DUKE UNIVERSITY

UNDERGRADUATE HONORS THESIS

Splitting Hairs or Splitting Regions: The Differential Democratic Impacts of Splitting ZIP Codes vs. Counties During Redistricting

---

Author:
Jacob Hervey

Faculty Advisors:
Dr. Patrick Bayer
Dr. Michelle Connolly

Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University.

April 2024

Duke University Department of Economics

Abstract

In light of the Supreme Court's holding in *Gill v. Whitford*, judicially-enforceable gerrymandering metrics must focus on democratic harms to individual citizens, instead of state-wide measures of proportionality. Previous literature has suggested that gerrymandering metrics should focus on the extent to which congressional districts split preexisting geographic boundaries (namely, ZIP codes and counties). This work compares the differential democratic harms caused by ZIP code versus county splitting during redistricting across two domains. First, we exploit the changes during the 2010 redistricting process to construct a difference-in-difference model that captures changes in voters' political knowledge as a function of their exposure to geographic splitting. Second, we predict district-level electoral outcomes from 2002-2018 based upon the extent of ZIP code and county splitting. Our results indicate that ZIP code and county splitting cause more significant democratic harms for different outcomes of interest. While county splitting has more negative consequences for constituents' political knowledge, ZIP code splitting is more detrimental with regards to voter turnout.

Keywords: Gerrymandering, Redistricting, Communities of Interest, Geographic Boundaries

JEL Codes: D72, K16, H11

Acknowledgements

First and foremost, I would like to thank Professor Michelle Connolly for all the time she has spent with me to iron out every detail of this thesis. She has been an instrumental mentor to me this past year, and she truly cares about our success in and out of the classroom. Additionally, I would like to thank Professor Patrick Bayer for his help as a faculty mentor on this thesis. Beyond the encouragement in the classroom, I am incredibly lucky to have such a supportive network of friends and family. I mean seriously—I have the best friends and family ever. Particularly, I would like to thank all my friends at Duke who offered their scarce time as soon-to-be graduates to help with my project: Alexia Jackson, Delaney Eisen, Dharshan Varia, Evan Chan, Heera Rajavel, Hunter Quigley, John Lee, Nelson Aknin, Nhu Bui, and Rebecca Edelman. My incredible friends are simultaneously the reason I have loved my time here at Duke, and why I am heartbroken to say goodbye.

Meanwhile, my family is the entire reason why I am where I am today. Thank you to my parents, Jan and Paul Hervey, who have sacrificed everything to bring me to where I am today. My mom even went as far as helping organize my data on spreadhseets for this thesis. As much as they tease me, my siblings Abi, Maddie, and Camryn are my built-in lifelong friends (and some of the biggest supporters of my gerrymandering obsession). I love them all to the moon and back and am amazed by all their unique passions. I cannot wait to see where life takes each of them.

And of course, I would like to thank all of the avid followers of my gerrymandering Instragm page: @jacobs_gerrys. Keep calm and gerry on.

Contents

List of Figures

List of Tables

1 Introduction

> The plaintiffs and their amici curiae promise us that the efficiency gap and similar measures of partisan asymmetry will allow the federal courts—armed with just "a pencil and paper or a hand calculator"—to finally solve the problem of partisan gerrymandering that has confounded the Court for decades. We need not doubt the plaintiffs' math. The difficulty for standing purposes is that these calculations are an average measure. They do not address the effect that a gerrymander has on the votes of particular citizens. Partisan-asymmetry metrics such as the efficiency gap measure something else entirely: the effect that a gerrymander has on the fortunes of political parties.
> — *Gill v. Whitford*, 585 U.S.U.S. ____ (2018)

For decades, the issue of partisan redistricting has vexed the Supreme Court. Prior to the Court's unanimous ruling in *Gill v. Whitford*, mathematicians and policymakers alike proposed formulaic calculations to test congressional maps' fairness. Many of these calculations quantify how well maps represent political parties at a statewide level (Bernstein and Duchin, 2017; Stephanopoulos and McGhee, 2015). In *Gill v. Whitford* and subsequent majority opinions, the Supreme Court repeatedly denied challenges using these calculations because they are "average measures" that do not explain how congressional maps affect the "votes of particular citizens" (*Gill v. Whitford*, 2018). The Court's holdings have made clear that proportional representation at a statewide level is not a constitutionally enshrined right; instead, gerrymandering metrics must pinpoint the specific democratic harms done to individual civilians (Warrington, 2018).

Given the Court's current doctrinal posture, scholars have offered several new proposals for quantifying the impact of redistricting at the individual level. After each decennial U.S. census, all 50 states begin the redistricting cycle to redraw the boundaries of all 435 congressional districts for the United States House of Representatives. States must account for population shifts over the last decade to ensure that every district has approximately the same population. The process of redistricting is ultimately a game of line drawing. In doing so, new congressional maps may split geographic communities among multiple congressional districts, whether intentionally or not. Contemporary statisticians have argued that this geographic splitting has the potential to cause serious democratic harms; heightened geographic splitting among multiple congressional districts breeds confusion among voters and reduce politicians' accountability to specific communities (Curiel and Steelman, 2018; Bowen, 2014; Bowen and Clark, 2014). In response, many states have im-posed vague guidelines that require new maps to preserve geographic communities to the fullest extent possible; however, these guidelines do not concretely define what constitutes a given geographic community (Arrington, 2010).

While there is certainly a broad consensus that maps should avoid splitting geographic communities, recent scholarship has proposed various definitions of such communities. In one camp, scholars advocate that individuals within the same ZIP code act as a natural community of interest that should not be split during the redistricting process (Curiel and Steelman, 2018). Alternatively, other researchers advocate that maps should prioritize preserving counties over all other geographic communities when redistricting (Bowen, 2014). On both sides, scholars argue that splitting geographic units between multiple congressional districts erodes representatives' accountability and civilian political involvement. This paper seeks to compare the impact of splitting different geographic units on the political process. Specifically, we catalogue the ways splitting ZIP codes and counties have differential impacts on individual constituents and their political behavior (namely, their election turnout, congressional approval ratings, knowledge of their representative, and overall political engagement).

To assess this research question, this paper traces constituent behavior along two primary levels of analysis. First, we look at changes in individual voters' sentiment and knowledge from 2010 to 2012, as captured by political surveys. This is a timeframe of interest given that all 50 states underwent redistricting between 2010 and 2012. We compare the changes in constituents' behavior between those whose counties were split and those whose ZIP codes were split. Harvard University's "Cooperative Congressional Election Survey (CCES): 2010 to 2012 Panel Study" interviewed over 19,000 Americans in both timeframes and will capture voters' sentiment in this study (Ansolabehere and Schaffner, 2014). Second, this paper traces district-level voting behavior patterns from 2002 to 2018 as a function of geographic splitting. In line with the current doctrinal position of the courts, this paper does not look at how fairly elections reflect partisan division; instead, this research focuses on voter turnout and electoral competitiveness as quantified by MIT's Election Data and Science Lab (MIT Election Data and Science Lab, 2024).

2 Literature Review

2.1 Gerrymandering and its Operations

In the United States, members of the House of Representatives are elected not in statewide elections, but at the district level. Redistricting is the decennial process whereby states redraw district boundaries to account for population shifts captured in the latest census (Rossiter, Wong, and Delamater, 2018). The maps developed in every redistricting cycle are then used for congressional elections throughout the next decade (barring a successful court challenge). While the number of districts nationwide has been fixed at 435 since the 1920s, the amount that each state receives varies decade to decade (Kromkowski and Kromkowski, 1991). States with large population growth tend to gain district(s), and those with declining populations may lose district(s). While the district size in each state has some variance, seats in 2020 were assigned such that each state has one district for approximately every 760,000 inhabitants.[1]

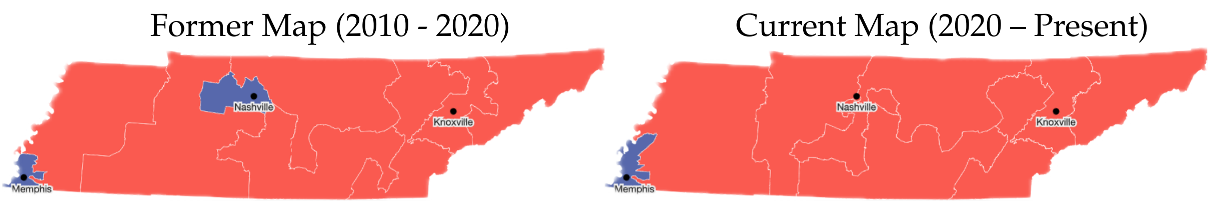
Federally, the only requirement that every state must adhere to during redistricting is to ensure that every district has an approximately equal population; otherwise, individuals in more populated districts would have diminished political impact (Congressional Research Service, 2021). Strikingly, these new electoral lines are often drawn by politicians themselves. Li, Limón, and Kirschenbaum (2021) find that in 38 states, district lines are determined either by politically appointed commissions or the state

[1]Some states have more or less populous districts than others by sheer mathematically construction. Rhode Island provides an extreme example. Rhode Island had a population of a little over one million people following the 2020 census. The state was assigned two congressional seats, meaning each district represented a bit more than half a million residents. If they were only given one representative, the district would represent all million plus Rhode Islanders. Because states can only be given a discrete number of congressional districts, smaller states tend to have more atypical district populations. Nonetheless, districts are assigned so that their populations are as close to the national average ($\sim$ 760,000) as possible.

legislators. Gerrymandering occurs when politicians redraw districts to benefit individual political parties. Engstrom (2020) defines gerrymandering as "the intentional manipulation of the boundaries of geographic election districts in order to facilitate the election of candidates of one political party."[2]

Gerrymandering primarily involves two techniques: cracking and packing. Cracking is the process wherein the party in power splits their opposition into separate districts to dilute their political voice (Stephanopoulos and McGhee, 2015). This strategy succeeds by splicing opposition votes among multiple districts such that they do not reach a majority in any of the districts. Figure 2.1 exemplifies this strategy. During the 2020 redistricting cycle, Tennessee Republicans were able to eliminate a Democrat-held seat centered in Nashville by cracking the liberal city among three districts. As the maps illustrate, state legislators split the former district centered on Nashville by assigning it to three largely rural districts (Wines, 2022). The left panel below illustrates the map prior to redistricting; the right panel shows Tennessee lawmakers' successful attempt to 'crack' Nashville and eliminate a Democrat-leaning seat:

