Support

Note: Figure format follows Figure 5.1 in showing Palm Beach county voters but recodes the reference category to the US Senate race, and includes Nader voters who voted for a Republican or Democratic Senate candidate.

voters exhibit different patterns of ticket splitting. Cluster 3, which is more likely to split their ticket than stick to their Senate choice, comprises 28 percent of Nader voters but only 11 percent of Bush voters and 8 percent of Gore voters. These findings are roughly consistent with the original paper's findings based on a IRT model, namely that Nader voters were not in fact predominantly "left" of Gore, and so it is not clear if Nader handed the Presidency to Bush by running. For example, Nader voters only supported the Democratic Senate candidate 60 - 40.

The picture of the electorate that emerges from these analyses is one in which 60 percent were party loyalists and about 5 percent roll off, but where a quarter of the vote can be considered as a reasonable swing bloc. Further, the clustering model's parameter estimates in both cluster 2 and cluster 3 help refute the null hypothesis that voter types vote straight regardless of the office. While a majority of voters were consistently straight partisans, the rest vote differently, with a particular difference between Congressional, state, and local offices.

5.4 Application to Survey Data

In my second application, I turn to the 2018 midterm election using survey data from the Cooperative Congressional Election Study (CCES). Survey data is limited by its small sample size and potential risk of measurement error, but it provides comparable data across states and a rich set of individual-level covariates that anonymous ballot images lack. Moreover, surveys have more wide-ranging uses in social science research and polling. Like the ballot data, the clustering algorithm can be applied to respondents who answer a series of questions that all roughly draw from the same set of categorical outcomes.

As illustration, I use the midterm ballot in 2018. Midterm elections are an interesting case for the study of the swing voter because many state-level offices are elected in midterm years. I first gather the post-election wave's self-report responses for Congress, Governor, and State Attorney General in ten of the largest states where those offices were on the ballot, shown in Table 5.1. Later, I also compare the results from the 2014 CCES, a year in which many Republicans flipped seats, in contrast to 2018 which saw a Democratic resurgence.

I code the respondent's vote choice as a straight / split vote with the voter's partisan self identification as the reference category, instead of presidential vote choice as in the Palm Beach example. I drop pure independents because the notion of ticket splitting is not well defined for this group, and make inferences about the voters who identify with or lean towards one party or the other. Therefore, the sample definition

Table 5.1: States Analyzed in the 2018 CCES

	Incumbent (running for re-election)				
State	US Senator	Governor	Attorney Gen.	CDs	CCES n
Massachusetts	Warren (D)	Baker (R)	Healey (D)	7 (5)	688
Maryland	Cardin (D)	Hogan (R)	Fosh (D)	8 (8)	663
Texas	Cruz (R)	Abbott (R)	Paxton (R)	36 (32)	1,128
New York	Gillibrand (D)	Cuomo (D)		27 (21)	1,990
Florida	Nelson (D)			27 (22)	2,817
Michigan	Stabenow (D)			14 (13)	1,156
Wisconsin	Baldwin (D)	Walker (R)	Schimel (R)	8 (7)	757
Ohio	Brown (D)			16 (16)	1,555
Arizona		Ducey (R)	Brnovich (R)	9 (8)	1,051
Minnesota	Klobuchar (D)			8 (8)	652

Note: The ten largest states (in terms of CCES sample size) that held contested elections for US Senate, Governor, and Attorney General in 2018. Incumbent columns show the name and party of the incumbent if running for re-election; open (but contested) seats are left blank. CDs indicate the number of US House districts in the state, with the number of contested districts in parentheses. States are ordered as in Figure 5.3.

is slightly different from the previous example, but still captures more than 80 percent of the electorate sampled. I model the uncontested races in the US House using the IIA assumption previously discussed. I also model cluster assignment with the respondent covariates for Democrat, White, high news interest, identification as an ideological moderate. I fit clusterCVR using three clusters,[6] generating 10 replicates and picking the model with the highest log likelihood.

Figure 5.3 shows the final estimated cluster sizes and cluster characteristics, by state and cluster, in the same fashion as before. To understand which types of partisans and demographic groups were more likely to be in particular clusters, we plot the estimated coefficients (γ in equation 5.4). States in the figure are ordered by the estimated size of the first cluster, which is evidently the cluster for straight ticket voters as it was in the Palm Beach county example.

[6] Estimating four clusters recovered similar findings for the swing voter bloc, but was sensitive due to much smaller samples in certain states compared to the ballot data example.

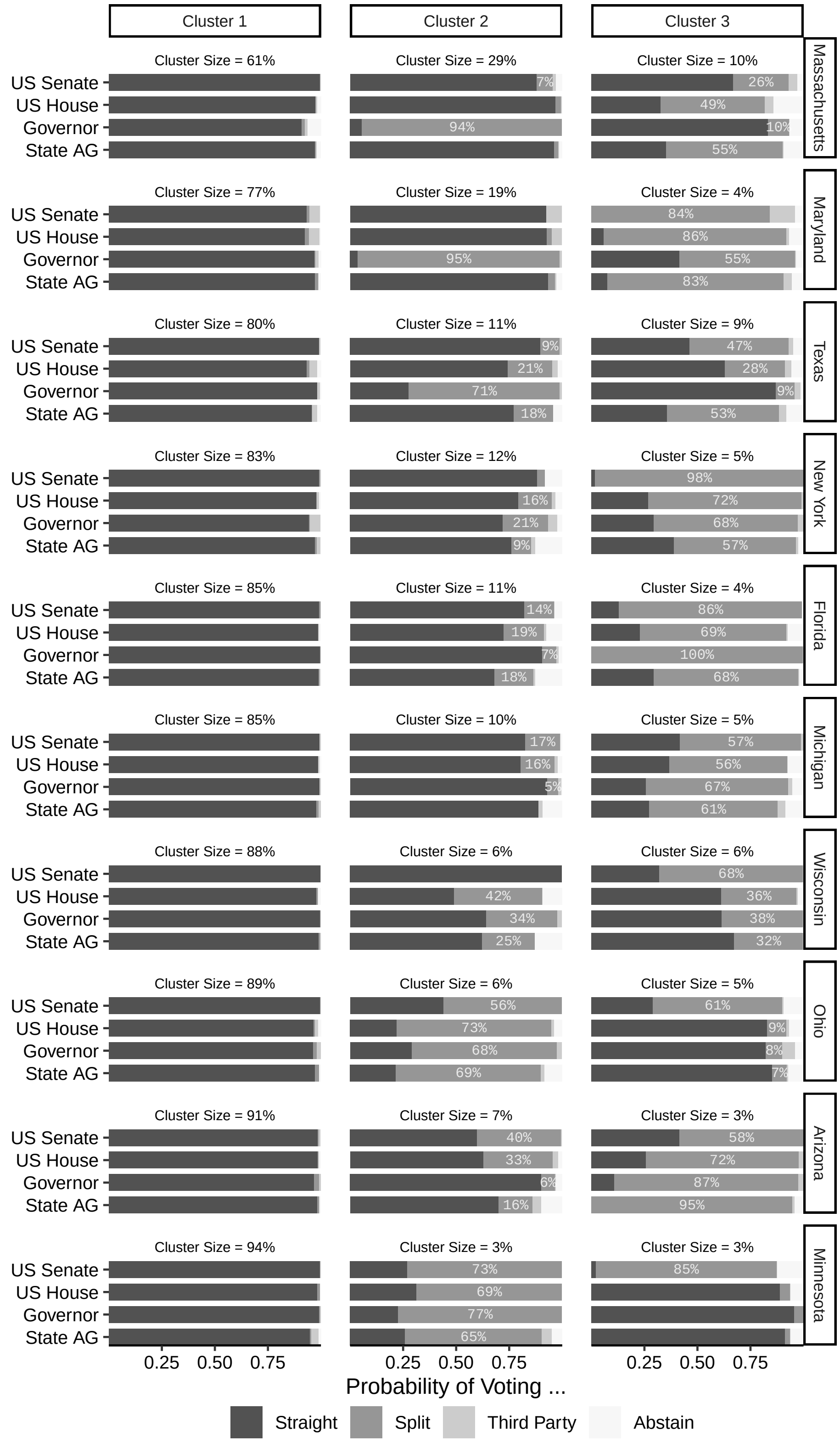


Figure 5.3: Clusters of Voters in 2018 Surveys

Note: Each facet shows the size and estimated latent voting patterns of an estimated cluster, for a given state and given cluster. Data: CCES 2018.

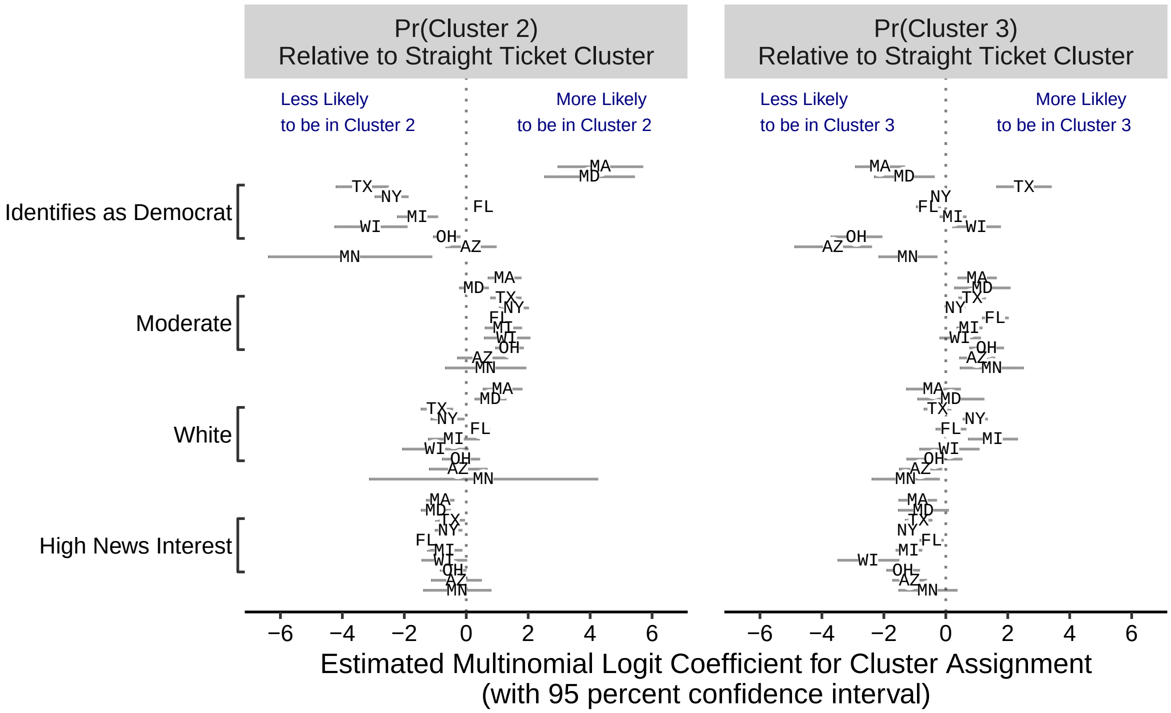


Figure 5.4: Demographic Predictors of Cluster Membership

Note: Each point is a coefficient estimate for predicting cluster membership estimated within the clustering process. In all states, moderates and low news interest voters are more likely to be in the two smaller clusters, but the correlation by party differs starkly by state. All four predictors are coded to be binary indicators. Letters correspond to the state being estimated and one clustering algorithm is estimated per state. Cluster numbers correspond to those in Figure 5.3. Data: CCES 2018.

Across all states, we find that the majority of the electorate are straight ticket voters, who are all but certain to vote for the same candidates as their party. But the proportion of these voters ranges from close to 95 percent in Minnesota to nearly 60 percent in Massachusetts. The remaining two clusters appear to contain various types of ticket splitters.

Interestingly, the pattern of ticket-splitting is state- and office-specific. Figure 5.3 shows that states where the core partisan bloc is smallest are Massachusetts, Maryland, and Texas. In all three, the Governor vote stands out as in the second cluster, approaching 30 percent of the voters in Massachusetts. As the table of candidates in

Table 5.1 shows, in all three states a Republican incumbent ran for re-election. The coefficients on the voter's Democratic partisan identification predicting cluster membership (Figure 5.4) indicate that in Massachusetts and Maryland, it was Democrats who split their ticket (for Baker and Hogan, respectively), whereas in Texas, it was Republicans who split their ticket (*against* Abbott). In Wisconsin, the only other state in Figure 5.3 in which a Republican Governor ran for re-election, there is no clear Governor-specific bloc, but both Republicans (cluster 2) and Democrats (cluster 3) were most equally likely to cross party lines in their vote for Governor. In New York state, where Cuomo ran for re-election there is no clear Governor vote. That Massachusetts and Maryland to the top in the Governor vote is not necessarily surprising, given that Baker and Hogan were the nation's two most popular Governors in the summer leading up to the election, and won with large margins.[7] The clustering estimates, however, add more insight than these standard statistics because they summarize the survey data into voter prototypes: it shows that Democrats supporting Baker and Hogan still voted for Democratic candidates in other offices.

We also see clear, if smaller, blocs of ticket splitting for incumbent Senators in cluster 3 of Ohio (Brown) and Minnesota (Klobuchar). In both of these states, Republicans were more likely to be in the cluster that were more likely to ticket-split, indicating that they voted for the Democratic incumbents. We do not see similar blocs of pro-incumbent Senator voting blocs in Massachusetts (Warren), Maryland (Cardin), Texas (Cruz), and Michigan (Stabenow). These patterns across different candidates of the same offices are consistent with the notion that popular incumbents tend to attract ticket splitting votes.