FIGURE 2.1: Cracking — Tennessee Congressional Maps Before and After 2020 Redistricting Process

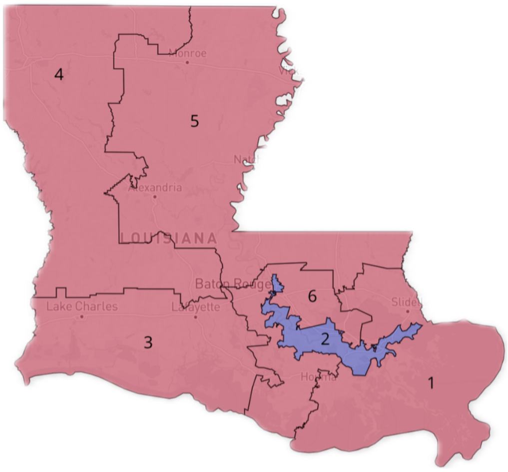


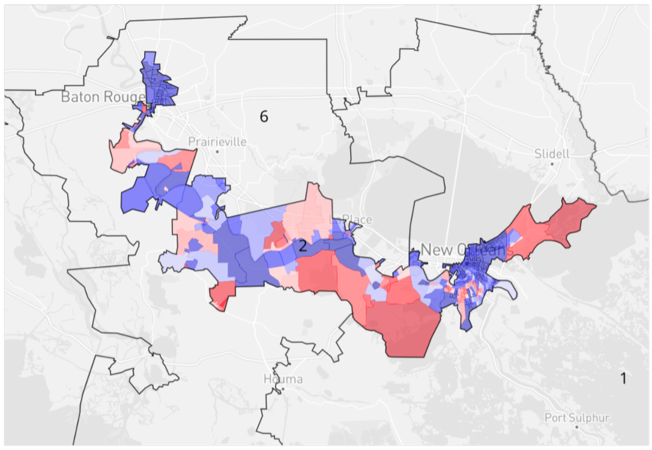
Conversely, mapmakers may attempt to 'pack' their opposition. Packing involves connecting several opposition strongholds into a singular district. In doing so, mapmakers can concentrate their opposition into one district and guarantee that the remaining districts will vote in their favor (Stephanopoulos and McGhee, 2015). Louisiana mapmakers have employed this strategy over the last several decades. Legislators have drawn a snake-like district (District 2) that connects the urban core of New Orleans to that of Baton Rouge. District 2 is expected to overwhelming vote for Democrats;

[2]Engstrom (2020), p. 23.

however, this leaves the remainder of the state as safely Republican territory (DeFord, Eubank, and Rodden, 2021). The left panel of Figure 2.2 shows Louisiana's current congressional map, and the right panel focuses in on District 2 to show how the map 'packs' Democrat-leaning precincts, shown in blue:

FIGURE 2.2: Packing — Louisiana's Current Congressional Map with a Focus on District 2





2.2 The Partisan Impacts of Gerrymandering

Via the process of cracking and packing, gerrymandering enables politicians to gamify elections to improve electoral outcomes in favor of their party. By comparing election results immediately before and after the redistricting cycle, Jeong and Shenoy (2022) demonstrate that the party in control of redistricting increases their chances of winning any individual race by 11% on average. However, the impacts of gerrymandering begin far before the final outcome at the ballot box. Stephanopoulos and Warshaw's (2020) employ a difference-in-difference approach to capture the downstream electoral effects before and after a redistricting cycle. When parties in power redraw the boundaries in their favor, minority parties are less likely to field a candidate at all, and even if they do, the candidate's credentials (measured in political experience) are significantly lower. Moreover, donor activity is significantly reduced when gerrymandering ensures safer seats (Stephanopoulos and Warshaw, 2020). By securing safe seats, gerrymandering can reduce the incentive to vote because any individual vote is less likely to sway the election. Jones, Silveus, and Urban (2023) demonstrate that gerrymandered maps are negatively associated with statewide voter turnout.

The impacts of gerrymandering have been especially pronounced in the last two decades. For example, despite Democratic candidates receiving 1.5 million more votes nationwide in the 2012 congressional elections, Republicans won a 33-seat majority. Engstrom (2020) attributes this disparity between representation and the popular vote to Republican-controlled state legislatures' attempts to gerrymander maps in the 2010 redistricting cycle. Gerrymandering is perpetrated by both major American parties. Before the 2020 redistricting cycle, Democrats regained majorities in several state legislative chambers and were empowered to redraw lines in their favor. After 2020, Democratic state legislators sought to fortify seats in states like Nevada, New Mexico, Illinois, Maryland, and New York. Although several of these Democratic-drawn maps were rejected by state courts, the party was able to flip several formerly Republican seats due redistricting (Rakich, 2022).

2.3 Traditional Methods of Quantifying Gerrymandering

Over the last several decades, statisticians have devised several measures for quantifying the extent of gerrymandering within a particular map. The most widely cited and utilized measure is referred to as the efficiency gap. Following the 2010 redistricting cycle, Nicholas Stephanopoulos and Eric McGee proposed the efficiency gap to quantify the disproportionate effect of gerrymandering on a given political party. Their metric relies on a term called "wasted votes". Wasted votes are ballots cast for candidates who did not benefit from that marginal vote. There are two types of wasted votes: (1) votes for a losing candidate, and (2) votes for a winning candidate above the 50% threshold (because whether a candidate receives 51% or 90% of the electorate, he or she will still be elected) (Bernstein and Duchin, 2017). The efficiency gap is calculated by netting the amount of wasted votes for each political party within a given state and dividing it by the total number of votes cast (Stephanopoulos and McGee, 2015):

$$\text{Efficiency Gap} = \frac{(\text{Wasted Votes}_{\text{Party A}} - \text{Wasted Votes}_{\text{Party B}})}{\text{Total Votes Cast}} \tag{2.1}$$

In Equation 2.1, a positive value would indicate that the given map is biased against Party A, and a negative value would indicate that it is biased against Party B. The efficiency gap has gained widespread support in the field because it offers a seemingly objective and normative guide to compare the differential effects of redistricting on political parties (Cover, 2018).

Beyond the efficiency gap, researchers have also proposed the median seat gap as a mechanism to quantify a map's bias. For every state, this approach takes a given party's electoral performance in each congressional district and compares its median performance to the average (or mean) of the state overall. A perfectly proportional map would have a median seat that matches the average (or mean) of the state as a whole. This measure holds that the discrepancy between the median seat and the state's average can be a measure of electoral skewness (Wang, 2016).

While both of these metrics (among others) have their merits, their success in the judicial system is minimal. Both measures appeal to (a) statewide discrepancies, and (b) the performance of political parties as a whole. The Supreme Court has repeatedly dismissed challenges that appeal to these statewide metrics. The Roberts Court has emphasized that such metrics merely demonstrate harms done to political parties and not individuals themselves, and only the latter group has constitutionally protected rights (Warrington, 2018). While the aforementioned measures are helpful in research settings to quantify gerrymandering, both measures are fruitless in efforts to judicially challenge gerrymandering (Rush, 2020).

2.4 Towards a Disaggregated Measure of Gerrymandering

While the only federal requirement states must adhere to while redistricting is to design districts of equal population, many states legislatures have enacted traditional districting principles (TDPs) to guide the mapmaking process for their own maps. TDPs are a set of guidelines that individual states instruct their mapmakers to consider while redistricting (Sabouni and Shelton, 2021). TDPs provide an initial springboard for motivating individualized metrics of gerrymandering. Although the language and presence of TDPs vary by state, most fall into five primary categories: (1) prioritizing compact district sizes, (2) ensuring all parts of the district are connected and continuous, (3) districts should preserve communities of interests within the same district, (4) districts should preserve prior electoral lines when possible, and (5) avoiding districts that force two incumbents to compete against each other (National Conference of State Legislatures, 2021). In practice, these principles are often loosely defined, without a clear benchmark for what qualifies as violating a TDP. Thus, Sabouni and Shelton (2021) document that many TDPs have been historically unenforced and politically motivated maps have persisted, even in states with robust TDPs.

Our paper falls into the broader context of gerrymandering literature in that it seeks to quantify one of these TDPs in particular: districts should preserve communities of interest (COIs) in the same district. Heading into the latest cycle of redistricting in 2020, 27 states had either passed laws or state constitutional amendments which require maps to preserve such COIs when possible (National Conference of State Legislatures, 2021). The objective of this redistricting principle is to prevent COIs from being split into separate congressional districts, as such a move would dilute their political voice (Rossiter, Wong, Delamater 2018). There is an intuitive appeal to this principle. Makse (2012) argues that the phrase “communities of interest” captures the idea that some geopolitical places “belong together"; however, he argues that COIs “are often in the eye of the beholder”, without any objective, enforceable definition of where one

COI ends and the other begins.[3] Arrington (2010) catalogues all the different ways academics, politicians, and lawyers have tried to define the term "communities of interest" in court and in scholarly work. COIs could be defined either by natural geographic boundaries or by some shared identify trait (i.e., by voting patterns or demographic features).

This paper focuses primarily on the disagreement over how to geographically define COIs boundaries (by county lines or by ZIP codes). Because these geographic boundaries rarely line up, mapmakers often must weigh the importance of preserving one geographic group (e.g., ZIP codes) versus another (e.g., counties). This paper works to compare the salience of these different geographic units with regards to redistricting to determine which subdivisions should be prioritized.

In an effort to define communities of interest, Curiel and Steelman (2018) argue that district lines should preserve ZIP codes as the primary community of interest. They argue that "the form, function, and rules behind creating ZIP codes creates a modern-day analogue to the ideal district."[4] Moreover, they argue that when ZIP codes are split among multiple districts, it confuses constituents as to who their representative is. Curiel and Steelman (2020) advocate that ZIP codes are the preferable subdivision to preserve because of their connection with the postal service; preserving ZIP codes as a political unit would "ensure that any one citizen would not find unnecessarily complicated impediments to identifying or receiving [mail] communication from their member of Congress."[5] Curiel and Steelman (2018) find a significantly negative relationship between the extent to which a ZIP code is split among multiple congressional districts and three outcomes of interest: constituents' (a) ability to correctly identify their representative, (b) prevalence of contact with their House member, and (c) perceived ideological divide with his or her representative. These deleterious impacts are more aligned with the Roberts Court's desire for individualized and local

[3]Makse (2012), p. 504.
[4]Curiel and Steelman (2018), p. 329.
[5]Curiel and Steelman (2020), p. 103.

metrics of gerrymandering. Curiel and Steelman (2018) work models constituent behavior in a merely correlational approach, controlling for demographic features like race, education, and income.