It may be the case that simple cross-state comparisons are confounded by fixed characteristics of the state's electorate. Therefore as a final step, I repeat this analysis

[7] "America's Most and Least Popular Governors". Morning Consult Poll, July 25, 2018. https://perma.cc/2XYN-NJZ7

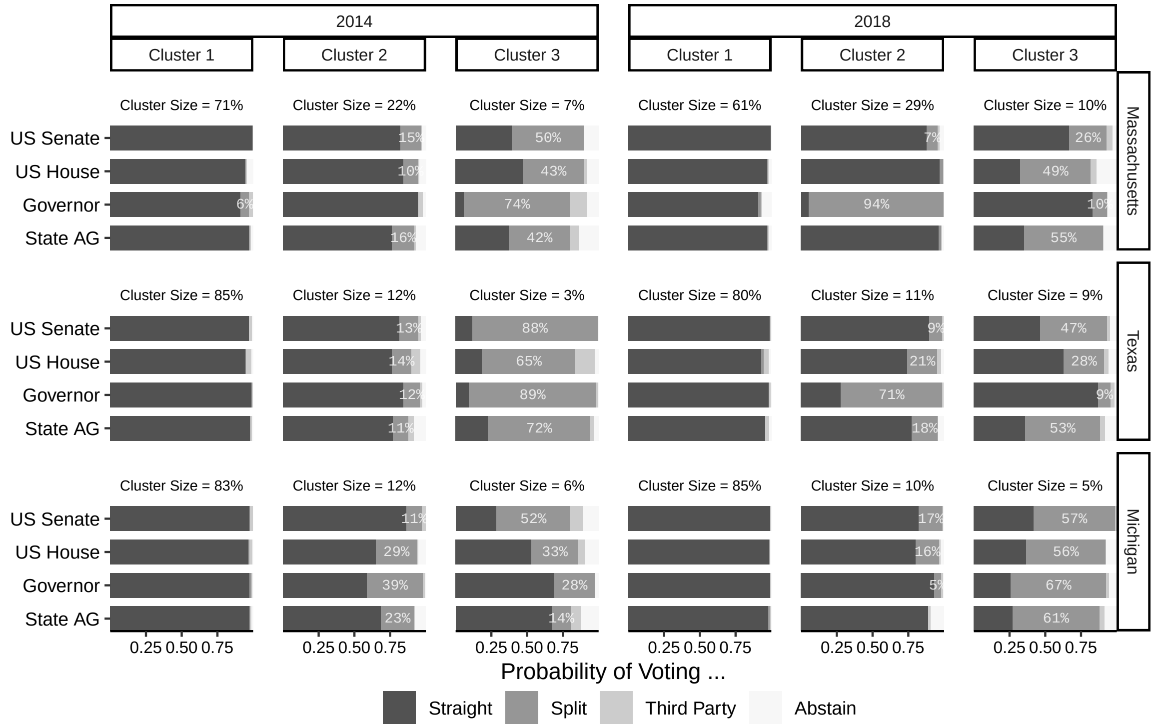


Figure 5.5: Clusters in Surveys between 2014 and 2018

Note: A separate analysis is run for all states which held a contested US Senate, Governor, and State Attorney General race in both 2014 and 2018.

4 years prior, in states where the same offices (but usually not the same candidates) were on the ballot. Three states out of the ten shown in Figure 5.3 meet that criteria: Massachusetts, Texas, and Michigan. I ran the clustering model on similar 2014 CCES data, and show the estimated model parameters side by side in Figure 5.5.

The Governor-specific ticket splitting found in Massachusetts and Texas in 2018 is not found in 2014, even though the same Republican candidate was on the ballot. In 2014, few voters were classifiable as swing, but in 2018 Baker netted more votes from Democrats and won 67 - 33 while Abbott suffered from Republican defections (but pulled off re-election despite it, winning 56 - 43). In Michigan, the composition of straight and split ticket voting blocs did not change significantly.

5.5 Conclusion

In this chapter, I introduced a clustering model for election observers to draw insights from large-N, high-dimensional data. I outlined a simple model of vote choice for partisan races on the long ballot, and offered some guidance on how data analysts should interpret the model estimates of latent quantities.

In an application to the crucial vote in Palm Beach County in Florida, the clustering method found that a majority of strong partisans (although not a super-majority), about 20-30 percent of potential swing voters who split disproportionately in state and local offices, and about 8 percent of rolloff voters who still voted for their members of Congress. In the second application to survey data in 2018, the method reveals that about 80-90 percent of voters who identify with one party are straight ticket voters. But popular Governors and some popular Senators drawing voters across the party line and effectively forming blocs that deliver their re-election.

The statistical model used here can benefit from several more additions in the future. First, one can include *choice*-specific covariates such as candidate ideology, candidate incumbency, and candidate gender, directly into the estimation. These coefficients are widely analyzed in multinomial logit models of consumer choice but a faster algorithm to solve such models must be derived to incorporate them into a EM algorithm for large datasets. Modeling district-specific characteristics as random effects by positing that they are drawn from a common distribution is another possible feature to add to the modeling process, although estimation of such models in multinomial regression also remains an active area of statistics research (Linderman, Johnson, and Adams 2015).

I have shown here that a straightforward application of a clustering model can be applied to illuminate patterns and identify groups from complex voting data. As I have discussed, these tools should not be a substitute for substantive theorizing and interpretation, but they can facilitate discovery, provide a more principled measurement of size and voting propensity, and improve theory building by providing data-based guidance.

A | Appendix to Chapter 3

Appendix A.1: Data Construction

Subsection A.1.1 describes the dataset construction. Subsection A.1.2 describes the search specifications used to collect the number of newspaper article hits for a given candidate. Subsection A.1.3 describes the data collection procedure and coverage for campaign finance data, and Subsection A.1.4 describes how both the newspaper measures and campaign finance data are summarized to form the values used in the regressions.

A.1.1 Data Construction

The final dataset I analyze is in wide-form, with one observation for a given voter in one South Carolina election. I compile this dataset in the following steps.[1]

1. I first process the set of raw cast vote records to standardize the names of offices across precincts. While the state commission oversees elections for all offices, county board of elections apparently finalize their ballots separately, leading to variations in spelling for offices, even statewide ones.

2. Second, I identify the affiliated party of candidates that voters vote for using a separate roster of certified candidates. Although party information is presented

[1] These procedures are implemented in an open-source R package.

in each touchscreen to the voter, only the chosen candidate's name appears in the log. I merge the party affiliation to each name chosen and, given the purposes of this study.

One of the more difficult tasks in data processing is to determine which races were available to which voters' ballots, and whether or not the race was contested. The combination of different legislative, school, and special purpose districts leads to a proliferation of different ballot styles (i.e., a layout for which contested for a given voter). Each entry in the logs contain a precinct identifier as well as a ballot style identifier unique to each precinct. Although there are around 2,000 to 2,200 precincts in each general election, there are at least 5,000 different ballot styles.

3. To infer the layout of each style, I aggregate the individual logs from the bottom-up. For each precinct and ballot style combination, I tabulate the votes cast for each candidate. When working with contests for offices that held elections for only a subset of voters, I denote that this office did not exist for a given precinct - ballot style if no voter in that set cast a vote for the office. This way, I distinguish abstentions from the lack of existence of the contest.

   One side-effect of this procedure is that absentee votes are *not* counted, because the voting machine codes them with a virtual precinct at the county-level, thereby effectively erasing information about the precinct of the absentee ballot. Until 2018, South Carolina voters had to be over 65 or have an "excuse" for not be able to vote on election day to apply for an absentee ballot. In the five general elections studied here, 17.7 percent of the 8.4 million ballots cast were absentee ballots.

4. I then aggregate votes at the district level, and declare a district as contested if

votes for both the Republican and Democrat exist.

A.1.2 Data on Newspaper Mentions

For a measure of name familiarity I compute the number of state newspaper articles that mention the candidate's name.

- I search the 86 newspapers in South Carolina available in NewsLibrary.com.
- I used the length of the office's term ending the day before the election. For example, for U.S. Senate candidates running against each other in the November 6, 2018 election, I search the dates November 5, 2012 to November 5, 2018, and for U.S. House candidates I would use a two-year timeframe. I do not include election day to prevent biasing counts towards the eventual winner.
- I search the official name on the ballot. In case of middle name or first name initials, I also include a version that removes the initial. For example, for "Nikki R Haley", I search for the term `("Nikki R Haley") OR ("Nikki Haley")`.
- I generally do not restrict to specific election-related or office-related terms, with the following exceptions: County Council members, Sheriffs, and Probate Judge searches are further restricted by the county of the office. This measure, then, aims to captures general name recognition with some filters added to prevent miscounting common names (like "David Smith") as mentions of candidates.

Figure A.1 shows the distribution of the counts, and some illustrative examples follow.

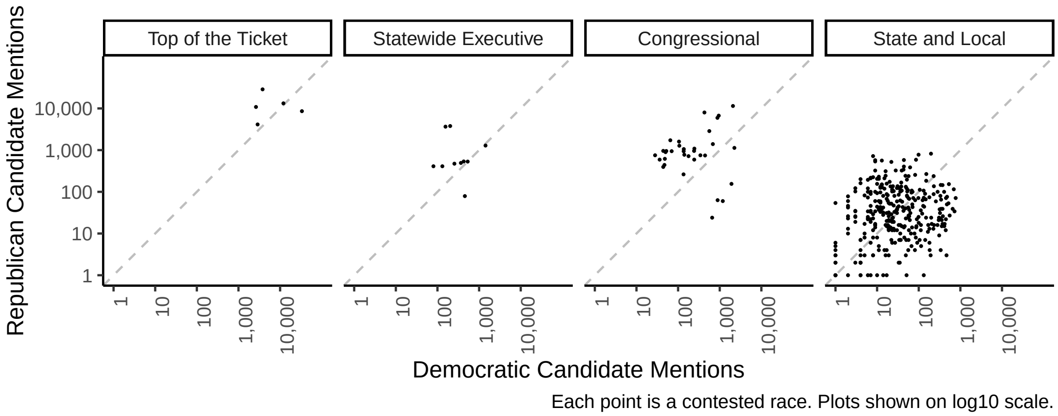


Figure A.1: Newspaper Mentions in Contested Elections.

Note: Each point represents the number of South Carolina newspaper articles mentioning the name of the Democrat (x-axis) or Republican (y-axis) candidate in a contested race.

Contests and candidates with the highest candidate counts in each facet in Figure A.1 are listed below.

Top of the Ticket

		Democrat		Republican	
Year	Office	Name	#	Name	#
2012	President	Barack Obama	33,384	Mitt Romney	8,588
2014	Governor	Vincent Sheheen	3,792	Nikki R Haley	28,660
2016	President	Hillary Rodham Clinton	12,021	Donald J Trump	13,237
2018	Governor	James Smith	2,633	Henry McMaster	10,834
2010	Governor	Vincent A Sheheen	2,882	Nikki R Haley	4,129

Congressional

		Democrat		Republican	
Year	Office	Name	#	Name	#
2014	US Senate	Brad Hutto	2,078	Lindsey Graham	11,431
2016	US Senate	Thomas Dixon	428	Tim Scott	7,991
2010	US Senate	Alvin M Greene	944	Jim DeMint	6,707
2014	US Senate	Joyce Dickerson	879	Tim Scott	5,945
2010	US House	Rob Miller	561	Joe Wilson	2,861
2010	US House	John Spratt	2,247	Mick Mulvaney	1,130
2010	US House	James E Jim Clyburn	1,912	Jim Pratt	154
2016	US House	Dimitri Cherny	64	Mark Sanford	1,721
2012	US House	Bobbie Rose	104	Tim Scott	1,601
2018	US House	Archie Parnell	683	Ralph W Norman	1,395

State and Local

		Democrat		Republican	
Year	Office	Name	#	Name	#
2012	State Senate	Paul Tinkler	196	Paul Thurmond	820
2012	State Senate	Robert Rikard	98	John Courson	776
2012	Sheriff	Barry Faile	770	Scott R Case	70
2016	State Senate	Nikki Setzler	723	Brad Lindsey	33
2016	Sheriff	William McCoy	7	Barry Faile	718
2016	Sheriff	Alex Underwood	694	Richard Smith	114
2016	Sheriff	David H Taylor	652	Jeff D Bailey	38
2012	County Council	Jack T Collins	85	Joel R Thrift	587
2016	Auditor	Peter J Tecklenburg	43	Elizabeth Moffly	578
2018	State House	Marty R Cotton	8	Thomas E Tommy Pope	572

A.1.3 Data on Campaign Contributions

Contribution data is collected from official sources but measured with error. These come from two sources, the organization Follow The Money and the South Carolina State Ethics Commission.

For the offices of US House, State Senate, and State House, I drew values curated by Follow The Money (`https://followthemoney.org`). This organization collects the contributions reported to the FEC and state election commissions, and for each candidate in each election cycle, reports the total campaign contribution a candidate received.

For the offices of Sheriff, Probate Judge, and County Council, I collected data from the state Ethics Commission (`https://apps.sc.gov/PublicReporting`). These contain all candidate finance disclosure reports a campaign for state office has filed. Candidates can file these reports at any time in the cycle, so when multiple reports I take the most recent value of "Total Contribution for Election Cycle" (including in-kind contributions) no later than the January following the election (all from the appropriate election cycle). I ignore finances listed under the primary election and set the total to zero if no record is found, which is the case in a handful of local races.

Some campaign finance reports in state and local races report zero contributions and zero expenditures. Perhaps those candidates in fact did no fundraising and spent no money on their campaigns. However, that state and local parties may have supported the candidate in ways that do not appear on the report, such as for booking meeting centers for candidate meetings with candidates for multiple offices. Campaign finances are difficult to track in state and local races anywhere but especially in South Carolina, where a 2010 district court case effectively exempted independent political action committees from disclosing their donors, which include party organizations (*South Carolina Citizens for Life v. Kenneth C. Krawcheck et al.*).

A.1.4 Summary Statistics for the Valence Advantage

This subsection formalizes how the measure of the valence advantage is constructed.

Incumbency is a binary variable, taking 1 if the Republican candidate is an incumbent and 0 otherwise. This variable is only used in contested races.

Newspaper coverage (measured in number of article hits) and contributions (measured in dollars) are denoted by the continuous variable $v^{a}_{d[j]}$: the valence measure of candidate a in district $d[j]$. We quantify the Republican advantage for both measures of valence as the natural log of the ratio, specifically

$$\begin{aligned} \text{Advantage}^{\text{R}} &= \log\left(\frac{v^{\text{R}}_{d[j]}+1}{v^{\text{D}}_{d[j]}+1}\right) \\ &= \log(v^{\text{R}}_{d[j]}+1) - \log(v^{\text{D}}_{d[j]}+1). \end{aligned}$$

Taking the ratio is appropriate because the relevant comparison is between two candidates competing against each other. Taking the log reduces the impact of outliers and allows for both a ratio and difference interpretation. Adding one to each value before taking the log prevents the few candidates that have zero news article hits or report no campaign contributions from causing divide-by-zero errors.

Table A.1 provides summary statistics of this measure for each office. Because log values are difficult to interpret substantively, the table also shows the exponentiated version of the summary statistics. This is approximately equivalent to the actual ratio,

$$\exp\left(\text{Advantage}^{\text{R}}\right) = \frac{v^{\text{R}}_{d[j]}+1}{v^{\text{D}}_{d[j]}+1} \approx \frac{v^{\text{R}}_{d[j]}}{v^{\text{D}}_{d[j]}}.$$

Table A.1: Summary Statistics Newspaper Coverage and the Fundraising Advantage

Republican Newspaper Coverage Advantage Summary Statistics

	Log Values					Exponentiated				
Office	Mean	10th	Median	90th	S.D.	Mean	10th	Median	90th	N
US House	1.24	-2.54	1.70	2.94	1.92	3.46	0.08	5.49	18.94	31
State House	0.42	-2.15	0.46	2.89	1.88	1.52	0.12	1.59	18.02	161
State Senate	-1.32	-3.29	-1.41	2.09	2.23	0.27	0.04	0.24	8.10	19
County Sheriff	-0.29	-2.88	-0.49	2.81	2.20	0.75	0.06	0.61	16.61	19
Probate Judge	0.42	-0.72	0.00	1.66	1.01	1.53	0.49	1.00	5.24	19
County Council	-0.30	-2.66	-0.46	2.43	1.96	0.74	0.07	0.63	11.37	101

Republican Fundraising Advantage Summary Statistics

	Log Values					Exponentiated				
Office	Mean	10th	Median	90th	S.D.	Mean	10th	Median	90th	N
US House	1.15	-4.51	2.49	4.26	5.45	3.17	0.01	12.04	70.86	31
State House	1.05	-2.46	0.95	4.37	3.41	2.86	0.09	2.58	78.82	161
State Senate	-0.86	-2.72	-1.20	1.34	2.01	0.42	0.07	0.30	3.82	19
County Sheriff	0.77	-0.78	0.52	2.19	1.24	2.16	0.46	1.67	8.96	19
Probate Judge	0.48	-1.18	0.08	1.78	2.35	1.61	0.31	1.08	5.90	19
County Council	0.39	-2.68	0.00	4.96	3.34	1.48	0.07	1.00	142.98	101

Note: Tables show mean, 10th percentile, median, 90th percentile, standard deviation, and sample size for the respective measure by office. Each observation is measured at the contest level. Exponentiated versions approximate the quantity in their original ratio form.