Other academics have proposed that instead of preserving ZIP codes, congressional maps should preserve county boundaries. Many academics argue that maps should avoid splitting counties because they "can be thought of as 'natural' communities of interest."[6] Bowen (2014) reveals that individuals are 8% more likely to recall their representative if their county is preserved in one district as opposed to being split among multiple. This conception of county preservation has grounding in actual law. For instance, Ohioans successfully passed a constitutional amendment which enumerated explicit ways to preserve geographic COIs; redistricted maps must respect county borders to the fullest extent possible. Concretely, the law requires that 65 of the 88 counties be contained wholly within one district, and only five of those 88 counties can be split between more than two districts (Ohio Laws and Administrative Rules 2021).

Bowen (2014) and Curiel and Steelman's (2018) work provides the motivation for this paper. While these researchers have demonstrated the advantages to preserving specific geographic boundaries, research has yet to establish which geopolitical boundary is most important to preserve. That is, research has not directly tested whether ZIP code splitting or county splitting is more detrimental for the electoral process and constituent engagement (Grofman and Cervas, 2021). This is a particularly important question to answer, given that county lines and ZIP code boundaries are not perfectly aligned. ZIP codes (drawn by the Postal Service) are designed completely independent of county borders (Rossiter, 2014). Figure 2.3 exemplifies how the ZIP codes in the Raleigh area do not perfectly align with Wake County's border. ZIP code boundaries are shown in grey, and the county line is in green. As you can see, over a quarter (14) of the county's 40 ZIP codes cross the county boundary.

[6] Bowen (2014), p. 864.

FIGURE 2.3: Imperfect Overlap Between 40 ZIP Codes and Wake County Line in Raleigh

Given that county lines and ZIP codes are not perfectly nested, congressional maps cannot wholly preserve both ZIP code and county borders. Our work seeks to compare the differential electoral impacts on constituents depending on whether their ZIP code or county is split in redistricting. In doing so, our research specifically will look at the differential ways ZIP code and county splitting affects individual voters' engagement with the political process. Moreover, our research expands upon the previous empirical framework of Curiel and Steelman's (2018) ZIP code work and Bowen's (2014) county work. Their previous works were merely correlational and did not take into account how recently a ZIP code or county had been split by gerrymandering. We begin by largely replicating their correlation work, but we also exploit recent changes in district lines and employ a difference-in-difference model to parse out potential timing effects. multirow

3 Data

To assess the impact of geographic splitting on political outcomes, this paper draws on four primary data sets: the Missouri Census Data Center, the "Cooperative Congressional Election Survey (CCES): 2010-2012 Panel Study", MIT Election and Science Lab's electoral database, and IPUMS American Community Survey data.

The Missouri Census Data Center is a local partner of the federal Census Bureau that is curated and overseen by the University of Missouri (The University of Missouri, 2022). This dataset enumerates the overlap between any two census-recognized geographic boundaries. For our purposes, the Missouri Census Data Center will produce several dyads or overlaps. From 2010 to 2012, we create county-district and ZIP code-district dyads.[1] These dyads define which district(s) occupy a specific county or ZIP code. If multiple districts occupy the same county or ZIP code, then we can label that region as "split" between multiple districts. Separately, we construct district-county and district-ZIP code dyads from 2002 to 2018. The difference in order matters. The first set of dyads tell us whether or not a county or ZIP code is split; whereas, district-county and district-ZIP code dyads tell us how many counties or ZIP codes a given congressional district occupies.

Responses to Harvard University's "Cooperative Congressional Election Survey (CCES): 2010-2012 Panel Study" provide many of our outcome variables of interest (Ansolabehere and Schaffner, 2014). This longitudinal study interviewed a pool of tens of thousands of respondents before and after both the 2010 and 2012 election cycles.

[1]Technically, the Missouri Census Data Center does not list the 2010 congressional boundaries as a potential geographic layer. In this research, we look at the 2008 borders and compare those to county and ZIP code lines for our 2010 analysis. This is appropriate given that no congressional lines were redrawn or modified between 2008 and 2010. They were redrawn between 2010 and 2012, and this dataset indeed has the adjusted district lines following redistricting.

All interviews were conducted online via the platform YouGov, and the initial survey population was confirmed to be nationally representative via sample matching. In total, approximately 19,500 Americans responded to the panel survey in both 2010 and 2012; however, there are some potential concerns with attrition. Approximately 44% of initial participants re-answered the 2012 interview questions. The issues of attrition were fairly uniform across all demographic groups, except for African Americans and non-voters whose attrition rates were noticeably higher than other groups. The unequal attrition rates across all demographics groups may introduce some degree of sampling bias, but the surveyors interviewed additional respondents of underrepresented groups during the 2012 wave of interviews. Survey participants voiced their responses to various questions that tracked their individual sentiments and knowledge with regards to contemporary politics. Our study looks at responses to a select few questions. First, surveyors provided respondents with their representative's name and asked for their political party; we track the extent to which they answered with the correct answer. Second, respondents provided their approval rating of Congress and their individual on a scale of 1-4. Finally, we track the extent to which respondents reported that they participated in any one of the following political activities within the previous year: attended a political meeting, put up a political sign, worked for a campaign, attended a protest or demonstration, contacted a public office, or donated money to a campaign.

Survey respondents were asked to provide geographic data on both their county and ZIP code, which allows us to connect their responses to the extent to which their region was split among congressional districts. Out of the roughly 19,000 respondents, nearly all respondents provided their county information in both timeframes; however, the same is not true for ZIP code data. All respondents provided their ZIP code data in 2012, but only a fraction ($\sim$ 3,800 respondents) provided their ZIP code in 2010. We employ an imputation method to infer the lion's share of these respondents' ZIP codes. Given that we have complete geographic data for $\sim$ 3,800 respondents, we look to see the extent to which county data can predict ZIP code data. Table 3.1 compares what fraction of respondents changed their ZIP code from 2010 to 2012 conditional on whether they moved counties. We observe that $\sim$ 90% of respondents who did not move counties also did not move between ZIP codes. We find this value sufficiently high to backwards induce respondents' ZIP codes in 2010: if a respondent did not change counties from 2010 to 2012, we impute their 2010 ZIP code as the same as their 2012 ZIP code.

TABLE 3.1: Cross Tabulation of Survey Respondents Who Changed ZIP Codes and Counties: 2010 to 2012

Does ZIP Code Change from 2010 to 2012?	Does County Change from 2010 to 2012?		
	0 (No)	1 (Yes)	Total
0 (No)	3,112	54	3,166
	89.71%	15.04%	82.71%
1 (Yes)	357	305	662
	10.29%	84.96%	17.29%
Total	3,469	359	3,828
	100.00%	100.00%	100.00%

First value in cell is raw count of regions.
Second value in cell is percent of column.

MIT Election and Science Lab's electoral database provides other informative outcome variables of interest. This database provides election data for the U.S. House of Representatives since 1976 (MIT Election Data and Science Lab, 2024). This dataset provides figures on congressional district level electoral results, chiefly the number of votes cast in each district and the percent of the vote received by the winning candidate. We extract electoral outcomes for all congressional races from 2002 to 2018. We merge this dataset with a set of district-level control variables (total population, median age, gender, race, and average income) provided by IPUMS American Community Survey data. We additionally merge the electoral dataset with data from the University of Missouri Data Center to see how many ZIP codes or counties each congressional district splits each year. This contrasts from the aforementioned use of the University of Missouri's data. Previously, we described measuring how many districts a region is split between. Because electoral data is only provided at the district level, our splitting data must be at the district level, not with respect to counties or ZIP codes. We therefore quantify how many regions a district splits (i.e., how many regions a district cuts through) for our electoral outcomes.

4 Methods and Approach

This paper tracks the impact of geographic splitting in redistricting along two primary levels of analysis: (1) changes in individual constituents' sentiment and knowledge, and (2) changes in localized voting patterns.

4.1 Changes in Individual Constituents' Sentiment and Knowledge

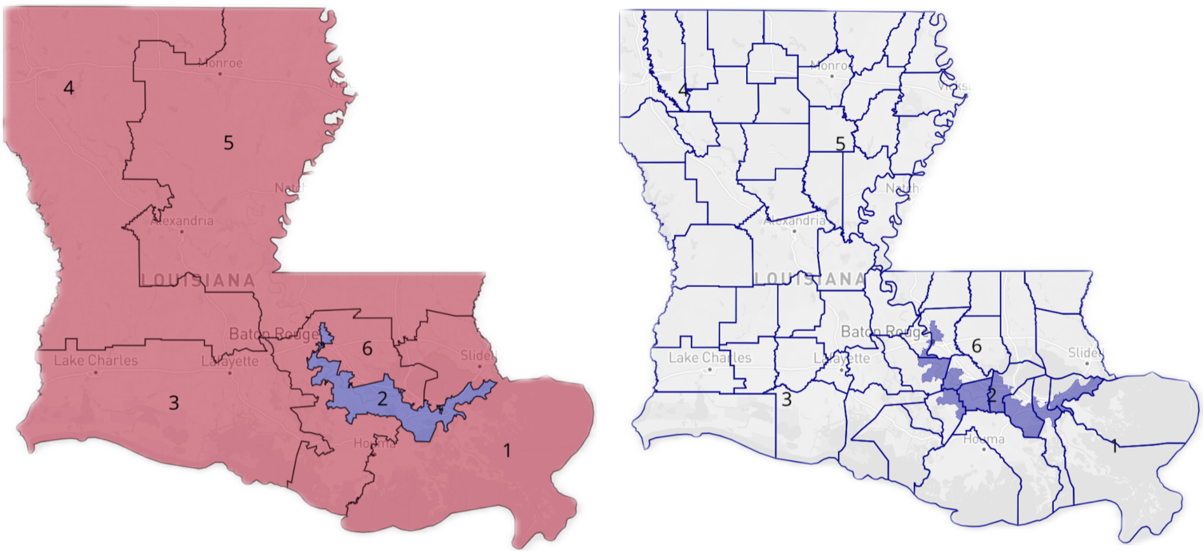
The first level of analysis of this paper studies how redistricting affects individual voters outside of the ballot box. In doing so, our primary independent variables are the extent to which an individual's ZIP code or county is split, and we look at survey data for our outcomes of interest.