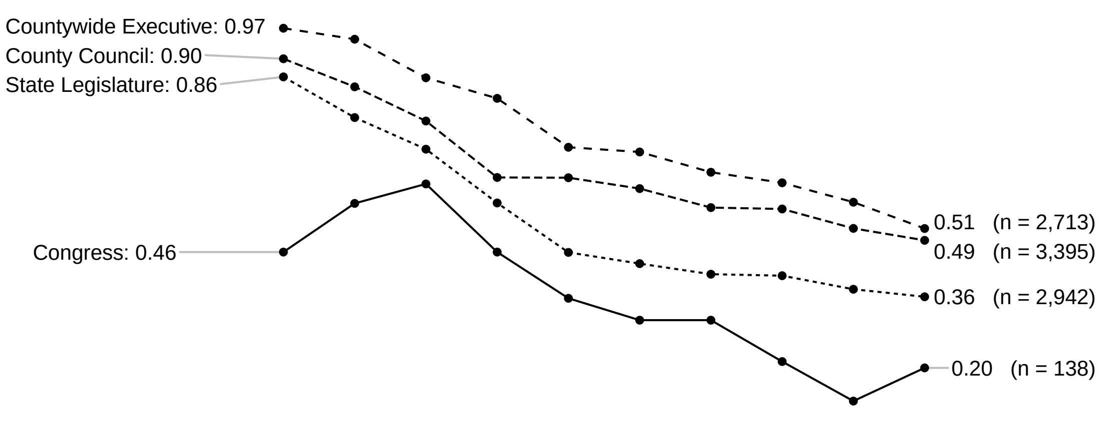
Appendix A.2: Elections in South Carolina

A.2.1 Historical Trends in South Carolina Elections

Table A.2 summarizes the election results from the past four decades in South Carolina. State legislative results (-2016) come from Klarner (2021), and countywide executive results between 1980-1996 come from Lublin (2007) and were generously provided by David Lublin.

Table A.2: Party Choice from Election Returns, South Carolina 1980 - 2018

	General Elections										
	1980 & 1982	1984 & 1986	1988 & 1990	1992 & 1994	1996 & 1998	2000 & 2002	2004 & 2006	2008 & 2010	2012 & 2014	2016 & 2018	
Statewide Vote Margin											
President	R+01	R+28	R+24	R+04	R+06	R+16	R+17	R+09	R+10	R+14	$n = 10$
Governor	D+40	R+03	R+42	R+02	D+08	R+06	R+11	R+04	R+14	R+08	$n = 10$
Proportion of Seats Captured by Democrats											



Note: *President* and *Governor* entries indicate the winning party and win margin in percentage points. Lineplot values are aligned with the table and each line's endpoints show values at the first and last set of elections. *Countywide Executive* offices are the six offices of sheriff, probate judge, clerk of court, auditor, treasurer, and coroner, all of which are elected countywide. The last column indicates the total number of contests comprising each row.

A.2.2 Elections and Offices Covered

Table A.3 shows the extent of the iVotronic data examined in this article. I show the offices up for election, the number of contested races, and the number of precincts and voters for each general election year.

Table A.3: Extent of Elections Analyzed through Cast Vote Records

	General Elections					
	2010	2012	2014	2016	2018	Total
Availability of Contested Statewide Races						
President		✓		✓		
Governor	✓		✓		✓	
US Senate	✓		✓	✓		
Attorney General	✓		✓		✓	
Secretary of State	✓		✓		✓	
State Superintendent	✓		✓			
Number of Contested District Races						
US House	6	5	5	7	7	30
State Senate	0	11	0	7	1	19
State House	34	20	29	32	44	159
County Sheriff	1	3	0	14	1	19
Probate Judge	6	2	4	0	7	19
County Council	26	8	20	27	19	100
Sample Size						
Counties	43	45	46	46	46	226
Precincts	2,001	2,115	2,216	2,232	2,245	10,809
Voters	1,101k	1,501k	1,058k	1,570k	1,414k	6,644k

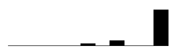
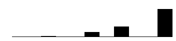
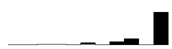
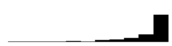
Note: Numbers are from data after pre-processing, detailed in Appendix A.2. The number of voters in each election are shown in the last row, thousands. Two counties from 2010 and one county from 2012 is missing from the files released by the state election commission.

Appendix A.3: Additional Findings

Subsection A.3.1 shows the distribution of proportion straight ticket with and without the people who pulled the party lever, Subsection A.3.2 shows additional descriptive statistics by office and election, Subsection A.3.3 shows diagnostics to choose the number of clusters, Subsection A.3.4 shows additional regression results for the incumbency advantage, and Subsection A.3.5 shows estimates of overtime change in straight ticket voting.

A.3.1 Distribution of Party Splits

Table A.4: Straight Ticket Rates by Number of Contested Contests

	(i) All Voters			(ii) Voters opting *out* of the Party Lever		
Contested Races	Straight	Distribution	n	Straight	Distribution	n
3	0.80		1,093k	0.66		570k
4	0.82		1,295k	0.64		559k
5	0.77		732k	0.56		322k
6	0.75		299k	0.52		134k
7	0.64		965k	0.39		505k
8	0.65		897k	0.42		471k
9 - 12	0.62		380k	0.35		199k

Note: Each proportion shows the fraction of voters who voted for the same party for all contested races on their ballot, with number of voters (n) counted in 1000s. Each histogram shows the distribution of a person's vote for a favored party as a fraction of the contested races on their ballot. In all graphs, axes range from 0 to 100 percent; therefore the height of the rightmost bar corresponds to the "Straight" proportion.

A.3.2 Split Ticket Voting by Election, Party, and Office.

Figure A.2 shows the breakdown of straight ticket, split ticket, third party voting and abstention by office, election, and party choice at the top of the ticket. They serve as a more granular version than Figure 3.2 and show sample sizes.

Each subplot shows the vote choice composition for a contested down-ballot office. Colors indicate Republican (red), Democrat (blue), third-party or write-in (yellow), and abstention (gray). White number shows the pairwise straight-ticket rate.

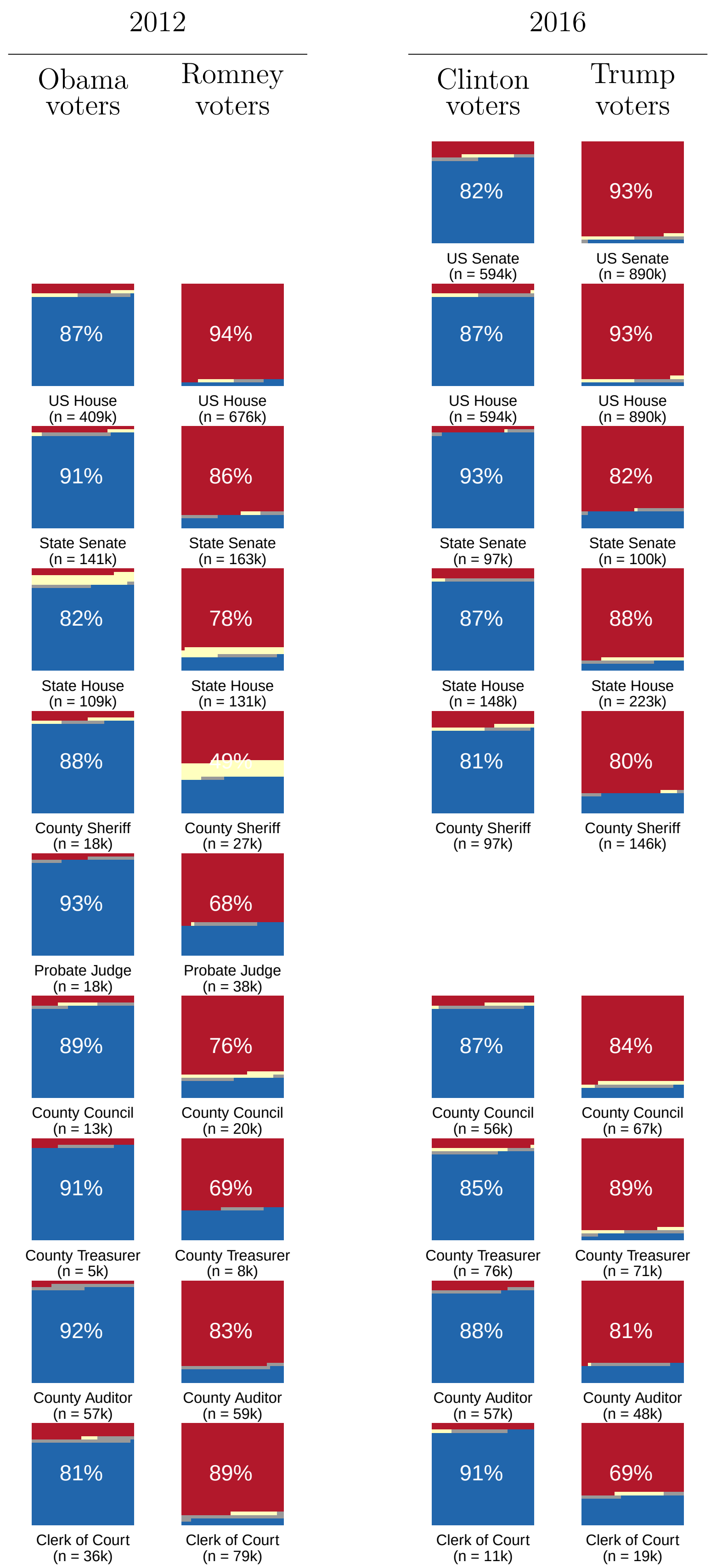
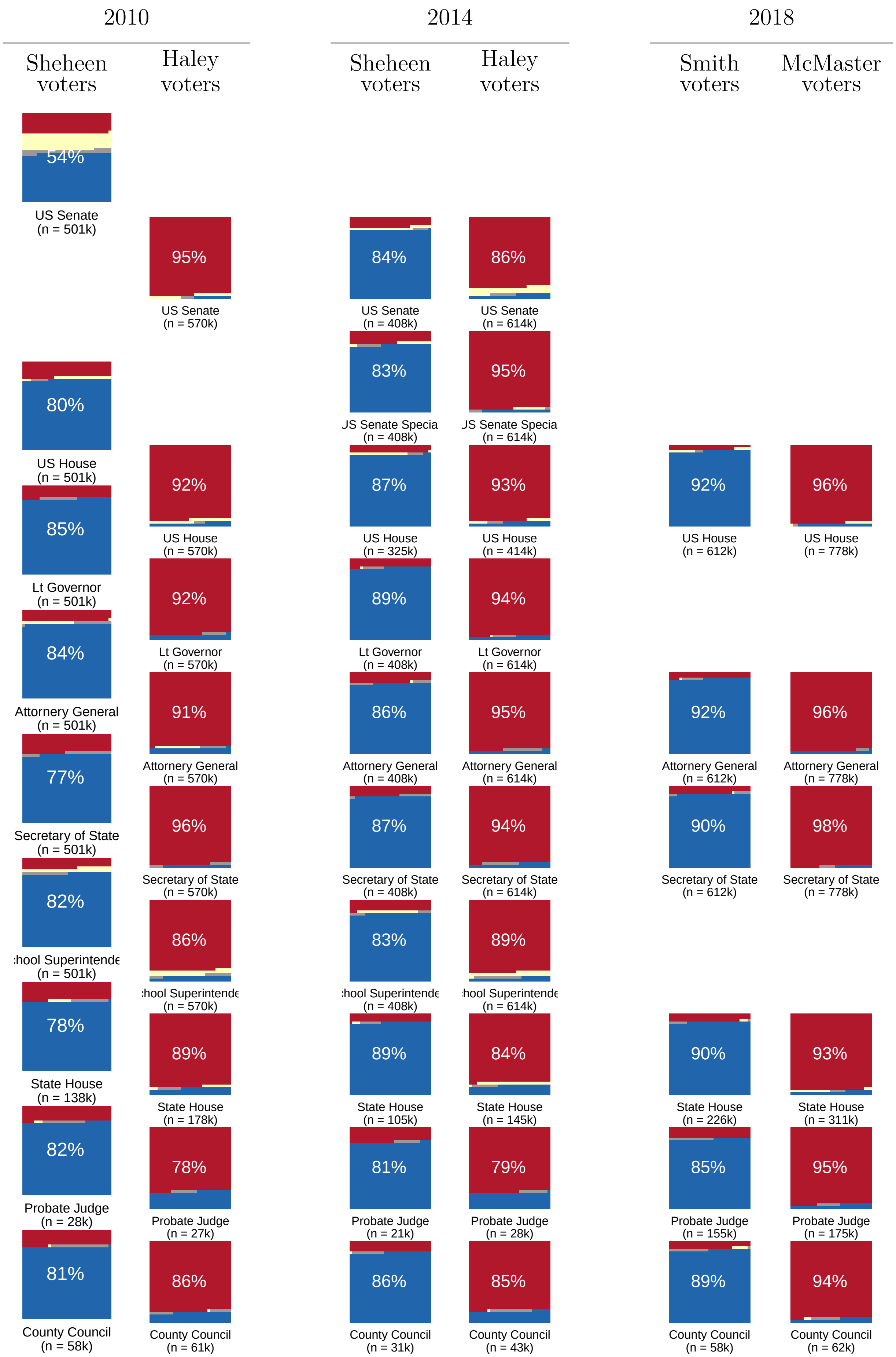


Figure A.2: Party Loyalty on the Long Ballot.

(Figure A.2 continued)



A.3.3 Clustering Algorithim

In Figure A.3 I show the change in the model fit statistic for varying levels of the number of clusters modeled. The statistic used is the Bayesian Information Criterion (BIC) defined in Fraley and Raftery (1998). For a given choice K,

$$\text{BIC}_K = 2\widehat{\mathcal{L}} - \log(K + JKL)\log(N) \tag{A.1}$$

where $\widehat{\mathcal{L}}$ is the observed log likelihood with the estimated parameters, J is the number of offices, L is the number of choices in a given race, and N is the number of voters. In this application, $K + JKL$ amounts to the number of parameters estimated.

Higher numbers indicate better fit, with a penalty for too many parameters. In the Figure, I draw a line at the value of K which roughly marks where the improvement in the BIC starts to plateau.