4.1.1 Quantifying Independent Variables: Geographic Splitting

Curiel and Steelman (2018) quantify the impact of redistricting by measuring how many times a ZIP code is split between multiple congressional districts (Curiel and Steelman, 2018). Here, we generalize this approach to both ZIP codes and counties to compare the differential effects of geographic splitting. For a region (either ZIP codes or counties) to be split, the region must be divided between multiple congressional districts. This approach is preferable to statewide measures because it allows for (1) localized quantifications of a given area's exposure to splitting and (2) greater consistency with the underlying strategies of cracking and packing.

Cracking and packing may be different redistricting strategies, but both approaches require geographic splitting. The relationship between geographic splitting and cracking is a bit more intuitive: cracking requires mapmakers to split certain regions to dilute voters' voices between multiple districts. While the intent behind packing is different, packing also requires splitting. This is because packing empowers mapmakers to carve through geographic regions to engulf pockets of like-minded voters. Figure 4.1 illustrates this phenomenon. The left panel shows Louisiana's current congressional map. Its second district epitomizes packing by carving through several downtown urban cores to concentrate Democrats into one heavily left-leaning district and leaving the remaining five districts safely red. Nonetheless, the right panel illustrates how that 'packed' district (District 2) must split several (9) counties along the way to attach several urban cores:

FIGURE 4.1: Packing Requires Geographic Splitting — Louisiana's 2022 Congressional Map



To uniformly trace the impact of redistricting, we construct an indicator variable (geography_split[1]) equal to one if a geographic region is split between multiple districts in a given year and zero otherwise. This allows us to label geographic regions as either "split" or "unified" in any individual time period.

[1]In this paper, variable names that include the word 'geography' are meant to capture the fact that the variable could either be with respect to ZIP codes or counties. In the results section, we will discuss the differential effects of the variables in the ZIP code and county dimension.

On average, ZIP codes are much smaller and less populous than counties. The average ZIP code had a population of ∼9,400 people after the 2010 census, and the average county had a population of ∼98,000 people. Therefore, extending Curiel and Steelman (2018) approach for ZIP codes to the county level of analysis introduces additional complexities. While ZIP codes are generally quite small in population, counties may be well over a million people. Because each congressional district is mandated by law to have roughly the same population, each district contained approximately 760,000 Americans in 2020 (Hyer, 2021). This presents a unique challenge: counties may be "split" simply because they're too populous to fit wholly within one congressional district.[2] Approximately 5% of all counties exceeded the size of an average congressional district in 2010 (Winburn and Wagner, 2010).

To distinguish between artificial and mathematically-required county splitting, we will use an adjustment factor. The adjustment procedure is as follows:

1. We take the population of a given county and divide it by the population size of the state's congressional districts for that decade.
2. We round that result up to the nearest integer (this value will describe the minimum number of districts mathematically-required to represent a given county).
3. If the county is split more than what is mathematically-required, it is labeled as "split." A county that is divided amongst multiple districts may still be labeled as "unified" if the county is split the minimum number of times.[3]

[2]For example, Los Angeles County is the largest county in the United States—over 10 million residents live within its borders. As such, it would be impossible to preserve this county in one congressional district of ∼760,000 people.

[3]An example here may help. With a population of over 1.1 million people, Wake County, North Carolina houses the state's capital: Raleigh. It is mathematically-impossible for Wake County's residents to all be within one district. The minimum required number of districts within Wake County is two (1.1 million divided by 760,000 rounds up to 2). In 2022, the state passed a congressional map that divides Wake County into only 2 districts. This means the county was split the mathematically-required minimum number of times. Therefore, Wake County would be marked as "not artificially split." On the flip side, take Los Angeles (LA) County. Dividing its population size by the district size, LA County must have at least 14 districts within its borders. The current map features 18 different districts within LA County. This means the map splits the county above and beyond what is mathematically-required.

Equation 4.1 defines the mathematically-required number of splits for each county. The "ceil" function rounds the value in the parenthesis up to the next integer. Only if a county's split count exceeds this value will the county be labeled as "split":

$$\text{Mathematically-Required Number of County Splits} = \text{ceil}\left[\frac{\text{County Population}}{\text{State's District Population Size}}\right] \tag{4.1}$$

The map in Figure 4.2 illustrates the amount of districts a county is split between following adjustments. Darker shaded counties are split by more districts, and lighter shaded districts are split by fewer districts.

4.1.2 Measuring Constituent Knowledge and Sentiment

This research analyzes the association between geographic splitting and voter behavior along three major dimensions: representative knowledge, approval ratings, and political engagement indices. Data for all outcome variables come from survey responses from Harvard University's "Cooperative Congressional Election Survey (CCES): 2010-2012 Panel Study" (Ansolabehere and Schaffner, 2014). All outcome variables listed below were recorded from the survey following the 2010 midterm elections and the 2012 elections.

First, respondents were given the name of his or her congressional representative, and they were asked if they recognized the person's name and knew their party identification. Respondents who recognized their representative's name and knew their correct party affiliation are coded as correctly identifying their representative. Previous research into geographic splitting for counties and ZIP codes both use this outcome variable to measure voters' knowledge (Curiel and Steelman, 2018; Bowen, 2014; Bowen and Clark, 2014). Second, survey responses capture voters' self-reported approval ratings of (1) their representative and (2) Congress as an institution. Respondents rated their approval on a scale from 1-4 (from strongly disapprove to strongly approve). Curiel and Steelman (2018) nor Bowen and Clark (2014) examine approval

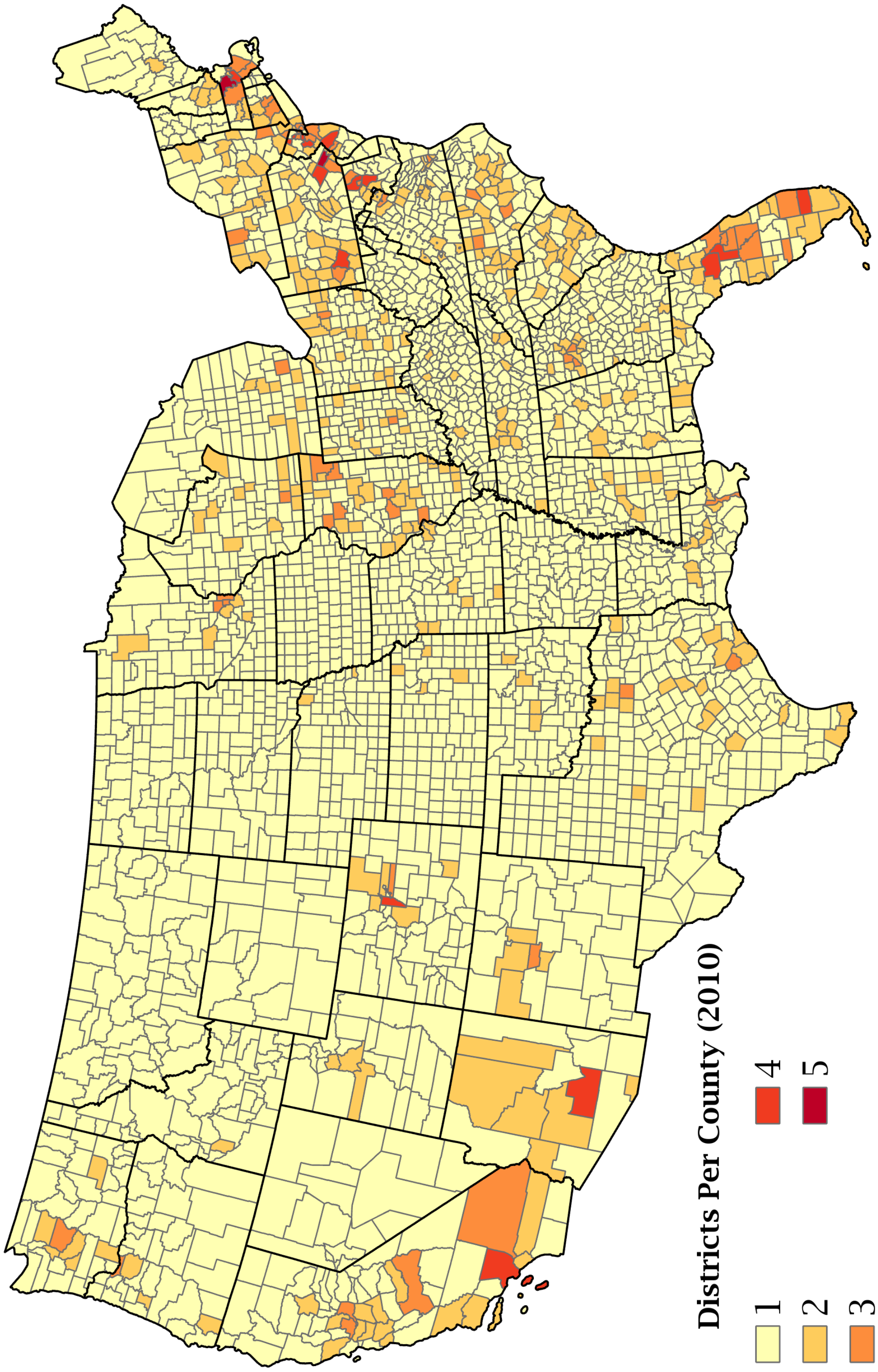


FIGURE 4.2: Map of Number of Districts a County is Split Between (After Adjustments)

ratings as a potential outcome of interest in their ZIP code and county splitting analysis, respectively. We are examining it here to see to what extent geographic splitting deteriorates constituents' approval of their legislators. Finally, we look at individuals' involvement in political activity as an outcome of interest. We create an indicator variable equal to 1 if the respondent participated in any one of the following activities in the year proceeding an election: attended a political meeting, put up a political sign, worked for a campaign, attended a protest or demonstration, contacted a public office, or donated money to a campaign.