A.3.4 The Valence Advantage: Additional Results

In this set of results, we estimate the following equation for each voter i making a choice for race j on their ballot:

$$\text{Y}_{ij}^{\text{REP}} = \alpha + \beta\text{D}_i + \gamma\text{V}_{i,d[ij]} + \varepsilon_{d[ij]}. \tag{A.2}$$

Here the binary outcome $\text{Y}_{ij}^{\text{REP}}$ is 1 if individual i votes for the Republican candidate in a contested election for an office j, and 0 if she votes for another party's candidate, writes in, or abstains. The predictor variables follow equation 3.1 in the main text. This choice of the outcome variable is slightly more convenient to present in table setting, because the expected direction of the party and valence variables are clear. We present the main results, by office, in Table A.5.

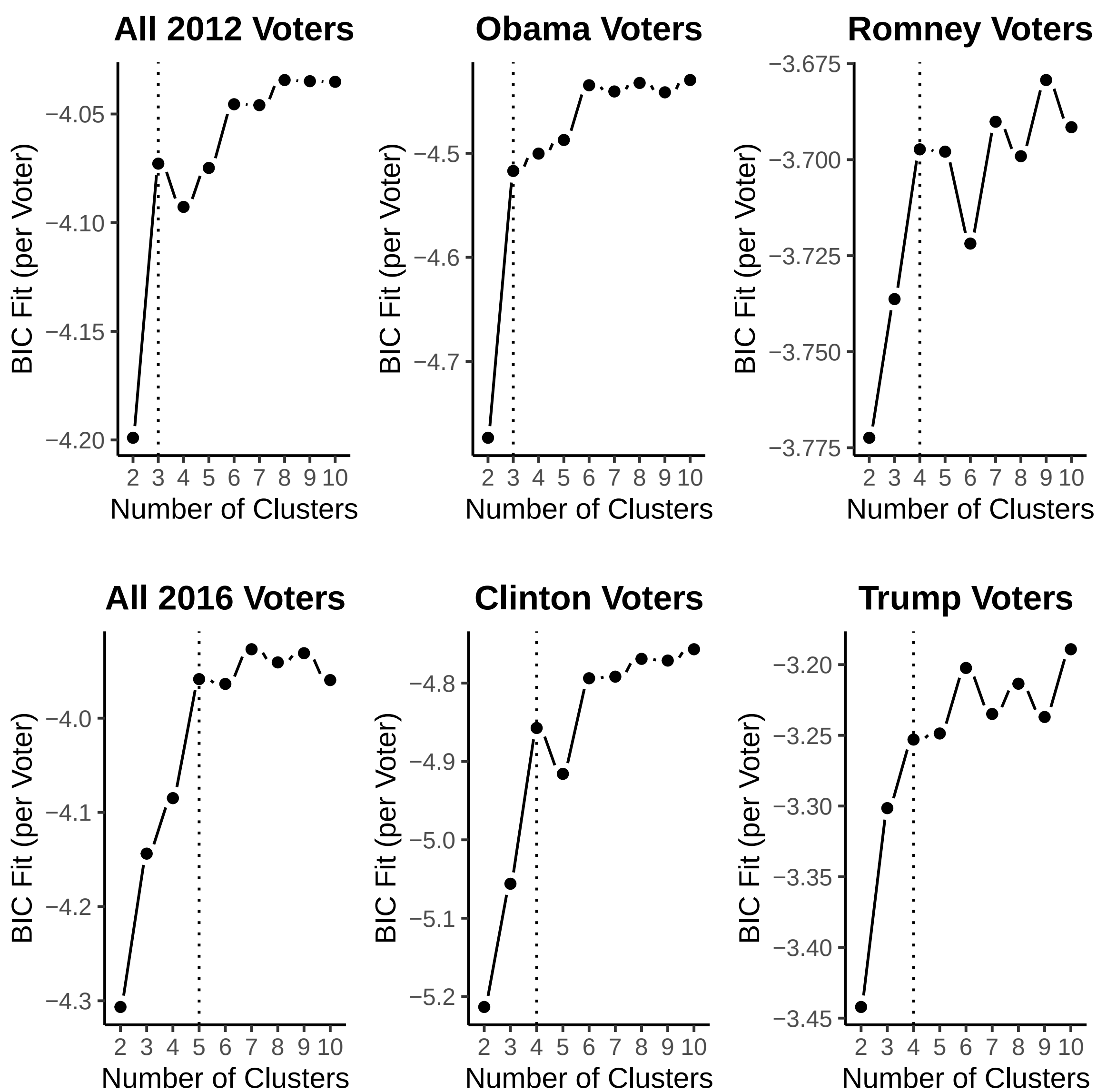


Figure A.3: Model Fit by Number of Clusters

Figure A.4 shows predicted probabilities from logit versions of the linear probability model presented in Table A.5. Table A.6 conducts similar regressions as that of Table A.5 but only among those who did not use the party lever.

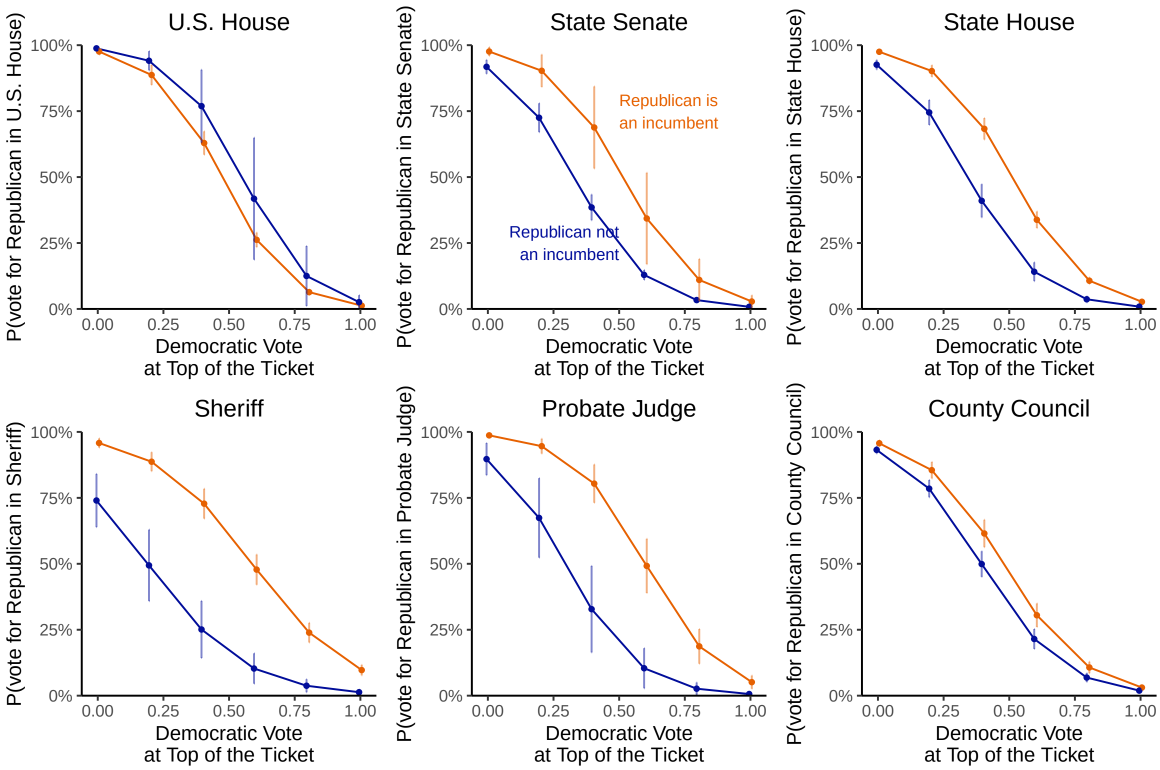


Figure A.4: Predicted Probabilities from Logit Regressions for Table A.5

Note: Each panel shows predicted probabilities from regressions of the same specification as the top panel of Table A.5, but using a logit regression instead of a linear probability model. I show predicted probabilities along equidistant values of the Democratic top of the ticket vote fraction (on the x-axis), separately for party affiliation of the incumbent (orange for Republican, blue for Democrat), and other predictors held at their mean. 95 percent confidence intervals are shown for each point.

Table A.5: Valence Advantages in State and Local Elections

(a) Contested Contests with an Incumbent

	Outcome: Vote for Republican for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Incumbent	0.024	0.11*	0.059*	0.25*	0.18*	0.034
	(0.02)	(0.05)	(0.02)	(0.04)	(0.02)	(0.02)
Republican Newspaper Coverage Advantage	-0.0078	-0.015	-0.0035	-0.0089	-0.0021	0.0068
	(0.007)	(0.01)	(0.003)	(0.01)	(0.006)	(0.004)
Republican Fundraising Advantage	0.010*	0.011	0.012*	0.020	-0.00078	0.0019
	(0.003)	(0.007)	(0.003)	(0.01)	(0.002)	(0.001)
Democratic Vote Top of the Ticket	-1.06*	-0.98*	-1.00*	-0.84*	-0.98*	-0.97*
	(0.005)	(0.02)	(0.008)	(0.04)	(0.02)	(0.01)
Mean of Outcome	0.60	0.46	0.57	0.54	0.54	0.53
R-squared	0.73	0.61	0.66	0.52	0.62	0.62
Clusters	20	12	113	12	10	70
Observations	4,114,796	346,529	1,246,364	238,949	369,369	333,860

(b) Open-Seat Races (No Incumbent)

	Outcome: Vote for Republican for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Newspaper Coverage Advantage	-0.0098	0.0037	-0.0092	0.046	0.022	0.0090
	(0.006)	(0.007)	(0.01)	(0.03)	(0.010)	(0.006)
Republican Fundraising Advantage	0.0079	0.0087	0.027*	0.099*	0.043	0.00014
	(0.004)	(0.006)	(0.007)	(0.03)	(0.02)	(0.001)
Democratic Vote Top of the Ticket	-1.05*	-1.04*	-1.01*	-0.85*	-0.99*	-1.01*
	(0.01)	(0.01)	(0.02)	(0.10)	(0.03)	(0.01)
Mean of Outcome	0.60	0.48	0.55	0.53	0.53	0.51
R-squared	0.68	0.71	0.64	0.55	0.66	0.64
Clusters	6	7	30	6	9	27
Observations	1,204,029	205,721	346,203	90,466	129,171	125,091

Note: Each column is a regression from a linear probability model for one down-ballot office (equation 3.1), standard errors clustered by district in parentheses. Intercept and fixed effects for each election year not shown. “Democratic Vote Top of the Ticket” is the proportion of Democratic votes in the office of President, U.S. Senate, Governor, and the party lever, where applicable. Valence attributes are explained in the main text. $^{*}p < 0.05$

Table A.6: Replication of Table A.5 among Voters Who Did Not Pull the Party Lever

(a) Contested Contests with an Incumbent

	Outcome: Vote for Democrat for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Incumbent	0.039	-0.21*	-0.20***	-0.42***	-0.33**	-0.088**
	(0.04)	(0.09)	(0.02)	(0.07)	(0.09)	(0.03)
Republican Newspaper Coverage Advantage	-0.025**	0.013	0.00063	-0.0070	-0.0082	-0.0066
	(0.007)	(0.02)	(0.004)	(0.008)	(0.02)	(0.006)
Democratic Vote Top of the Ticket	1.33***	1.11***	1.19***	0.68***	1.05***	1.06***
	(0.02)	(0.07)	(0.02)	(0.07)	(0.08)	(0.03)
Mean of Outcome	0.31	0.47	0.37	0.38	0.43	0.39
R-squared	0.52	0.33	0.40	0.30	0.29	0.31
Clusters	23	12	123	13	10	70
Observations	2,257,726	177,912	675,629	137,497	167,559	163,615

(b) Open Races (No Incumbent)

	Outcome: Vote for Democrat for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Newspaper Coverage Advantage	0.039	-0.0083	-0.041	-0.15**	-0.013	-0.022*
	(0.07)	(0.01)	(0.03)	(0.02)	(0.02)	(0.009)
Democratic Vote Top of the Ticket	0.37***	0.35***	0.33***	0.16	0.27***	0.35***
	(0.007)	(0.02)	(0.02)	(0.09)	(0.04)	(0.02)
Mean of Outcome	0.32	0.41	0.34	0.40	0.39	0.39
R-squared	0.48	0.46	0.35	0.24	0.27	0.34
Clusters	6	7	31	5	9	28
Observations	586,211	101,163	190,803	36,753	58,798	68,010

Note: See Table A.5.

Table A.7: Replication of Table A.5 excluding Abstentions and Third Party Vote

(a) Contested Contests with an Incumbent

	Outcome: Vote for Republican for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Incumbent	-0.027	0.12*	0.056*	0.26*	0.17*	0.042*
	(0.02)	(0.05)	(0.02)	(0.05)	(0.03)	(0.02)
Republican Newspaper Coverage Advantage	0.0051	-0.018	-0.0044	-0.010	0.00057	0.0042
	(0.004)	(0.01)	(0.003)	(0.01)	(0.009)	(0.003)
Republican Fundraising Advantage	0.0049*	0.013	0.013*	0.023	-0.0015	0.0023
	(0.001)	(0.008)	(0.002)	(0.01)	(0.002)	(0.002)
Democratic Vote Top of the Ticket	-1.08*	-0.99*	-1.01*	-0.85*	-0.98*	-0.98*
	(0.006)	(0.02)	(0.008)	(0.04)	(0.02)	(0.01)
Mean of Outcome	0.62	0.47	0.58	0.55	0.55	0.55
R-squared	0.78	0.63	0.69	0.53	0.64	0.66
Clusters	20	12	113	12	10	70
Observations	3,985,095	339,066	1,221,472	235,150	362,663	325,231

(b) Open-Seat Races (No Incumbent)

	Outcome: Vote for Republican for					
	U.S. House	State Senate	State House	Sheriff	Probate Judge	County Council
Republican Newspaper Coverage Advantage	-0.014*	-0.0055	-0.0073	-0.026*	0.0069	0.010
	(0.00006)	(0.006)	(0.01)	(0.006)	(0.007)	(0.007)
Republican Fundraising Advantage	0.020*	0.016*	0.025*	0.22*	0.012	0.0013
	(0.00005)	(0.006)	(0.007)	(0.00004)	(0.03)	(0.001)
Democratic Vote Top of the Ticket	-1.07*	-1.06*	-1.03*	-0.86*	-0.99*	-1.02*
	(0.006)	(0.01)	(0.02)	(0.10)	(0.03)	(0.01)
Mean of Outcome	0.62	0.50	0.58	0.55	0.54	0.52
R-squared	0.75	0.74	0.69	0.58	0.67	0.66
Clusters	6	7	30	6	9	27
Observations	1,157,884	199,425	327,247	86,382	127,615	121,987

Note: See Table A.5.

A.3.5 Overtime Change in Party Loyalty

The ballot image data is relatively limited in scope for fully test whether elections are nationalizing, because only certain elections occur simultaneously in the same election. Moreover, changes in candidates as well as electorates across elections make it difficult to attribute changes in district level partisan voting to changes in any individual voters' preferences. I therefore examine the same-party voting rates between the U.S. House and State House, which are up for election every general election.

Table A.8 shows how the degree of same-party voting in the U.S. House and the State House has changed over time for four generals elections. Both specifications show a strong increasing trend of partisan allegiance at the house district level of about 2 to 3 percentage points every general election cycle.

Table A.8: Straight-Ticket Voting Overtime

	(1)	(2)
Time (2-Year Increment)	0.032*	0.018*
	(0.009)	(0.005)
Midterm Year	0.050*	0.042*
	(0.02)	(0.009)
Constant	0.76*	0.80*
	(0.02)	(0.010)
Fixed Effects by	Congressional District	House District
Average of Outcome	0.86	0.86
Std. Dev. of Outcome	0.11	0.11
R-squared	0.22	0.94
Observations	151	151

Note: Each column is a linear regression with the pairwise straight-ticket rate between the U.S. House and State House in each *State House* district. The "time" variable is computed by (year - 2012) / 2, and thus its coefficient indicates the change in the rate in the next general election. Midterm year is a binary indicator. Standard errors in parentheses. $^*p < 0.05$.

Appendix A.4: Cast Vote Records from Other States

Maryland, 2018 Table A.9 shows aggregate ticket splitting rates in the 2018 General Election in the state of Maryland, using only contested races. We see that ticket splitting rates tend to be higher in state and local offices than in the US House. Two other patterns stand out. First, the ticket splitting rate in the race for Governor is remarkably high, owing to the popularity of the incumbent Governor Larry Hogan. Second, the pattern in state and local races is more prominent among supporters of the Democratic US Senate candidate.

Table A.9: Ticket Splitting Rates in the Maryland 2018 General Election

	Ticket Splitting Rates				
Office	All	among Democrats	among Republicans	Districts	Total Voters
US House	0.040	0.036	0.050	8	1,669,493
Governor	0.224	0.338	0.007	1	1,961,532
State Attorney General	0.044	0.051	0.033	1	1,961,825
State Comptroller	0.094	0.040	0.214	1	1,960,895
State Senate	0.068	0.092	0.037	32	965,741
State Attorney (Solicitor)	0.093	0.127	0.056	6	541,090
Register of Wills	0.081	0.096	0.071	9	1,007,618
Sheriff	0.104	0.136	0.068	12	1,097,558
County Council	0.088	0.126	0.048	36	699,013

Note: Reference category is the race for US Senate, where incumbent Ben Cardin won with 65 percent of the vote. "Democrat" and "Republican" in the headers are shorthand for the party vote in the reference category.

Palm Beach County Florida, 2000 Figure 5.1 shows the cluster analysis results of voting patterns in the 2000 General Elections in the state of Florida, using only contested statewide or countywide races. Though from an earlier time period than the one studied in this paper, this electorate also saw higher levels of ticket splitting in state and local offices.

B | Appendix to Chapter 4

Appendix B.1: Survey Variables and Census Codes

The ACS Table `B01001` provides distributions of `[age x sex x education x CD]`. The Table `B15001` provides the distribution of `[age x sex x education x CD]`. These are then collapsed to match the values of how the CCES asks race and education. Education is collapsed to four categories (`High School or Less`, `Some College` (including 2-Year Degrees), `4-Year College Graduate`, and `Post-Graduate Degree`). Race is collapsed to four categories (`(Non-Hispanic) White`, `Black`, `Hispanic`, and `All Others`). Party registration, in states where it is recorded, is collapsed into `Registered Democrats`, `Registered Republicans`, and `All Others` (including third parties and non-affiliated voters).

Data recoding, download, and reshaping is performed by the `ccesMRPprep` package (Kuriwaki 2021a), which partly relies on the `tidycensus` package (Walker and Herman 2021) for drawing ACS data. All survey recodings are available in the `ccesMRPprep` package.

Appendix B.2: Multilevel Regression

For the outcome model, I use a multilevel logit regression estimated through the brms package (Bürkner 2021). For the baseline specification, I run

```
trump ~ female + race + educ + age + (1 | state/cd)        (B.1)
```

in the brms or lme4 notation. Here the parentheses `(|)` indicate varying intercepts estimated by random effects, and the forward slash `/` indicates a nested random effect such that both a state-specific varying intercept and a state-congressional district varying intercept are estimated.

I set standard normal priors for all coefficients, and take 2,000 MCMC samples taken across 4 chains. I construct credible intervals by these 2000 posterior samples.

The "no partial pooling" model (Figure 4.5 Model 3) includes no demographic variables, but estimates the random effect model

```
trump ~ (1 | state/cd)        (B.2)
```

and predicts on the general voting age population.

The "off the shelf" model (Figure 4.5 Model 4) includes the three variables whose joint distribution is given by ACS standard datasets, and fits on the synthetic population target subsetted to the turnout electorate.

```
trump ~ female + age + educ + (1 | state/cd)        (B.3)
```

The party registration model (Figure 4.5 Model 6) simply adds party registration to the list of covariates

$$\texttt{trump} \sim \texttt{female + race + educ + age + partyreg + (1 | state/cd)} \tag{B.4}$$

and also predicts on to the synthetic turnout electorate.

C | Appendix to Chapter 5

Appendix C.1: Package Speed Performance

Figure C.1 shows the time required to complete the clustering algorithm on datasets of varying sizes and four alternatives for the number of clusters. I take the entire dataset by Herron and Lewis (2007), which includes data on 3 million voters in 10 Florida counties. I only consider four offices ($J = 4$) and vary the number of voters to analyze.

Time measurements in seconds are plotted on a log-log scale. Both clusterCVR (the method I use in this paper) and the R package poLCA scale relatively well, but poLCA can be an order of magnitude slower in very large datasets.

Because of the collapsing procedure I describe below, the algorithm iterates through unique profiles, instead of individual voters. Without the collapsing, both algorithms will take orders of magnitude longer in large datasets.

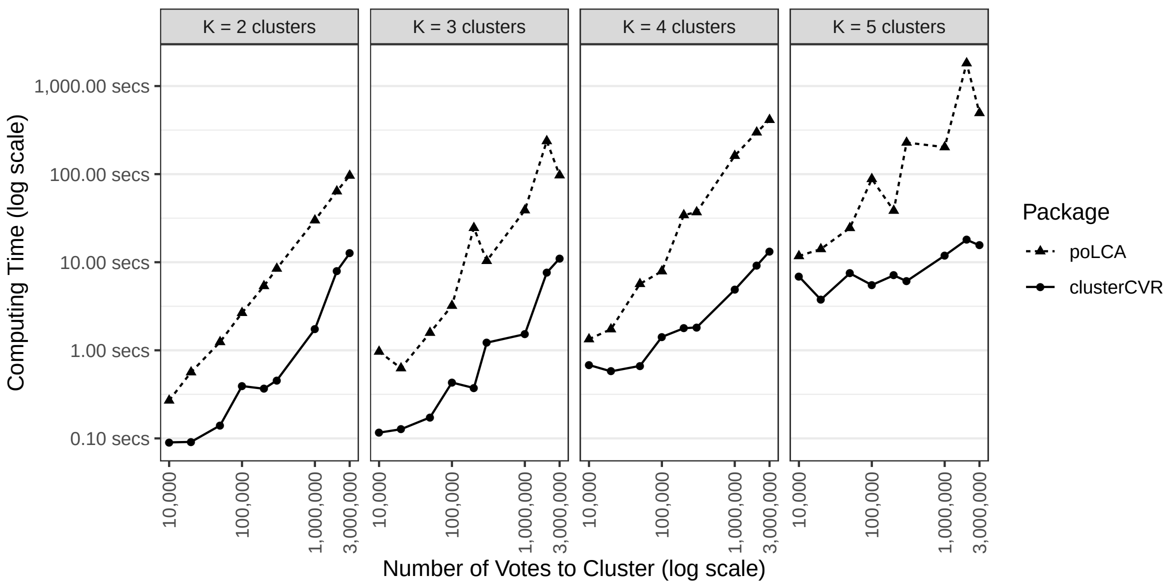


Figure C.1: Scalability of the Clustering Algorithm

Appendix C.2: EM Derivations

This appendix formalizes and derives the key steps of the clustering model and EM algorithm. The notation follows the main text.

C.2.1 Complete Likelihood

If we knew the cluster assignment, we would be able to write the complete log-likelihood ($\mathcal{L}_{\text{comp}}$). First start with the joint probability of the outcome data and the cluster assignment:

$$\begin{aligned}\Pr(\mathbf{Y}, \mathbf{Z} \mid \boldsymbol{\mu}, \boldsymbol{\pi}) &= \Pr(\mathbf{Y} \mid \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\pi}) \Pr(\mathbf{Z} \mid \boldsymbol{\pi}) \\ &= \prod_{i=1}^{N} \prod_{j=1}^{J} \Pr(Y_{ij} \mid \mathbf{Z}, \boldsymbol{\mu}) \prod_{i=1}^{N} \Pr(Z_i \mid \boldsymbol{\pi}) \\ &= \prod_{i=1}^{N} \prod_{j=1}^{J} \prod_{k=1}^{K} \left\{ \prod_{\ell=0}^{L} \Pr\left(Y_{ij} = \ell | Z_i = k\right)^{\mathbf{1}(Y_{ij}=\ell)} \right\}^{\mathbf{1}(Z_i=k)} \prod_{i=1}^{N} \prod_{k=1}^{K} \Pr\left(Z_i = k \mid \boldsymbol{\pi}\right)^{\mathbf{1}(Z_i=k)}\end{aligned}$$

Therefore, the complete log-likelihood is:

$$\begin{aligned}\mathcal{L}_{\text{comp}}(\boldsymbol{\mu}, \boldsymbol{\pi} | \mathbf{Y}, \mathbf{Z}) = &\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=0}^{L} \mathbf{1}\{Y_{ij} = \ell, Z_i = k\} \log \Pr(Y_{ij} = \ell | Z_i = k, \boldsymbol{\mu}) \\ &+ \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbf{1}\{Z_i = k\} \log \Pr(Z_i = k | \boldsymbol{\pi}) \end{aligned} \tag{C.1}$$

To derive the EM algorithm, we first take expectations over the latent variable Z_i,

$$\begin{aligned}\mathbb{E}\left\{\mathcal{L}_{\text{comp}}\right\} = &\sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=0}^{L} \mathbf{1}\left(Y_{ij} = \ell\right) \mathbb{E}\left\{\mathbf{1}\left(Z_i = k\right)\right\} \underbrace{\log \Pr(Y_{ij} = \ell | Z_i = k, \boldsymbol{\mu})}_{\equiv \log \boldsymbol{\mu}_{kj\ell}} \\ &+ \sum_{i=1}^{N} \sum_{k=1}^{K} \mathbb{E}\left\{\mathbf{1}\left(Z_i = k\right)\right\} \underbrace{\log \Pr(Z_i = k | \boldsymbol{\pi})}_{= \log \boldsymbol{\pi}_k} \end{aligned} \tag{C.2}$$

We represent this unknown quantity as

$$\zeta_{ik} \equiv \mathbb{E}\left\{\mathbf{1}\left(Z_i = k\right)\right\}.$$

Then the E-step can be the normalized version of the posterior probability marginalized by the mixing proportion,

$$\widehat{\zeta}_{ik} \propto \pi_k \prod_{j=1}^{D} \underbrace{\prod_{\ell=0}^{L} \left(\mu_{kj\ell}\right)^{\mathbf{1}(Y_{ij}=\ell)}}_{\equiv \boldsymbol{\mu}_{kj,Y_{ij}}} \tag{C.3}$$

E-step For each voter i, compute the probability that they belong in cluster k:

$$\zeta_{ik} \leftarrow \frac{\boldsymbol{\pi}_k \prod_{j=1}^{J} \boldsymbol{\mu}_{kj,Y_{ij}}}{\sum_{k'=1}^{K} \boldsymbol{\pi}_{k'} \prod_{j=1}^{J} \boldsymbol{\mu}_{k'j,Y_{ij}}} \tag{C.4}$$

The M-step is derived by taking the derivatives of $\mathbb{E}\left\{\mathcal{L}_{\text{comp}}\right\}$ with respect to the model parameters $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$. This leads to a MLE-like M-step:

M-step Take the MLE, as derived in section C.2.4. For updating π_k, take the simple average of $\widehat{\zeta}_{ik}$ across all i. For updating $\widehat{\mu}_{kj\ell}$, take for each k and ℓ the sample proportion of the occurrence of $Y_{ij} = \ell$, but weighted by $\widehat{\zeta}_{ik}$:

$$\text{for each } k, \text{ update:} \quad \widehat{\pi}_k \leftarrow \frac{1}{N} \sum_{i=1}^{N} \widehat{\zeta}_{ik} \tag{C.5}$$

$$\text{for each } k, j, \ell, \text{ update:} \quad \widehat{\mu}_{kj\ell} \leftarrow \frac{\sum_{i=1}^{N} \mathbf{1}\left(Y_{ij} = \ell\right) \widehat{\zeta}_{ik}}{\sum_{i=1}^{N} \widehat{\zeta}_{ik}}, \tag{C.6}$$

We iterate through these two steps until convergence.

C.2.2 Evaluating Convergence

We evaluate convergence by the observed log likelihood,

$$\mathbf{L}_{\text{obs}} = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \boldsymbol{\mu}_{kj,Y_{ij}}$$

So the observed log-likelihood is

$$\mathcal{L}_{\text{obs}} = \sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \boldsymbol{\mu}_{kj,Y_{ij}} \right\} = \sum_{i=1}^{N} \log \left\{ \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \prod_{\ell=0}^{L} (\mu_{kj\ell})^{\mathbf{1}(Y_{ij}=\ell)} \right\} \tag{C.7}$$

C.2.3 Speed-Up by Collapsing to Unique Profiles

Because this EM algorithm deals with discrete data, the algorithm needs only sufficient statistics. In our setting the unique number of voting profiles is much smaller than the number of observations, because vote vectors follow a systematic pattern and most votes are straight-ticket votes. Therefore, we can re-format the dataset so that each row is a unique combination.

Let $u \in \{1, ..., U\}$ index the unique voting profiles, and n_u be the number of such profiles in the data. We re-cycle the objects $\mathbf{Y}$ and $\boldsymbol{\zeta}$ so that each row indexes profiles rather than voters.