4.1.3 Empirical Approach

Curiel and Steelman (2018) and Bowen (2014) employ a purely correlational approach to model geographic splitting's impact on constituent behavior. Here, we follow similar procedures to estimate the relationship between splitting and survey responses across all outcomes of interest. We do this to expand their work to outcomes of interests not yet explored (namely, approval ratings), and second, to compare the relative significance of these results across the ZIP code and county levels of analysis. Our regression of interest is outlined below in Equation 4.2:

$$
\begin{aligned}
Y_{igt} = \; & \beta_0 + \beta_1 \text{Geography is Split}_{gt} + \beta_2 \text{Black}_i + \beta_3 \text{Hispanic}_i + \beta_4 \text{Asian}_i + \\
& \beta_5 \text{Native American}_i + \beta_6 \text{Mixed Race}_i + \beta_7 \text{Other Race}_i + \\
& \beta_8 \text{Middle Eastern}_i + \beta_9 \text{Age}_{it} + \beta_{10} \text{High School Graduate}_{it} + \\
& \beta_{11} \text{4 Year Degree}_{it} + \beta_{12} \text{Professional Degree}_{it} + \beta_{13} \text{Female}_{it} + \\
& \beta_{15} \text{Mobility}_{it} + \beta_{16} \text{Lower Class}_{it} + \beta_{17} \text{Upper Class}_{it} + \\
& \text{State Fixed Effect} + \text{Year Fixed Effect} + \epsilon_{igt}
\end{aligned}
\tag{4.2}
$$

In Equation 4.2, Y is our survey responses for individual i in geographic region g during time t. We include many of the same controls as Bowen (2014) and Curiel and Steelman (2018). We include respondent-specific controls to account for their self-reported race, age (in years), level of education (measured with what degrees they have attained), gender, mobility (measured in number of years respondent had lived in their reported address), and income. Because the surveyors provided different bands of income in the 2010 and 2012 surveys, we cannot directly control for a respondent's income. Instead, we define a respondent as being "Lower Class" if they reported earning less than $30,000 a year and "Upper Class" if they reported earning above $150,000 a year. The lower bound represents roughly 200% the federal poverty line, and the upper bound represents the top 20% of earners nationally. Given the previous work and data, we hypothesize that outcomes of civic engagement will generally be higher among white, educated, upper class males who have lived in their area for longer periods of time. We also include state and year fixed effects to net out the average trends over time (from 2010 to 2012) and within states.

The first covariate will indicate the general relationship between survey responses and geographic splitting. We will compare both the sign and the significance of this coefficient across the ZIP code and county models for all outcomes of interest.

We recognize that this correlational approach likely does not capture the true causal relationship due to potential problems with endogeneity. Particularly, mapmakers often redraw district boundaries with the intent of political gain. As such, map makers may split regions with higher levels of civic engagement to dilute their voices. If this is the case, Equation 4.2 would be biased. As such, we expand upon the correlational approach outlined above and executed by Bowen (2014) and Curiel and Steelman (2018). We expand upon their work by utilizing a difference-in-difference approach. Our research exploits the 2010 redistricting process to examine how changes in district lines affected survey respondents' political behavior between the 2010 and 2012 election cycle.

As aforementioned, congressional lines are redrawn every decade. This creates a quasi-experiment: from 2010 to 2012, all 435 congressional boundaries were redrawn. During this redistricting process, many ZIP codes and counties that were formerly "unified" become "split" (or vice versa). The redistricting process therefore creates four classes of regions:

1. Regions who remained unified before and after redistricting.
2. Regions who remained split before and after redistricting.
3. Regions who were initially unified but became split due to redistricting.
4. Regions who were split but became unified during redistricting.

Table 4.1 summarizes the distribution of counties and ZIP codes between all four potential classes.

TABLE 4.1: Distribution of Regions by Split Status from 2010 to 2012

Split Status	**Counties**	**Zip Codes**
Never Split	2,545 81.03%	24,960 75.76%
Always Split	242 7.71%	3,457 10.48%
Split to Unified	184 5.86%	2,373 7.20%
Unified to Split	171 5.44%	2,158 6.56%
Total	3,142 100%	32,948 100%

*F*irst value in cell is raw count of regions.
*S*econd value in cell is percent of total regions.

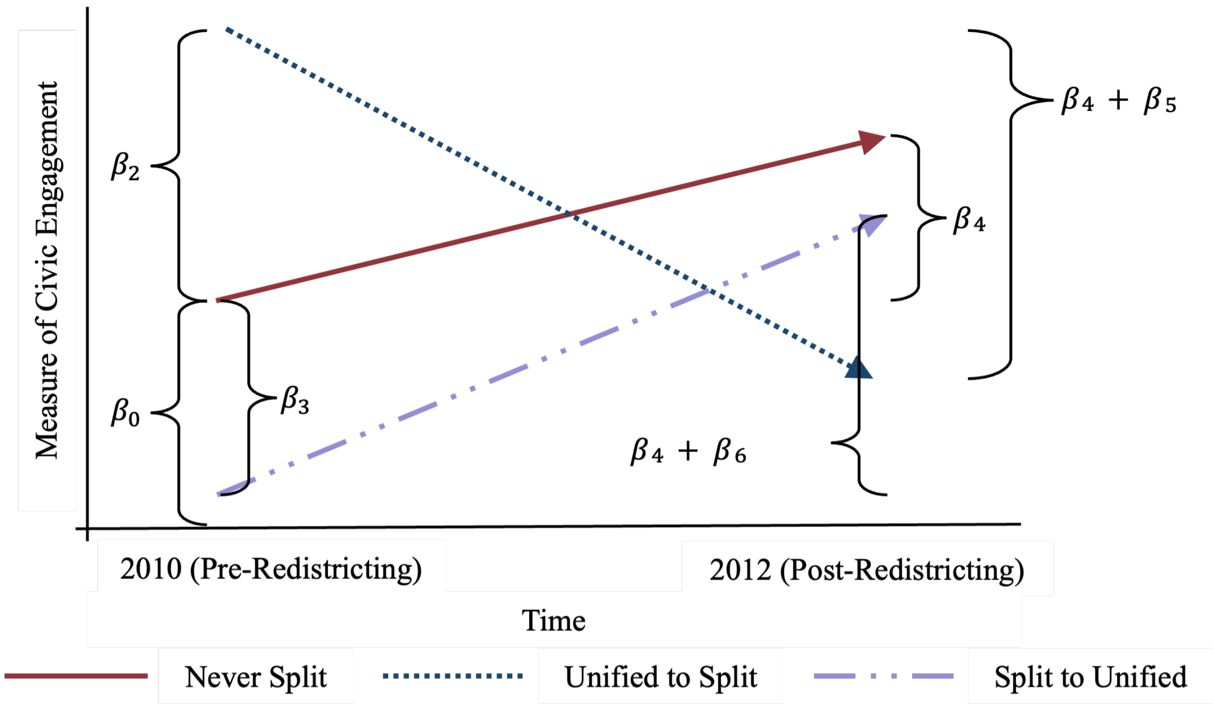
Our paper utilizes a difference-in-difference approach to compare how survey respondents' behavior changed as a function of their exposure to geographic splitting. In our work, the act of splitting a region via redistricting between 2010 and 2012 will operate as our treatment. There are several nuances with this approach. First, several counties (7.7%) and ZIP codes (10.5%) are split in both timeframes. Because our treatment is getting split during this redistricting cycle, these regions are not considered "treated" for our approach because they were already split before the 2010 redistricting cycle commenced. Second, the exposure to treatment can happen in both directions. That is, some respondents' counties or ZIP codes will become split, but others will become unsplit. Our framework recognizes the dual-directionality of this approach. We refer to the act of getting split (going from unified to split) as undergoing the hypothesized "negative treatment." Inversely, we refer to the act of getting unsplit (going from split to unified) as undergoing the hypothesized "positive treatment." For our difference-in-difference approach, we execute Equation 4.3:

$$\begin{aligned} Y_{igt} = \; & \beta_0 + \beta_1 \text{Region Always Split}_g + \beta_2 \text{Region Unified to Split}_g + \\ & \beta_3 \text{Region Split to Unified}_g + \beta_4 \text{Post Redistricting}_t + \\ & \beta_5 \text{Region Unified to Split}_g \times \text{Post Redistricting}_t + \\ & \beta_6 \text{Region Split to Unified}_g \times \text{Post Redistricting}_{t+} \\ & \text{State Fixed Effect} + \mu_i + \epsilon_{igt} \end{aligned} \tag{4.3}$$

In Equation 4.3, the subscripts refer to individual i, residing in geographic region g, at time t. This regression will be run twice for each outcome of interest, once with ZIP codes as the geographic region at issue and once for counties. The term μ_i describes the individual fixed effect, and we will include a state fixed effect. As aforementioned, each geographic region will have an indicator variable to denote whether it has been split. The variable "Region Unified to Split" will be an indicator variable that equals one if the region went from unified in 2010 to split after the redistricting process concluded in 2012; inversely, the variable "Region Split to Unified" will equal one if the region went from split to unified after the redistricting process. The variable "Region Always Split" captures regions that were split both before and after redistricting. This implies that the omitted (baseline) group for this regression will be counties or ZIP codes who remained unified before and after the 2010 redistricting cycle. The fourth covariate ("Post Redistricting") will capture the difference over time for our untreated regions. The fifth and sixth covariates interact these time trends with our negative and positive treatment groups, respectively.

The regression in Equation 4.3 is unique in that it allows the magnitude of the negative treatment's impact to be different from that of the positive treatment ($\beta_5 \neq -\beta_6$ by construction). We will look at the significance of β_5 and β_6 to determine whether ZIP code or county splitting significantly affects the outcomes of interests. From there, we will compare the significance of these coefficients in the ZIP code vs. county models. Figure 4.3 is an idealization of Equation 4.3:

FIGURE 4.3: Graphical Idealization of Regression Equation 4.3



As Figure 4.3 illustrates, we expect most measures of voter engagement to increase from 2010 (pre-redistricting) to 2012 (post-redistricting) simply because 2010 was a midterm election and 2012 was a presidential election. Voter behavior and civic engagement across a host of metrics are generally muted in midterm elections compared to presidential ones (Jackson, 2000). By using regions that were never split as a baseline, we can effectively control for these time trends to see the differential behavior for voters whose regions went from unified to split or vice versa. As such, β_5 and β_6 will be the primary coefficients of interests because they describe how such voters acted in comparison to those who were not split during both periods. These coefficients will describe whether or not the treatment was associated with a significant deviation from baseline time trends.

For β_5 and β_6 to completely capture the casual relationship, we must assume (1) parallel trends between the treated and untreated groups and (2) that respondents could not anticipate such changes in their split status. With regards to the parallel trends assumption, our panel data is limited to just 2010 to 2012. As such, we cannot directly test nor observe the trends before the redistricting cycle in question. However, we find this assumption reasonable. Prior to the 2010-2012 redistricting cycle, the last nation-wide redistricting occurred a decade prior (from 2000-2002). That means that constituents in these regions would have had a decade to internalize the preexisting impacts of redistricting prior to the timeframe in question. We find it reasonable that a decade is enough time for voters' behavior to stabilize and for constituents of these impacted regions to have parallel survey response trends. Moreover, both the positive and negative treatment regions were broadly diverse in geographic breadth and population size.