We repeat the EM algorithm described earlier. For each profile u, compute the probability that it belong in type k:

$$\text{for each } u, k, \text{ update:} \quad \widehat{\zeta}_{uk} \leftarrow \frac{\boldsymbol{\pi}_k \prod_{j=1}^{J} \boldsymbol{\mu}_{kj,Y_{uj}}}{\sum_{k'=1}^{K} \boldsymbol{\pi}_{k'} \prod_{j=1}^{J} \boldsymbol{\mu}_{k'j,Y_{uj}}} \tag{C.8}$$

Then given those type probabilities, update with

$$\text{for each } k \text{, update:} \quad \widehat{\pi}_k \leftarrow \frac{1}{N} \sum_{u=1}^{U} n_u \widehat{\zeta}_{uk} \tag{C.9}$$

$$\text{for each } k, j, \ell \text{, update:} \quad \widehat{\mu}_{kj\ell} \leftarrow \frac{\sum_{u=1}^{U} n_u \mathbf{1}\left(Y_{uj} = \ell\right) \widehat{\zeta}_{uk}}{\sum_{u=1}^{U} n_u \widehat{\zeta}_{uk}} \tag{C.10}$$

$$\tag{C.11}$$

And the observed log-likelihood will also only require looping through the profiles:

$$\mathcal{L}_{\text{obs}} = \sum_{u=1}^{U} \log n_u + \sum_{u=1}^{U} \log \left\{ \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \boldsymbol{\mu}_{kj,Y_{ij}} \right\} \tag{C.12}$$

C.2.4 Derivation of M-step

Recall that the expectation of the likelihood from equation C.2 is

$$\mathbb{E}\left\{\mathcal{L}_{\text{comp}}\right\} = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{\ell=0}^{L} \mathbf{1}\left(Y_{ij} = \ell\right) \zeta_{ik} \log \mu_{kj\ell} + \sum_{i=1}^{N} \sum_{k=1}^{K} \zeta_{ik} \log \pi_k$$

so to optimize we introduce Langrange multipliers λ and $\boldsymbol{\eta}$ for the constraints on $\boldsymbol{\pi}$ and $\boldsymbol{\mu}_{kj}$, respectively:

$$\widetilde{\mathcal{L}} = \mathbb{E}\left\{\mathcal{L}_{\text{comp}}\right\} - \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) - \sum_{k=1}^{K} \sum_{j=1}^{J} \eta_{kj} \left(\sum_{\ell=0}^{L} \mu_{kj\ell} - 1 \right) \tag{C.13}$$

Then, for $\boldsymbol{\pi}$ we have that

$$\frac{\partial}{\partial \pi_k} \widetilde{\mathcal{L}} = \frac{\sum_{i=1}^{N} \zeta_{ik}}{\pi_k} - \lambda = 0$$

along with the constraint $\sum_{k=1}^{K} \pi_k = 1$. Notice that when we sum the FOC for $\boldsymbol{\pi}$ across k, the first condition becomes $\sum_{k=1}^{K} \pi_k = \frac{1}{\lambda} \sum_{k=1}^{K} \sum_{i=i}^{N} \zeta_{ik}$, and because the LHS sums to 1 due to the constraint and in the RHS $\sum_{i=1}^{N} \sum_{k'=1}^{K} \zeta_{ik'}$ sums to N, we have $\lambda = N$.

Separately, for $\boldsymbol{\mu}_{kj}$ we have that

$$\frac{\partial}{\partial \mu_{kj\ell}} \widetilde{\mathcal{L}} = \frac{\sum_{i=1}^{N} \mathbf{1}\left(Y_{ij} = \ell\right) \zeta_{ik}}{\boldsymbol{\mu}_{kj\ell}} - \eta_{kj} = 0,$$

along with constraint $\sum_{\ell=0}^{L} \mu_{kj\ell} = 1$. Once we sum the FOC for $\boldsymbol{\mu}$ across ℓ the first condition becomes $\sum_{\ell=0}^{L} \mu_{kj\ell} = \frac{1}{\eta_{kj}} \sum_{i=1} \sum_{\ell=0} \mathbf{1}\left(Y_{ij} = \ell\right) \zeta_{ik}$, and because the LHS again sums to 1 and in the RHS $\sum_{i=1}^{N} \sum_{\ell=0}^{L} \mathbf{1}\left(Y_{ij} = \ell\right) \zeta_{ik}$ sums to the prevalence of the weights $\sum_{i=1}^{N} \zeta_{ik}$, we get $\eta_{kj} = \sum_{i=1}^{N} \zeta_{ik}$.

Together, the above imply that

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} \zeta_{ik} \quad \text{and} \quad \mu_{kj\ell} = \frac{\sum_{i=1}^{N} \mathbf{1}\{Y_{ij} = \ell\} \zeta_{ik}}{\sum_{i=1}^{N} \zeta_{ik}} \tag{C.14}$$

C.2.5 Estimation with Varying Choice Sets

Because $\exp(\psi_{kj\ell}) = 1$ for $\ell = 0$, which exists in all three components, each component is analogous to a simple multinomial logit. In the first two cases, since we consider only two possibilities, it reduces to a simple intercept-only logit. Also notice that we use the same set of parameters $\boldsymbol{\psi}_{kj}$ regardless of M_{ij}. This represents the well-known independence of irrelevant alternatives (IIA) assumption in multinomial logit. The choice probabilities when one option is not on the "menu" is assumed to follow the same type of decision rule as the ratio between the existing options.

We use this new representation of the parameter μ in the EM algorithm, replacing the weighted average M-step for μ with a weighted multinomial logit:

$$\text{for each } k, \text{ update:} \quad \widehat{\pi}_k \leftarrow \frac{1}{N}\sum_{i=1}^{N}\widehat{\zeta}_{ik} \tag{C.15}$$

$$\text{for each } k, j, \ell, \text{ update:} \quad \widehat{\mu}_{kj\ell} \leftarrow \frac{\exp\left(\widehat{\psi}_{kj\ell}\right)}{1+\exp\left(\widehat{\psi}_{kj1}\right)+\exp\left(\widehat{\psi}_{kj2}\right)}, \tag{C.16}$$

where the $\boldsymbol{\psi}_{kj}$ vector is estimated from the coefficients of a multinomial logit, of the form

```
mlogit(Y[[j]] ~ 1, data, weights = zeta_k).
```

In other words, for each k, j, we estimate intercepts from regressing a vector of categorical votes for office $\mathbf{Y}_j$, using the estimates of $\boldsymbol{\zeta}_k$ as the weight `zeta_k`. R packages of multinomial logit typically presume IIA if an outcome value is missing and implicitly do the kind of three-way subsetting as in equation 5.5.

We can also solve the mlogit with varying choice sets by coding the MLE directly. In this paper, I opt for this option because it is considerably faster than using a built-in multinomial package.

To formalize this, I introduce new notation $m_{ij\ell} \in \{0, 1\}$, for whether option ℓ is available for individual i in office j. Clearly, therefore, $m_{ij\ell}$ is a direct a mapping from M_{ij}.

$$
\begin{array}{rc}
 & \begin{array}{ccc} \ell=0 & \ell=1 & \ell=2 \end{array} \\
m_{ij\ell} = \begin{array}{r} \text{if } M_{ij}=1 \\ \text{if } M_{ij}=2 \\ \text{if } M_{ij}=3 \end{array} & \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}
\end{array}
$$

The log likelihood for the parameter of interest $\boldsymbol{\psi}_{jk} = \{\psi_{jk0}, \psi_{jk1}, ..., \psi_{jkL}\}$ is, for a fixed office j when considering the kth cluster:

$$
\mathcal{L}(\boldsymbol{\psi}_{jk}) = \sum_{i=1}^{n} \sum_{\ell=0}^{L} m_{ij\ell}\zeta_{ik}\mathbf{1}\left(Y_{ij}=\ell\right) \log\left(\frac{\exp(\psi_{jk\ell})}{\sum_{\ell'=0}^{L} m_{ij\ell'} \exp(\psi_{jk\ell'})}\right) \tag{C.17}
$$

Then we can solve the parameters numerically, i.e.,

$$
\widehat{\boldsymbol{\psi}}_{jk}^{\text{MLE}} = \arg_{\boldsymbol{\psi}} \max \mathcal{L}(\boldsymbol{\psi}_{jk}) \tag{C.18}
$$

by software like `optim`.

C.2.6 The gradient for varying multinomial logit

The optimization program to find the MLE numerically will converge faster if we provide a gradient function. To derive this take the partial derivative of the log likelihood, which returns a length-$(L+1)$ vector $\nabla\mathcal{L}(\boldsymbol{\psi}_{jk})$ where the $(\ell+1)$th element is derived in the following subsection. All of significantly reduces time by reducing the overhead introduced in off-the-shelf packages like `mlogit`.

Our goal is to take the partial derivative of the likelihood in eq. C.17 with respect to ψ_{jk1} and ψ_{jk2}:

$$\mathcal{L}(\boldsymbol{\psi}_{jk}) = \sum_{i=1}^{n} \sum_{\ell=0}^{L} m_{ij\ell} \zeta_{ik} \mathbf{1}\left(Y_{ij} = \ell\right) \log \left(\frac{\exp(\psi_{jk\ell})}{\sum_{\ell'=0}^{L} m_{ij\ell'} \exp(\psi_{jk\ell'})} \right) \tag{C.19}$$

It is easier to consider the gradient at i, because the rest will be the sum of the individual gradients.

$$\mathcal{L}(\boldsymbol{\psi}_{jk})_i = \zeta_{ik} \sum_{\ell=0}^{L} \left\{ m_{i\ell} \mathbf{1}\left(Y_{ij} = \ell\right) \log \left(\frac{\exp(\psi_{jk\ell})}{\sum_{\ell'=0}^{L} \exp_{\psi_{jk\ell'}}} \right) \right\}$$

$$\text{Let } c_i = \sum_{\ell'=0}^{L} \exp_{\psi_{jk\ell'}} \text{ to abbreviate}$$

$$\begin{aligned}
\nabla \mathcal{L}(\psi_{jk1})_i =& \zeta_{ik} \Bigg\{ m_{i0} \mathbf{1}\left(Y_{ij} = 1\right) \frac{\partial}{\partial \psi_{jk1}} \log \left(\frac{1}{c_i} \right) + \\
& m_{i1} \mathbf{1}\left(Y_{ij} = 1\right) \frac{\partial}{\partial \psi_{jk1}} \log \left(\frac{\exp \psi_{jk1}}{c_i} \right) + m_{i2} \mathbf{1}\left(Y_{ij} = 1\right) \frac{\partial}{\partial \psi_{jk1}} \log \left(\frac{\exp \psi_{jk2}}{c_i} \right) \Bigg\} \\
=& \zeta_{ik} \Bigg\{ m_{i0} \mathbf{1}\left(Y_{ij} = 1\right) \left(-c_i \left(\frac{1}{c_i} \right)^2 \exp \psi_{jk1} \right) + \\
& m_{i1} \mathbf{1}\left(Y_{ij} = 1\right) \left(1 - \frac{\exp(\psi_{jk1})}{c_i} \right) + m_{i2} \mathbf{1}\left(Y_{ij} = 1\right) \left(- \frac{\exp \psi_{jk1}}{c_i} \right) \Bigg\} \\
=& \zeta_{ik} \left\{ m_{i1} \mathbf{1}\left(Y_{ij} = 1\right) \left(1 - \frac{\exp \psi_{jk1}}{c_i} \right) + \sum_{\ell' \neq 1} m_{i\ell'} \mathbf{1}\left(Y_{ij} = 1\right) \left(\frac{\exp \psi_{jk\ell}}{c_i} \right) \right\}
\end{aligned}$$

So generally, the $\ell + 1$th gradient is

$$\nabla \mathcal{L}(\psi_{jk\ell}) = \sum_{i=1}^{n} \left[\zeta_{ik} \left\{ m_{i\ell} \mathbf{1}\left(Y_{ij} = \ell\right) \left(1 - \frac{\exp \psi_{jk\ell}}{c_i} \right) + \sum_{\ell' \neq \ell} m_{i\ell'} \mathbf{1}\left(Y_{ij} = \ell\right) \left(\frac{\exp \psi_{jk\ell}}{c_i} \right) \right\} \right]$$

C.2.7 Evaluating Convergence with missingness

When following the EM algorithm on this data affected by uncontested choices, the observed log likelihood changes. Recall that in the no-missing case, we have equation C.7. However, in cases of missingness, the contribution of a data point also depends on the contestedness class.

$$\mathcal{L}^{\star}_{\text{obs}} = \sum_{i=1}^{N} \log \left[\sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} \prod_{\ell \in S_{M_{ij}}} \left\{ \left(\frac{\mu_{kj\ell}}{\sum_{\ell' \in S_{M_{ij}}} \mu_{kj\ell'}} \right)^{\mathbf{1}(Y_{ij}=\ell)} \right\} \right] \tag{C.20}$$

Bibliography

Abramowitz, Alan I. and Kyle L. Saunders (1998). "Ideological Realignment in the U.S. Electorate". *The Journal of Politics* 60 (3), pp. 634–652. DOI: 10.2307/2647642.

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, pp. 12–22. DOI: 10.1016/j.electstud.2015.11.001.

Ahlquist, John S. and Christian Breunig (2012). "Model-based clustering and typologies in the social sciences". *Political Analysis* 20 (1), pp. 92–112. DOI: 10.1093/pan/mpr039.

Aldrich, John H. (2000). "Presidential Address: Southern Parties in State and Nation". *Journal of Politics* 62 (3), pp. 643–670. DOI: 10.1111/0022-3816.00028.

Alvarez, R. Michael, Thad E. Hall, and Ines Levin (2018). "Low-Information Voting: Evidence From Instant-Runoff Elections". *American Politics Research*. DOI: 10.1177/1532673X18759643.

American Political Science Association (1950). "Toward a More Responsible Two-Party System: A Report of the Committee on Political Parties". *American Political Science Review* 44 (3), pp. 1–14. DOI: 10.2307/1950998.

Ansolabehere, Stephen. and Eitan Hersh (2012). "Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate". *Political Analysis* 20 (4), pp. 437–459. DOI: 10.1093/pan/mps023.

Ansolabehere, Stephen and Douglas Rivers (2013). "Cooperative Survey Research". *Annual Review of Political Science* 16 (1), pp. 307–329. DOI: 10.1146/annurev-polisci-022811-160625.

Ansolabehere, Stephen, Brian Schaffner, and Sam Luks (2017). "Guide to the 2016 Cooperative Congressional Election Survey". *Harvard Dataverse*. DOI: 10.7910/DVN/GDF6Z0.

Ansolabehere, Stephen and James M. Snyder (2000). "Valence Politics and Equilibrium in Spatial Election Models". *Public Choice* 103 (3), pp. 327–336. DOI: 10.1023/A:1005020606153.

Ansolabehere, Stephen and James M. Snyder (2002). "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942 - 2000". *Election Law Journal* 1 (3), pp. 315–338. DOI: 10.1089/153312902760137578.

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart (2001). "Candidate Positioning in U.S. House Elections". *American Journal of Political Science* 45 (1), pp. 136–159. DOI: 10.2307/2669364.

Arceneaux, Kevin (2006). "The Federal Face of Voting: Are Elected Officials Held Accountable for the Functions Relevant to their Office?" *Political Psychology* 27 (5), pp. 731–754. DOI: 10.1111/j.1467-9221.2006.00530.x.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg (2019). "Incumbency and Information". URL: https://perma.cc/FD7A-9V24.

Bafumi, Joseph, Michael C. Herron, Seth J. Hill, and Jeffrey B. Lewis (2012). "Alvin Greene? Who? How Did He Win the United States Senate Nomination in South Carolina?" *Election Law Journal* 11 (4), pp. 358–379. DOI: 10.1089/elj.2011.0137.

Bailey, Michael A. and Forrest Maltzman (2008). "Does Legal Doctrine Matter? Unpacking Law and Policy Preferences on the U.S. Supreme Court". *American Political Science Review* 102 (3), pp. 369–384. DOI: 10.1017/S0003055408080283.

Beck, Paul Allen, Lawrence Baum, Aage R. Clausen, and Charles E. Smith (1992). "Patterns and Sources of Ticket Splitting in Subpresidential Voting". *American Political Science Review* 86 (4), pp. 916–928. DOI: 10.2307/1964344.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein (2020). "Varying impacts of letters of recommendation on college admissions: Approximate balancing weights for subgroup effects in observational studies". *arXiv preprint arXiv:2008.04394*.

de Benedictis-Kessner, Justin (2017). "Off-Cycle and Out of Office: Election Timing and the Incumbency Advantage". *Journal of Politics* 80 (1), pp. 119–132. DOI: 10.1086/694396.

de Benedictis-Kessner, Justin and Christopher Warshaw (2020). "Politics in Forgotten Governments: The Partisan Composition of County Legislatures and County Fiscal Policies". *Journal of Politics* 82 (2). DOI: 10.1086/706458.

Besley, Timothy and Stephen Coate (2008). "Issue Unbundling via Citizens' Initiatives". *Quarterly Journal of Political Science* 3 (4), pp. 379–397. DOI: 10.1561/100.00008059.

Bisbee, James (2019). "BARP: Improving Mister P Using Bayesian Additive Regression Trees". *American Political Science Review* 113 (4), pp. 1060–1065.

Broockman, David E. (2016). "Approaches to Studying Policy Representation". *Legislative Studies Quarterly* 41 (1), pp. 181–215. DOI: 10.1111/lsq.12110.

Broockman, David E and Christopher Skovron (2018). "Bias in perceptions of public opinion among American political elites". *American Political Science Review* 112 (3), pp. 542–563.

Bucchianeri, Peter (2020). "Party Competition and Coalitional Stability: Evidence from American Local Government". *American Political Science Review*, pp. 1–16. DOI: 10.1017/S0003055420000386.

Buell, Duncan A., Eleanor Hare, Frank Heindel, Chip Moore, and Barbara Zia (2011). "Auditing a DRE-Based Election in South Carolina". *Proceedings of the 2011 conference on Electronic voting technology / workshop on trustworthy elections*.

Burden, Barry C. and Gretchen Helmke (2009). "The Comparative Study of Split-Ticket Voting". *Electoral Studies* 28 (1), pp. 1–7. DOI: 10.1016/j.electstud.2008.06.004.

Burden, Barry C. and David C. Kimball (1998). "A New Approach to the Study of Ticket Splitting". *American Political Science Review* 92 (3), pp. 533–544.

Burden, Barry C. and David C. Kimball (2002). *Why Americans Split Their Tickets: Campaigns, Competition, and Divided Government*. University of Michigan Press.

Bürkner, Paul-Christian (2021). *brms: Bayesian Regression Models using 'Stan'*. URL: https://cran.r-project.org/package=brms.

Buttice, Matthew K and Benjamin Highton (2013). "How does multilevel regression and poststratification perform with conventional national surveys?" *Political Analysis*, pp. 449–467. DOI: 10.1093/pan/mpt017.

Campbell, Angus (1960). "Surge and Decline: A Study of Electoral Change". *Public Opinion Quarterly* 24, pp. 397–418.

Campbell, Angus and Warren E. Miller (1957). "The Motivational Basis of Straight and Split Ticket Voting". *The American Political Science Review* 13 (3), pp. 400–414. DOI: 10.2307/1952193.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan (2002). "Out of Step, out of Office: Electoral Accountability and House Members' Voting Author(s):" *American Political Science Review* 96 (1), pp. 127–140. DOI: 10.1017/S0003055402004276.

Carnes, Nicholas and Noam Lupu (2020). “The White Working Class and the 2016 Election”. *Perspectives on Politics*. DOI: 10.1017/S1537592720001267.

Carroll, Susan J. (1999). “The Disempowerment of the Gender Gap: Soccer Moms and the 1996 Elections”. *PS: Political Science and Politics* 32 (1), p. 7. DOI: 10.2307/420743.

Carson, Jamie L., Joel Sievert, and Ryan D. Williamson (2019). “Nationalization and the Incumbency Advantage”. *Political Research Quarterly*. DOI: 10.1177/1065912919883696.

Caughey, Devin et al. (2020). *Target Estimation and Adjustment Weighting for Survey Nonresponse and Sampling Bias*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.

Chattopadhyay, Ambarish, Christopher H Hase, and José R Zubizarreta (2020). “Balancing vs modeling approaches to weighting in practice”. *Statistics in Medicine* 39 (24), pp. 3227–3254.

Cohn, Nate (2018). “Big Republican Advantages Are Eroding in the Race for House Control.” *The New York Times* (Feb. 12, 2018). URL: https://perma.cc/7VWJ-3Z9B.

Cohn, Nate (2019a). *A Sliver of the Electorate Could Decide 2020. Here’s What These Voters Want.* Nov. 5, 2019. URL: https://nyti.ms/2NEC2sY.

Cohn, Nate (2019b). “One Year From Election, Trump Trails Biden but Leads Warren in Battlegrounds.” *The New York Times* (Nov. 4, 2019). URL: https://perma.cc/2HE5-8YDA.

Coppock, Alexander, Seth J. Hill, and Lynn Vavreck (2020). “The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments”. *Science advances* 6 (36), eabc4046.

Cox, Gary W and Jonathan N. Katz (1996). “Why Did the Incumbency Advantage in U. S. House Elections Grow?” *American Journal of Political Science* 40 (2), pp. 478–497.

Daily Kos (2019). “The Daily Kos Elections naming guide to the nation’s congressional districts”. URL: https://perma.cc/P5C9-DRN8.

Daily Kos (2021). “Calculating Statewide Election Results by Congressional and Legislative District: Statement of Methodology”. URL: https://docs.google.com/document/d/1efGEHGAQGaDGnIrdedwxVg1uUrjZYz06IPLTHRCg9zM.

Darr, Joshua P., Matthew P. Hitt, and Johanna L. Dunaway (2018). "Newspaper Closures Polarize Voting Behavior". *Journal of Communication* 68, pp. 1007–1028. DOI: 10.1093/joc/jqy051.

Davis, Nicholas T. and Lilliana Mason (2016). "Sorting and the Split-Ticket: Evidence from Presidential and Subpresidential Elections". *Political Behavior* 38 (2), pp. 337–354. DOI: 10.1007/s11109-015-9315-7.

Dehejia, Rajeev H. and Sadek Wahba (1999). "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs". *Journal of the American statistical Association* 94 (448), pp. 1053–1062. DOI: 10.1080/01621459.1999.10473858.

Deming, W. Edwards and Frederick F. Stephan (1940). "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known". *The Annals of Mathematical Statistics* 11 (4), pp. 427–444.

Drutman, Lee (2018). "America has local political institutions but nationalized politics. This is a problem." *FiveThirtyEight* (May 31, 2018). URL: https://perma.cc/Z4X5-J8K9.

Drutman, Lee (2020). *Breaking the Two-Party Doom Loop: The Case for Multiparty Democracy in America*. Oxford University Press, USA.

Dubin, Jeffrey A. and Elisabeth R. Gerber (1992). "Patterns of Voting on Ballot Propositions: A Mixture Model of Voter Types".

Eggers, Andrew C. (2017). "Quality-Based Explanations of Incumbency Effects". *Journal of Politics* 79 (4), pp. 1315–1328. DOI: 10.1086/692667.

Eggers, Andrew C., Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder (2015). "On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races". *American Journal of Political Science* 59 (1), pp. 259–274. DOI: 10.1111/ajps.12127.

Enns, Peter K. and Patrick C. Wohlfarth (2013). "The Swing Justice". *Journal of Politics* 75 (4), pp. 1089–1107. DOI: 10.1017/S0022381613001035.

Enten, Harry (2016). "There Were No Purple States On Tuesday". *FiveThirtyEight* (Nov. 10, 2016). URL: https://perma.cc/VD6J-XPXC.

Erikson, Robert S., Gerald C. Wright, and John P. McIver (1993). *Statehouse democracy: Public opinion and policy in the American states*. Cambridge University Press.

Estevez-Abe, Margarita, Torben Iversen, and David Soskice (2010). "Social Protection and the Formation of Skills-A Reinterpretation of the Welfare State". *95th American Political Association Meeting*, pp. 1–80. DOI: 10.1093/0199247757.001.0001.

Farris, Emily M. and Mirya R. Holman (2017). "All Politics Is Local? County Sheriffs and Localized Policies of Immigration Enforcement". *Political Research Quarterly* 70 (1), pp. 142–154. DOI: 10.1177/1065912916680035.

Fiorina, Morris (1992). *Divided Government*. Pearson.

Fiorina, Morris P (2017). *Unstable Majorities: Polarization, Party Sorting, and Political Stalemate*. Hoover Press.

Fiorina, Morris P. and Samuel J. Abrams (2008). "Political Polarization in the American Public". *Annual Review of Political Science* 11, pp. 563–588. DOI: 10.1146/annurev.polisci.11.053106.153836.

Fiske, Susan T., Amy J. C. Cuddy, Peter Glick, and Jun Xu (2002). "A Model of (Often Mixed) Stereotype Content: Competence and Warmth Respectively Follow From Perceived Status and Competition". *Journal of Personality and Social Psychology* 82 (6), pp. 878–902. DOI: 10.1037/0022-3514.82.6.878.

Fraga, Bernard L. and Eitan D. Hersh (2018). "Are Americans Stuck in Uncompetitive Enclaves? An Appraisal of U.S. Electoral Competition". *Quarterly Journal of Political Science* 13 (3), pp. 617–627. DOI: 10.1561/100.00017161.

Fraley, Chirs and Adrian E. Raftery (1998). "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis". *The Computer Journal* 41 (8), pp. 578–588. DOI: 10.1093/comjnl/41.8.578.

Gelman, Andrew (2007). "Struggles with Survey Weighting and Regression Modeling". *Statistical Science* 22 (2), pp. 184–188. DOI: 10.1214/088342307000000203.

Gelman, Andrew and Thomas C. Little (1997). "Poststratification into many categories using hierarchical logistic regression". *Survey Methodology* 23 (2), pp. 127–135.

Gerber, Elisabeth R. and Jeffrey B. Lewis (2004). "Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation". *Journal of Political Economy* 112 (6), pp. 1364–1383. DOI: 10.1086/424737.

Ghitza, Yair (2019). *Revisiting What Happened in the 2018 Election: An Analysis of the Catalist Voter Registration Database*. Tech. rep. URL: https://perma.cc/G7A9-DPMT.

Ghitza, Yair and Andrew Gelman (2013). "Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups". *American Journal of Political Science* 57 (3), pp. 762–776. DOI: 10.1111/ajps.12004.

Ghitza, Yair and Mark Steitz (2020). "DEEP-MAPS Model of the Labor Force". *Working Paper*.

Goplerud, Max (2019). "A Multinomial Framework for Ideal Point Estimation". *Political Analysis* 27 (1), pp. 69–89. DOI: 10.1017/pan.2018.31.

Goplerud, Max, Shiro Kuriwaki, Marc Ratkovic, and Dustin Tingley (2018). "Sparse Multilevel Regression (and Poststratification (sMRP))". *Unpublished manuscript, Harvard University.*

Greiner, James D. and Kevin M. Quinn (2009). "R x C Ecological Inference: Bounds, Correlations, Flexibility and Transparency of Assumptions". *Journal of the Royal Statisitical Society Series A* 172 (1), pp. 67–81.

Grimmer, Justin and Gary King (2011). "General purpose computer-assisted clustering and conceptualization". *Proceedings of the National Academy of Sciences* 108 (7), pp. 2643–2650. DOI: 10.1073/pnas.1018067108.

Grimmer, Justin and William Marble (2019). "Who Put Trump in the White House? Explaining the Contribution of Voting Blocs to Trump's Victory", Working Paper. URL: https://williammarble.co/docs/vb.pdf.

Groseclose, Tim (2007). "A Model of Candidate Location When One Candidate Has a Valence Advantage". *American Journal of Political Science* 45 (4), p. 862. DOI: 10.2307/2669329.

Hainmueller, Jens (2012). "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies". *Political Analysis*, pp. 25–46. DOI: 10.1093/pan/mpr025.

Hall, Andrew B. (2015). "What Happens When Extremists Win Primaries?" *American Political Science Review* 109 (1), pp. 18–42. DOI: 10.1017/S0003055414000641.

Hall, Andrew B. and Daniel M. Thompson (2018). "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections". *American Political Science Review* 112 (03), pp. 509–524. DOI: 10.1017/s0003055418000023.

Hanretty, Chris, Benjamin E. Lauderdale, and Nick Vivyan (2016). "Comparing Strategies for Estimating Constituency Opinion from National Survey Samples". *Political Science Research and Methods* 6 (3), pp. 1–21. DOI: 10.1017/psrm.2015.79.

Hansen, Jeremy A. (2015). "Ballot-Level Observations about Vermont's 2014 General Election". *New England Journal of Political Science* 8 (2), pp. 186–217.

Herron, Michael C. and Jeffrey B. Lewis (2007). "Did Ralph Nader Spoil a Gore presidency? A Ballot-level Study of Green and Reform Party Voters in the 2000 Presidential Election". *Quarterly Journal of Political Science* 2 (3), pp. 205–226. DOI: 10.1561/100.00005039.

Hersh, Eitan D. (2015). *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press.

Hertel-Fernandez, Alexander, Matto Mildenberger, and Leah C. Stokes (2019). "Legislative staff and representation in Congress". *American Political Science Review* 113 (1), pp. 1–18.

Hill, Jennifer L. and Hanspeter Kriesi (2001). "Classification by Opinion-Changing Behavior: A Mixture Model Approach". *Political Analysis* 9 (4), pp. 301–324. DOI: 10.1093/oxfordjournals.pan.a004872.

Hill, Seth J., Daniel J. Hopkins, and Gregory A. Huber (2021). "Not by turnout alone: Measuring the sources of electoral change, 2012 to 2016". *Science Advances* 7 (17). DOI: 10.1126/sciadv.abe3272.

Hillygus, D. Sunshine and Todd G. Shields (2009). "The Persuadable Voter". *The Persuadable Voter*. DOI: 10.2307/j.ctt7rxq8.

Hirano, Shigeo and James M. Snyder (2019). *Primary Elections in the United States*. Cambridge University Press. DOI: 10.1017/9781139946537.

Hopkins, Daniel J. (2018). *The Increasingly United States: How and Why American Political Behavior Nationalized*. University of Chicago Press.

Imai, Kosuke and In Song Kim (2019). "When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" *American Journal of Political Science* 63 (2), pp. 467–490. DOI: 10.1111/ajps.12417.

Imai, Kosuke and In Song Kim (2020). "On the Use of Two-way Fixed Effects Regression Models for Causal Inference".

Imai, Kosuke, James Lo, and Jonathan Olmsted (2016). "Fast Estimation of Ideal Points with Massive Data". *American Political Science Review* 110 (4), pp. 631–656. DOI: 10.1017/S000305541600037X.

Imai, Kosuke and Marc Ratkovic (2014). "Covariate balancing propensity score". *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pp. 243–263. DOI: 10.1111/rssb.12027.

Imai, Kosuke and Aaron Strauss (2011). "Estimation of heterogeneous treatment effects from randomized experiments, with application to the optimal planning of the get-out-the-vote campaign". *Political Analysis* 19, pp. 1–19. DOI: 10.1093/pan/mpq035.