The second required assumption is that respondents could not anticipate redistricting changes. We also find this assumption reasonable because the redistricting cycle could not begin by law until after the 2010 election cycle had concluded. This is because lawmakers had to wait for census data to be released and for new legislatures to be elected before they could design and approve prospective maps. As such, there would be no widespread mechanism for the public to know their future split status when new maps had yet to even be designed.

4.2 Changes in District-Level Voting Patterns

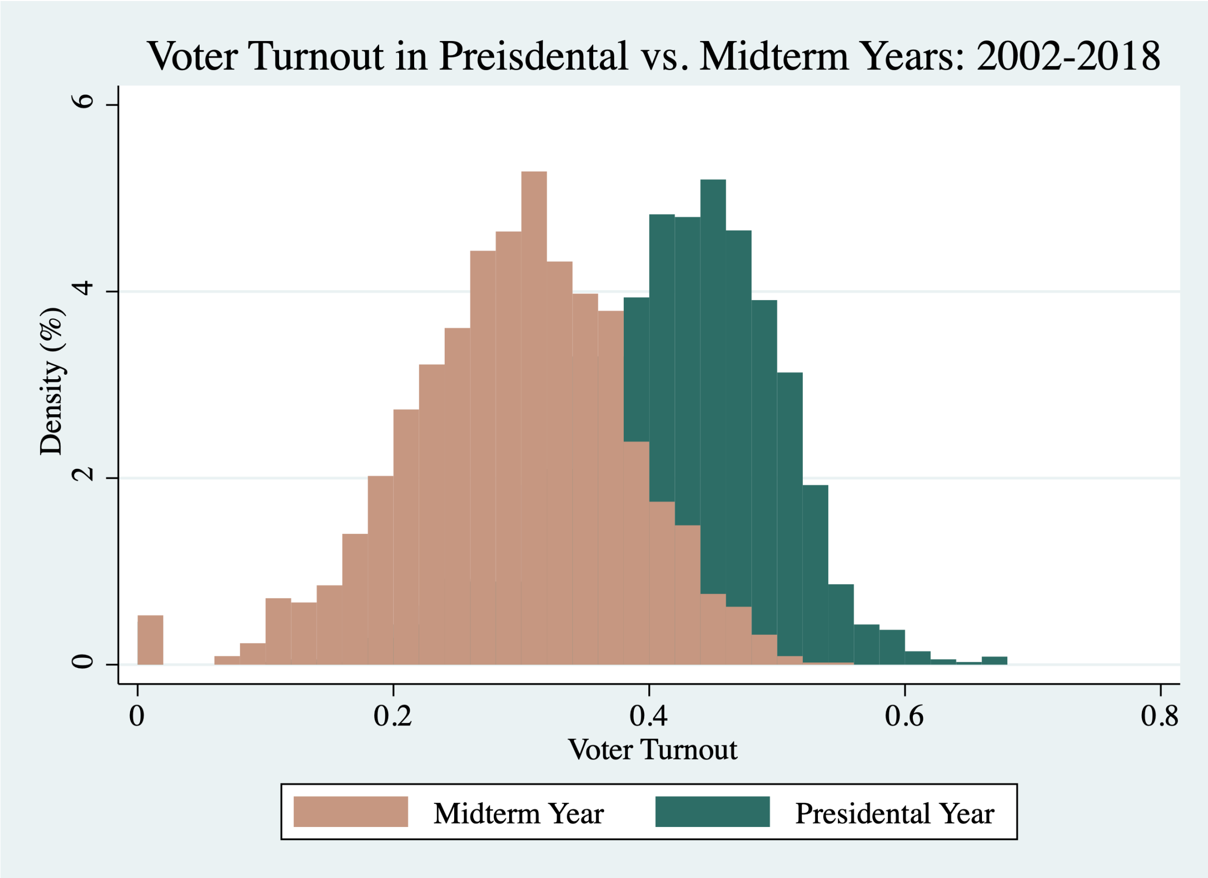
The second level of our analysis looks at the changes in district-level voting patterns from 2002 to 2018. Unlike the other level of analysis, this portion allows us to look at a larger timeframe in question and examine actual electoral results. As such, the unit of observation in this section of analysis will be congressional districts themselves, not individual voters.

4.2.1 Electoral Results at Issue

The Court's current doctrinal approach has made clear that they are not interested in the extent to which gerrymandering distorts electoral outcomes. That is, the Court has rejected metrics that look at how proportionally elections reflect the partisan distribution of voters. In light of their doctrine, we do not examine the outcomes of elections but the procedure itself. We look at two measures of electoral performance that gauge democratic health and civic engagement at the local congressional district level: voter turnout and winning percentages. MIT's Election and Science Lab provides the electoral data necessary to calculate both of these outcome variables.

Voter turnout is used here as a metric to capture civic engagement in given elections. We will interpret higher levels of voter turnout as an indicium for heightened civic engagement. The overlaid histograms in Figure 4.4 highlight the distribution of voter turnout in congressional elections from 2002 to 2018, segmented by midterm and presidential years. Evidently, we expect voter turnout to be higher in presidential election years, as compared to midterm election years. We also look at elections' winning percentages as an indicator for electoral competitiveness. A given district's winning percentage will be defined by the percentage of the vote that the winning candidate received in a given election. If a candidate runs unopposed, the winning percentage will by construction be 1 (for 100% of the vote). We look at this outcome variable to see the extent to which geographic splitting leads to "safer" elections. Given that competitive elections require candidates to fight more for constituents' votes, we will interpret lower winning percentage as democratically preferable.

FIGURE 4.4: Distribution of Congressional-Election Voter Turnout in Midterm vs. Presidential Election Cycles



4.2.2 Choices of Controls

Previous works have considered a laundry list of different idiosyncratic characteristics that partially explain voter behavior. In a comprehensive literature review of previous work in voting behavior, Kulachai et. al (2023) identifies several demographic traits as some of the most salient factors influencing voter behavior: race, gender, income, education, and age. Given the work of Kulachai et. al (2023), we expect wealthier, older, and more educated individuals to have higher levels of political involvement, regardless of their exposure to geographic splitting.

Here, we control for these factors. Using IPUMS American Community Survey (ACS) Data, we control for a congressional district's median age (in years), racial breakdown, average income, and gender split (in percent female). With regards to the racial breakdown, we control for the percent of each district that is White and Black separately. We normalize the average income of all districts to be in terms of 10,000 2018 USD. Given that district lines remained largely the same from 2002-2010, the control variables for districts remain constant from 2002-2010. However, state courts required various maps to be redrawn from 2012-2018. We therefore adjust the control factors for each district after every election. Due to that fact that the ACS data does not provide the average years of education at the congressional-district level, we are unable to control for education in this model.

4.2.3 Empirical Approach

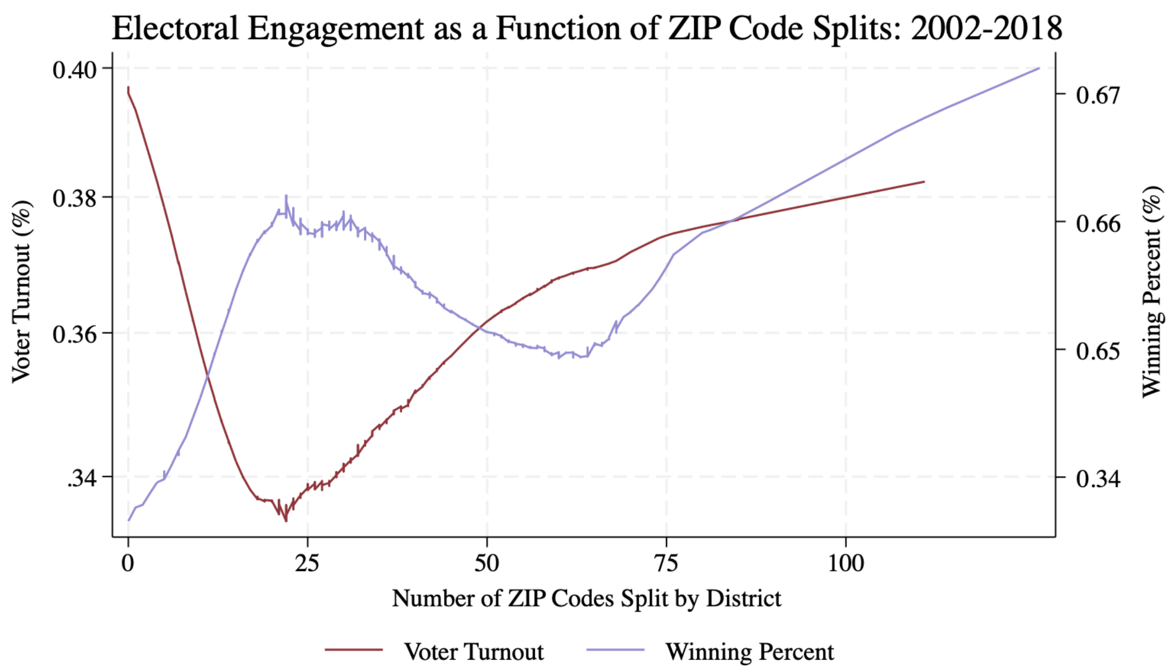
Unlike previous work into election turnout, we examine the extent to which localized measures of gerrymandering may influence electoral outcomes. Specifically, we test whether the number of counties or ZIP codes a district splits is predictive of its electoral outcomes. The difference between this approach and the previous one with survey data is minor but important. Here, we look at the number of regions a district splits up (i.e., any region a district cuts through but does not wholly contain); whereas in the previous section of analysis, we look at the number of districts a region is split between. We do this because election data is only available nationwide at the congressional district level, not at the county or ZIP code level. Below are the primary equations of interest in this level of analysis:

$$\begin{aligned} Y_{dt} = \; & \beta_0 + \beta_1 \text{Number of Counties Split}_{dt} + \beta_2 \text{Number of Counties Split}^2_{dt} + \\ & \beta_3 \text{Median Age}_{dt} + \beta_4 \text{Black Percent}_{dt} + \beta_5 \text{White Percent}_{dt} + \\ & \beta_6 \text{Female Percent}_{dt} + \beta_7 \text{Per Capita Income}_{dt} + \text{Year Fixed Effect}_{\,t} + \epsilon_{dt} \end{aligned} \tag{4.4}$$

$$
\begin{aligned}
Y_{dt} = \; & \beta_0 + \beta_1 \text{Number of ZIP Codes Split}_{dt} + \beta_2 \text{Number of ZIP Codes Split}^2_{dt} + \\
& \beta_3 \text{Median Age}_{dt} + \beta_4 \text{Black Percent}_{dt} + \beta_5 \text{White Percent}_{dt} + \\
& \beta_6 \text{Female Percent}_{dt} + \beta_7 \text{Per Capita Income}_{dt} + \text{Year Fixed Effect}_{\,t} + \epsilon_{dt}
\end{aligned} \tag{4.5}
$$

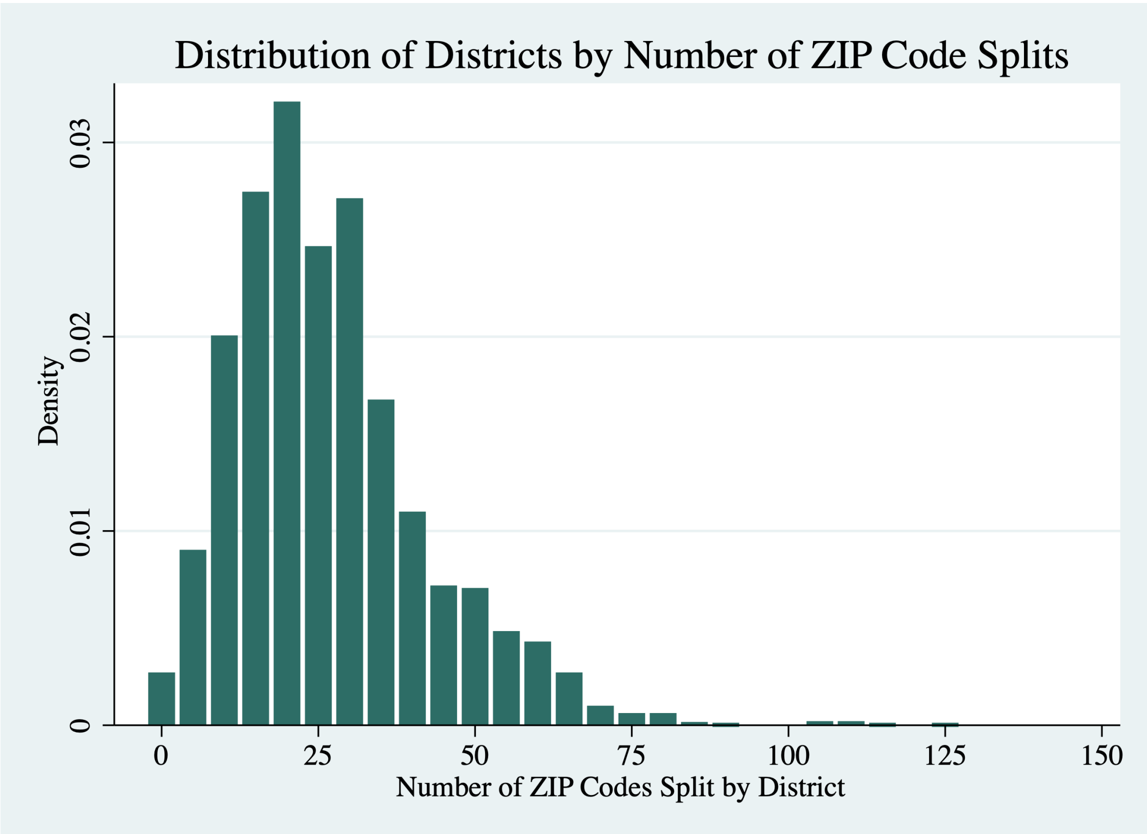
The equations above track electoral results Y for district d in time t. We construct a quadratic regression given the non-linear relationship we observe in the data. In Figure 4.5 below, we plot a smoothed scatter plot that helps visualize the relationship between ZIP code splitting and electoral outcomes. An analogous plot with regards to county splitting is displayed in Appendix A. The smoother scatter plot below demonstrates the seemingly non-linear relationship between geographic splitting and electoral outcomes:

FIGURE 4.5: Smoothed Scatter Plot of Electoral Engagement as a Function of the Number of ZIP Codes Split by District



We recognize that the relationship in the Figure 4.5 changes directions quite frequently. We particularly observe that at higher levels of ZIP code splitting (> 60 splits) the relationship takes an unanticipated skew. We include a histogram visualizing the distribution of ZIP code splits by district in Figure 4.6. ZIP code splits above 60 are quite rare. Less than 1% of districts feature such high levels of ZIP code splitting. We likewise include a similar histogram for counties split by districts in Appendix B.

FIGURE 4.6: Histogram of Number of ZIP Codes Split by District



Given that geographic splitting is a vital component to both cracking and packing, we hypothesize that geographic splitting will impact both electoral turnout and winning percentages. First, we theorize that more geographic splitting will reduce voter turnout. Second, we hypothesize that more splitting will be associated with an increase in winning percentages, as splitting may be used as a tool to secure safe seats and eliminate competitive ones.

5 Results

5.1 Changes in Individual Constituents' Sentiment and Knowledge

First, we examine the extent to which geographic splitting affects individual voters' behavior and knowledge. For this section, all the outcome variables of interest were captured in the 2010-2012 CCES Panel Survey. Below, we segment our results between (1) the correlational model akin to Bowen (2014) and Curiel and Steelman (2018) and (2) the difference-in-difference approach described in Equation 4.3 above.

5.1.1 Correlational Results

We first regressed CCES survey results on the extent to which a respondent was exposed to geographic splitting. Particularly, we summarize the results of Equation 4.2 in Tables 5.1 and 5.2 below. Table 5.1 summarizes the correlational results with regards to county splitting, and Table 5.2 summarizes the results with regards to ZIP code splitting.

We begin with the county model. We observe a statistically significant relationship at the 1% significance level between county splitting and three of the four CCES survey results of interest: the rate at which respondents could correctly identify the party of their representative, their reported approval of Congress, and their likeliness to engage in political activity. We do not observe a significant relationship between county splitting and representative approval ratings. Curiously, we find negative and significant relationships between county splitting and (1) respondents' involvement in

TABLE 5.1: Correlational Regression Results for CCES Survey Results on County Splitting

	(1) Correct Rep Party ID	(2) Approval of Rep	(3) Approval of Congress	(4) Political Activity
County is Split	-0.0123***	-0.0204	0.0290***	-0.0188***
	(-2.69)	(-1.50)	(2.95)	(-3.13)
Black	-0.0664***	0.175***	0.500***	-0.0506***
	(-8.72)	(7.51)	(29.99)	(-5.04)
Hispanic	-0.0845***	0.0107	0.244***	-0.0548***
	(-9.37)	(0.38)	(12.43)	(-4.57)
Asian	-0.0897***	0.149**	0.202***	-0.146***
	(-4.51)	(2.44)	(4.70)	(-5.59)
Native American	-0.0338	-0.0407	-0.0386	0.0211
	(-1.38)	(-0.56)	(-0.74)	(0.66)
Mixed	-0.0389**	-0.0253	0.0519	0.0197
	(-2.27)	(-0.48)	(1.40)	(0.88)
Other Race	0.0315**	-0.0949**	-0.0953***	0.109***
	(2.22)	(-2.26)	(-3.16)	(5.88)
Middle Eastern	0.0922	0.234	-0.0791	0.0986
	(1.46)	(1.25)	(-0.59)	(1.19)
Age in Years	0.00431***	0.00261***	-0.00700***	0.00664***
	(24.24)	(4.83)	(-18.13)	(28.34)
High School Graduate	0.178***	0.0377	-0.0476	0.122***
	(10.29)	(0.68)	(-1.21)	(5.27)
Four-Year College Degree	0.0751***	0.0163	-0.0296***	0.111***
	(16.02)	(1.17)	(-2.95)	(17.98)
Post-Grad Degree	0.00556	-0.0168	0.0450***	0.0688***
	(0.82)	(-0.84)	(3.14)	(7.77)
Female	-0.108***	0.0640***	0.153***	-0.0915***
	(-26.29)	(5.20)	(17.28)	(-16.88)
Mobility	0.00272***	-0.000251	0.000450	0.000872***
	(14.39)	(-0.45)	(1.11)	(3.51)
Lower Class	-0.0861***	-0.00204	0.0596***	-0.126***
	(-11.39)	(-0.09)	(3.62)	(-12.73)
Upper Class	0.0406***	0.000437	-0.0134	0.124***
	(5.14)	(0.02)	(-0.80)	(11.93)
Year=2012	-0.0111***	0.0874***	-0.0220**	-0.00271
	(-2.62)	(6.92)	(-2.42)	(-0.49)
Constant	0.422***	2.267***	1.818***	-0.0308
	(15.42)	(27.12)	(30.07)	(-0.85)
Observations	32,684	30,29	31,454	32,589
State Fixed Effect	Yes	Yes	Yes	Yes
R^2	0.109	0.013	0.072	0.095

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

political activity and (2) their ability to identify their representative's party. However, we observe a positive and significant relationship between county splitting and their approval of Congress.

With regards to the control variables of interest, the trends are not exactly clear-cut. As Table 5.1 demonstrates, some controls were positively associated with certain outcomes but negatively associated with others. We observe that non-White respondents (with the exception of those who identified their race as "Other") tended to have lower likelihoods to engage in political activity or correctly identify their representative's party. Meanwhile, those same respondents largely had higher reported approval ratings of both their representatives and Congress as a whole. Inversely, we observe that older and more educated respondents tended to have higher rates of a correct identification and political involvement, but lower levels of congressional approval.

Table 5.2 captures this relationship with regards to ZIP code splitting. Whether a respondent's ZIP code was split had a statistically significant relationship at the 1% significance level on two of the four outcomes of interest: (1) the correct identification rate and (2) representative approval ratings. Particularly, we observe that respondents whose ZIP codes were split at the time of the survey were associated with a 1.68% lower correct identification rate and a 0.0437 points lower approval rating of their representative (out of 4). The results with regards to the control variables of interest are mixed, similar to Table 5.1.

The goal of this correlational work is to compare the significance of results between the county and ZIP code models. As such, Table 5.3 compares just the magnitude and t-statistics associated with the splitting coefficient between the county and ZIP code models. This table does not feature any new regressions, but merely compares the primary coefficients of interests in Tables 5.1 and 5.2.