Imai, Kosuke and Dustin Tingley (2012). "A statistical method for empirical testing of competing theories". *American Journal of Political Science* 56 (1), pp. 218–236. DOI: 10.1111/j.1540-5907.2011.00555.x.

Jacobson, Gary C. (2015). "It's Nothing Personal: The Decline of the Incumbency Advantage in US House Elections". *Journal of Politics* 77 (3), pp. 861–873. DOI: 10.1086/681670.

Jacobson, Gary C. (2019). "Extreme Referendum: Donald Trump and the 2018 Midterm Elections". *Political Science Quarterly* 134 (1), pp. 9–38. DOI: 10.1002/polq.12866.

Jensen, Amalie, William Marble, Kenneth Scheve, and Matthew J. Slaughter (2019). "City Limits to Partisan Polarizationin the American Public". URL: https://perma.cc/7CDV-9Z8H.

Kalla, Joshua L. and Ethan Porter (2020). "Correcting Bias in Perceptions of Public Opinion Among American Elected Officials: Results from Two Field Experiments". *British Journal of Political Science*, pp. 1–9. DOI: 10.1017/S0007123419000711.

Kastellec, Jonathan P., Jeffrey R. Lax, Michael Malecki, and Justin H. Phillips (2015). "Polarizing the electoral connection: Partisan representation in Supreme Court confirmation politics". *The Journal of Politics* 77 (3), pp. 787–804. DOI: https://doi.org/10.1086/681261.

Key, V. O. (1948). *Southern Politics in State and Nation*. University of Tennessee Press.

Key, V. O. (1963). *American State Politics: An Introduction*. Knopf.

Key, V. O. (1966). *The Responsible Electorate: Rationality in Presidential Voting: 1936-1960*. Harvard University Press.

King, Gary (1997). *A Solution to the Ecological Inference Problem*. Princeton University Press.

King, Gary and Langche Zeng (2001). "Logistic Regression in Rare Events Data". *Political Analysis* 55 (3), pp. 693–715.

Klar, Samara and Yanna Krupnikov (2016). *Independent Politics: How American Disdain for Parties Leads to Political Inaction*. Cambridge University Press.

Klarner, Carl (2021). *State Legislative Election Returns, 1967-2018, Version 120*. DOI: 10.7910/DVN/DRSACA. URL: http://klarnerpolitics.org/.

Künzel, Sören R, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu (2019). "Metalearners for estimating heterogeneous treatment effects using machine learning". *Proceedings of the national academy of sciences* 116 (10), pp. 4156–4165.

Kuriwaki, Shiro (2020). "The Administration of Cast Vote Records in U.S. States". URL: https://bit.ly/3dxc0TD.

Kuriwaki, Shiro (2021a). *ccesMRPprep: Functions and Data to Prepare CCES data for MRP*. URL: https://github.com/kuriwaki/ccesMRPprep.

Kuriwaki, Shiro (2021b). *clusterCVR: A Clustering Algorithm for Cast Vote Records*. URL: https://github.com/kuriwaki/clusterCVR.

Kuriwaki, Shiro (2021c). *Replication Data for: "The Swing Voter Paradox: Electoral Politics in a Nationalized Era"*. DOI: 10.7910/DVN/GEA0CI.

Kuriwaki, Shiro and Soichiro Yamauchi (2021). "Synthetic Area Weighting for Measuring Public Opinion in Small Areas". arXiv: 2105.05829 [stat.ME].

Kustov, Alexander, Maikol Cerda, Akhil Rajan, Frances Rosenbluth, and Ian Shapiro (2021). "The Rise of Safe Seats and Party Indiscipline in the U.S. Congress". *American Political Association Meeting*.

Larcinese, Valentino, James M. Snyder, and Cecilia Testa (2013). "Testing models of distributive politics using exit polls to measure voters' preferences and partisanship". *British Journal of Political Science* 43 (4), pp. 1–31. DOI: 10.1017/S0007123412000245.

Lax, Jeffrey R. and Justin H. Phillips (2009). "How should we estimate public opinion in the states?" *American Journal of Political Science* 53 (1), pp. 107–121.

Lee, Frances (2016). *Insecure Majorities: Congress and the Perpetual Campaign*. University Chicago Press.

Leemann, Lucas and Fabio Wasserfallen (2017). "Extending the use and prediction precision of subnational public opinion estimation". *American Journal of Political Science* 61 (4), pp. 1003–1022.

Levitt, Steven D. and Catherine D. Wolfram (1997). "Decomposing the Sources of Incumbency Advantage in the U. S. House". *Legislative Studies Quarterly* 22 (1), p. 45. DOI: 10.2307/440290.

Lewis, Jeffrey B. (2001). "Estimating Voter Preference Distributions from Individual-Level Voting Data". *Political Analysis* 9 (3), pp. 275–297. DOI: 10.1093/polana/9.3.275.

Linderman, Scott W., Matthew J. Johnson, and Ryan P. Adams (2015). "Dependent multinomial models made easy: Stick breaking with the Pólya-gamma augmentation". arXiv: arXiv:1506.05843v1.

Linzer, Drew A (2012). "The Relationship between Seats and Votes in Multiparty Systems". *Political Analysis*, pp. 400–416.

Linzer, Drew A. and Jefrey B. Lewis (2011). "poLCA: An R Package for Polytomous Variable Latent Class Analysis". *Journal of Statistical Software* 42 (10), pp. 1–18. DOI: 10.4018/978-1-4666-2931-8.ch001.

Lublin, David (2007). *The Republican South: Democratization and Partisan Change*. Princeton University Press.

Mayer, William G. (2008). "What Exactly is a Swing voter? Definition and Measurement". Brookings Institution Press, pp. 1–31.

Mayhew, David R. (1974). "Congressional Elections : The Case of the Vanishing Marginals". *Polity* 6 (3), pp. 295–317.

Mayhew, David R. (1986). *Placing Parties in American Politics: Organization, Electoral Settings, and Government Activity in the Twentieth century*. Princeton University Press.

McCarthy, John, Neal McBurnett, Harvie Branscomb, Ron Rivest, and Philip Stark (2018). *Preserving Anonymity of Cast Vote Record*. Tech. rep., pp. 1–11. URL: http://bit.ly/2LcuCOQ.

McLachlan, Geoffrey J., Sharon X. Lee, and Suren I. Rathnayake (2019). "Finite Mixture Models". *Annual Review of Statistics and Its Application* 6 (1), pp. 355–378. DOI: 10.1146/annurev-statistics-031017-100325.

Meng, Xiao-Li (2018). "Statistical paradises and paradoxes in Big Data (I): Law of large populations, Big Data paradox, and 2016 US presidential election". *Annals of Applied Statistics*. DOI: 10.1214/18-AOAS1161SF.

Mickey, Robert (2015). *Paths Out of Dixie: The Democratization of Authoritarian Enclaves in America's Deep South, 1944-1972*. Princeton University Press.

Morse, Michael (2021). "The Future of Felon Disenfranchisement Reform: How Partisanship and Poverty Shape the Restoration of Voting Rights in Florida". *California Law Review* 109.

Moskowitz, Daniel J. (2020). "Local News, Information, and the Nationalization of U.S. Elections". *American Political Science Review* 115 (1), pp. 114 –129. DOI: 10.1017/S0003055420000829.

Müllner, Daniel (2013). "fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python". *Journal of Statistical Software* 53 (9), pp. 1–18. DOI: 10.18637/jss.v053.i09.

Oliver, J. Eric, Shang E. Ha, and Zachary Callen (2012). *Local Elections and the Politics of Small-Scale Democracy*. Princeton University Press.

Ornstein, Joseph T. (2020). “Stacked Regression and Poststratification”. *Political Analysis* 28 (2), pp. 293–301.

Park, Won-ho, Michael J. Hanmer, and Daniel R. Biggers (2014). “Ecological Inference under Unfavorable Conditions: Straight and Split-ticket Voting in Diverse Settings and Small Samples”. *Electoral Studies* 36, pp. 192–203. DOI: 10.1016/j.electstud.2014.08.006.

Persson, Torsten and Guido Tabellini (2000). *Political Economics: Explaining Economic Policy*. MIT Press.

Pesendorfer, Wolfgang and Timothy Feddersen (1996). “The Swing Voter’s Curse”. *American Economic Review* 86 (3), pp. 408–424.

Peterson, Erik (2017). “The Role of the Information Environment in Partisan Voting”. *Journal of Politics* 79 (4), pp. 1191–1204. DOI: 10.1086/692740.

Pew Research Center (2015). *A Deep Dive Into Party Affiliation*. URL: https://perma.cc/RHU8-928W.

Polborn, Mattias K. and Jr. Snyder James M. (2017). “Party Polarization in Legislatures with Office-Motivated Candidates*”. *The Quarterly Journal of Economics* 132 (3), pp. 1509–1550. DOI: 10.1093/qje/qjx012.

Prat, Andrea, Riccardo Puglisi, and James M. Snyder (2010). “Is Private Campaign Finance a Good Thing? Estimates of the Potential Informational Benefits”. *Quarterly Journal of Political Science* 5 (3), pp. 291–318. DOI: 10.1561/100.00008081.

Rakich, Nathaniel and Ryan Best (2020). “There Wasn’t That Much Split-Ticket Voting In 2020”. *FiveThirtyEight* (Dec. 2, 2020). URL: https://perma.cc/J3BD-YMQX.

Rentsch, Anthony, Brian F. Schaffner, and Justin H. Gross (2019). “The Elusive Likely Voter: Improving Electoral Predictions With More Informed Vote-Propensity Models”. *Public Opinion Quarterly* 83 (4), pp. 782–804. DOI: 10.1093/poq/nfz052.

Rivers, Douglas (2007). “Sampling for web surveys”. *Joint Statistical Meetings*, p. 4. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.468.9645.

Rogers, Steven (2015). “Strategic Challenger Entry in a Federal System: The Role of Economic and Political Conditions in State Legislative Competition”. *Legislative Studies Quarterly* 40 (4), pp. 539–570. DOI: 10.1111/lsq.12088.

Rogers, Steven (2016). “National Forces in State Legislative Elections”. *The ANNALS of the American Academy of Political and Social Science* 667 (1), pp. 207–225. DOI: 10.1177/0002716216662454.

Rogers, Steven (2017). "Electoral Accountability for State Legislative Roll Calls and Ideological Representation". *American Political Science Review* 111 (3), pp. 555–571. DOI: 10.1017/S0003055417000156.

Rogers, Steven (2021). *Accountability in American Legislatures*.

Schickler, Eric (2016). *Racial Realignment: The Transformation of American Liberalism, 1932–1965*. Princeton University Press.

Sewell, Daniel K., Yuguo Chen, William Bernhard, and Tracy Sulkin (2016). "Model-based longitudinal clustering with varying cluster assignments". *Statistica Sinica* 26 (1), pp. 205–233. DOI: 10.5705/ss.2014.205.

Shiraito, Yuki (2016). "Uncovering Heterogeneous Treatment Effects".

Shor, Boris (2018). *Aggregate State Legislator Shor-McCarty Ideology Data, May 2018 update*. DOI: 10.7910/DVN/BSLEFD.

Shor, Boris and Nolan M. McCarty (2011). "The Ideological Mapping of American Legislatures". *American Political Science Review* 105 (August), pp. 530–551. DOI: 10.1017/S0003055411000153.

Si, Yajuan (2020). *On the Use of Auxiliary Variables in Multilevel Regression and Poststratification*. arXiv: 2011.00360 [stat.ME].

Sievert, Joel and Seth C. McKee (2018). "Nationalization in U.S. Senate and Gubernatorial Elections". *American Politics Research*, pp. 1–26. DOI: 10.1177/1532673X18792694.

Skelley, Geoffrey (2018). "Split-Ticket Voting Hit A New Low In 2018 Senate And Governor Races." *FiveThirtyEight* (Nov. 19, 2018). URL: https://perma.cc/5LKP-7A4S.

Smidt, Corwin D. (2017). "Polarization and the Decline of the American Floating Voter". *American Journal of Political Science* 61 (2), pp. 365–381. DOI: 10.1111/ajps.12218.

Stark, Philip B. (2008). "Conservative Statistical Post-Election Audits". *Annals of Applied Statistics* 2 (2), pp. 550–581. DOI: 10.1214/08-AOAS161.

Stein, Jeff (2016). "Ticket splitting is dead. National parties are now everything." *Vox* (Nov. 17, 2016). URL: https://perma.cc/4M47-A69T.

TargetSmart (2021). *National 2020 Historic General Election Early and Absentee Vote Report*. URL: https://perma.cc/DMU7-HFEM.

Tausanovitch, Chris and Christopher Warshaw (2014). "Representation in Municipal Government". *American Political Science Review* 108 (03), pp. 605–641. DOI: 10.1017/S0003055414000318.

Thompson, Daniel M. (2019). "How Partisan Is Local Law Enforcement? Evidence from Sheriff Cooperation with Immigration Authorities". *American Political Science Review*, pp. 222–236. DOI: 10.1017/S0003055419000613.

Trounstine, Jessica (2011). "Evidence of a Local Incumbency Advantage". *Legislative Studies Quarterly* 36 (2), pp. 255–280. DOI: 10.1111/j.1939-9162.2011.00013.x.

Trounstine, Jessica (2018). "Political Schizophrenics? Factors Affecting Aggregate Partisan Choice at the Local Versus National Level". *American Politics Research* 46 (1), pp. 26–46. DOI: 10.1177/1532673X17733800.

Wakefield, Jon (2004). "Ecological Inference for 2 x 2 Tables". *Journal of the Royal Statistical Society Series A*, pp. 385–445. DOI: 10.1111/j.1467-985x.2004.02046_1.x.

Walker, Kyle and Matt Herman (2021). *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. URL: https://cran.r-project.org/web/packages/tidycensus/index.html.

Warshaw, Christopher (2019). "Local Elections and Representation in the United States". *Annual Review of Political Science* 22, pp. 461–479. DOI: 10.1146/annurev-polisci-050317-071108.

Warshaw, Christopher and Jonathan Rodden (2012). "How should we measure district-level public opinion on individual issues?" *Journal of Politics* 74 (1), pp. 203–219. DOI: 10.1017/S0022381611001204.

Wedel, Michel and Wagner A. Kamakura (2000). *Market Segmentation: Conceptual and Methodological Foundations*. 2nd ed. Kluwer Academic Publishers.

Yamamoto, Teppei (2014). "A Multinomial Response Model for Varying Choice Sets, with Application to Partially Contested Multiparty Elections". URL: https://perma.cc/8MGK-5TLH.

Yamauchi, Soichiro (2021). *emlogit: An ECM algorithm for the Multinomial Logit Model*. URL: https://github.com/soichiroy/emlogit.

Yamauchi, Soichiro and Shiro Kuriwaki (2021). *bmlogit: A Multinomial logit with balancing constraints*. R package version 0.0.3. URL: https://github.com/soichiroy/bmlogit.

Zoorob, Michael (2020). "There's (rarely) A New Sheriff in Town: Incumbency Advantage in Sheriff Elections". URL: https://perma.cc/7R6M-BKHL.

Zubizarreta, José R. (2015). “Stable weights that balance covariates for estimation with incomplete outcome data”. *Journal of the American Statistical Association* 110 (511), pp. 910–922.