The table allows us to directly compare the relative significance and magnitude of the relationships between county vs. ZIP code splitting on CCES survey responses. For the first outcome of interest (correct party identification rate), both county and ZIP code splitting had a negative and significant relationship. However, ZIP code splitting's

TABLE 5.2: Correlational Regression Results for CCES Survey Results on ZIP Code Splitting

	(1) Correct Rep Party ID	(2) Approval of Rep	(3) Approval of Congress	(4) Political Activity
ZIP Code is Split	-0.0168***	-0.0437***	0.0119	-0.00872
	(-3.72)	(-3.23)	(1.23)	(-1.47)
Black	-0.0664***	0.175***	0.506***	-0.0515***
	(-8.72)	(7.50)	(30.23)	(-5.11)
Hispanic	-0.0858***	0.0116	0.251***	-0.0563***
	(-9.44)	(0.41)	(12.64)	(-4.63)
Asian	-0.0942***	0.153**	0.211***	-0.143***
	(-4.71)	(2.46)	(4.87)	(-5.44)
Native American	-0.0430*	-0.0677	-0.0458	0.0146
	(-1.73)	(-0.90)	(-0.86)	(0.44)
Mixed	-0.0431**	-0.0240	0.0623*	0.0259
	(-2.50)	(-0.45)	(1.66)	(1.14)
Other Race	0.0305**	-0.0952**	-0.0887***	0.107***
	(2.14)	(-2.24)	(-2.92)	(5.72)
Middle Eastern	0.0929	0.223	-0.128	0.129
	(1.45)	(1.17)	(-0.94)	(1.53)
Age in Years	0.00429***	0.00258***	-0.00684***	0.00671***
	(23.85)	(4.72)	(-17.46)	(28.21)
High School Graduate	0.179***	0.0400	-0.0499	0.126***
	(10.23)	(0.71)	(-1.26)	(5.36)
Four-Year College Degree	0.0758***	0.0150	-0.0286***	0.111***
	(16.06)	(1.07)	(-2.82)	(17.76)
Post-Grad Degree	0.00435	-0.0205	0.0413***	0.0671***
	(0.64)	(-1.01)	(2.84)	(7.47)
Female	-0.107***	0.0630***	0.152***	-0.0926***
	(-25.68)	(5.07)	(16.97)	(-16.89)
Mobility	0.00262***	-0.000325	0.000505	0.000845***
	(13.77)	(-0.57)	(1.24)	(3.37)
Lower Class	-0.0898***	-0.00431	0.0659***	-0.133***
	(-11.64)	(-0.18)	(3.91)	(-13.12)
Upper Class	0.0400***	-0.0000628	-0.00926	0.123***
	(5.03)	(-0.00)	(-0.55)	(11.77)
Year=2012	-0.0138***	0.0843***	-0.0231**	-0.00357
	(-3.22)	(6.60)	(-2.52)	(-0.63)
Constant	0.428***	2.280***	1.815***	-0.0394
	(15.56)	(27.06)	(29.74)	(-1.08)
Observations	31,941	29,489	30,747	31,846
State Fixed Effect	Yes	Yes	Yes	Yes
R^2	0.108	0.013	0.072	0.096

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

TABLE 5.3: Correlational Regression Results Comparison Between County and ZIP Code Models

	(1) Correct Rep Party ID	(2) Approval of Rep	(3) Approval of Congress	(4) Political Activity
County is Split	-0.0123***	-0.0204	0.0290***	-0.0188***
	(-2.69)	(-1.50)	(2.95)	(-3.13)
ZIP Code is Split	-0.0168***	-0.0437***	0.0119	-0.00872
	(-3.72)	(-3.23)	(1.23)	(-1.47)

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

association is both larger in magnitude and significance (county splitting is significant only at the 1% level, but ZIP code splitting is significant at the 0.1% level). For the remaining three outcomes, only one geographic splitting indicator is significant for each. ZIP code splitting alone has a negative association with representative approval ratings. County splitting alone has a positive association with congressional approval ratings and a negative relationship with involvement in political activity.

The preliminary results comparing ZIP code and county splitting are therefore mixed. Each indicator of geographic splitting is more significant for different outcomes of interests. Thus, we next utilize a difference-in-difference approach to better identify the true effect of geographic splitting on voters' behavior.

5.1.2 Difference-in-Difference Results

We next move away from Bowen (2014) and Curiel and Steelman's (2018) correlational approach. Instead, we utilize Equation 4.3 above to conduct a difference-in-difference approach. This allows us to see if the act of getting split or unified (our "treatments") was associated with significantly different trends over time. Tables 5.4 and 5.5 below summarize the difference-in-difference approach for county splitting and ZIP code splitting, respectively.

For the county model in Table 5.4, we first look at the coefficient on the "Post Redistricting (2012)" term. We observe that survey responses generally improved from

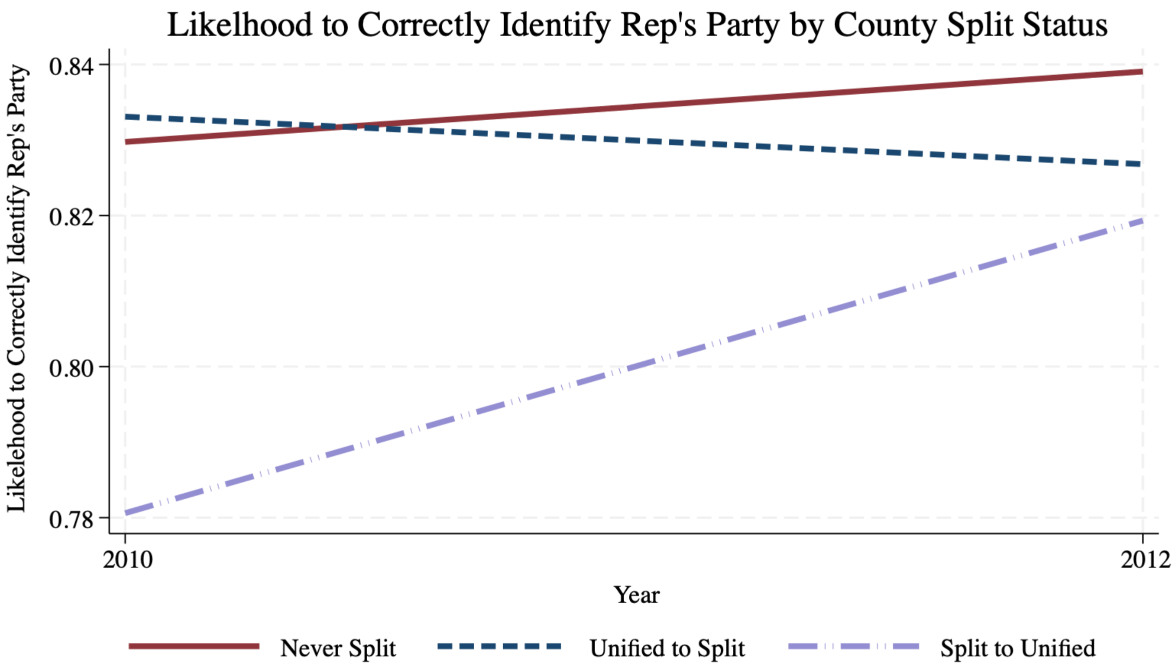
TABLE 5.4: Difference-in-Difference Estimates for the Effect of County Splitting on CCES Responses

	(1) Correct Rep Party ID	(2) Approval of Rep	(3) Approval of Congress	(4) Political Activity
County Always Split	-0.00378	0.0451	0.00292	0.0210
	(-0.18)	(0.63)	(0.05)	(0.81)
County Unified to Split	-0.0172	-0.215	-0.0921	-0.0181
	(-0.48)	(-1.62)	(-0.89)	(-0.40)
County Split to Unified	-0.00257	0.141	-0.0322	-0.00295
	(-0.08)	(1.34)	(-0.36)	(-0.08)
Post-Redistricting (2012)	0.0215***	0.0958***	-0.0372***	0.0388***
	(7.07)	(10.82)	(-4.30)	(10.18)
County Unified to Split x 2012	-0.0227**	-0.0144	0.0276	0.00631
	(-2.11)	(-0.46)	(0.90)	(0.47)
County Split to Unified x 2012	0.0273***	0.00317	0.00325	0.0166
	(2.95)	(0.12)	(0.12)	(1.43)
Constant	0.565***	2.364***	2.000***	0.705***
	(4.43)	(4.92)	(5.63)	(4.40)
Observations	38,297	35,206	36,878	38,170
State Fixed Effect	Yes	Yes	Yes	Yes
Individual Fixed Effect	Yes	Yes	Yes	Yes
R^2	0.011	0.013	0.004	0.012

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

2010 to 2012 (except for congressional approval ratings, which declined). The interaction terms indicate whether respondents who were exposed to the positive or negative treatment (getting unified or split during redistricting) experienced a significantly different change over time. The difference-in-difference estimates are captured in these interaction terms. We do not observe a significant relationship for the latter three results. However, we do observe a significant relationship on the first outcome: the rate at which respondents correctly identified their representative's party affiliation. Particularly, we find that the negative treatment (going from unified to split) is associated with a significant 2.27% reduction in the rate of correct identifications. Inversely, we observe a significant 2.73% increase among residents who experienced the positive treatment (going from split to unified). Figure 5.1 below visualizes these results.

FIGURE 5.1: Respondents' Ability to Identify Representative's Party from 2010 to 2012, Segmented by Their Exposure to County Splitting



As Figure 5.1 demonstrates, untreated respondents (maroon solid line) experienced an improvement in their identification rates from 2010 to 2012. Respondents who underwent the positive treatment (navy dashed line) had a significantly steeper increase in their identification rates; whereas, respondents who underwent the negative treatment (lavender dashed and dotted line) had a statistically significantly shallower (and indeed negative) slope. Given the insignificant terms on the rest of the interaction terms on other outcome variables, we could not rule out parallel slopes between treated and untreated groups for those other outcomes.

We then re-run these models but with regards to ZIP code splitting. These results are summarized in Table 5.5. Like the county model, we generally observe an improvement in survey responses from 2010 to 2012 (with the exception of congressional approval ratings). The interaction terms across all four outcome variables here are insignificant, both for the negative and positive treatment groups. This undermines the previous correlational data that suggested a significant negative impact of ZIP code splitting on survey responses. We therefore cannot conclude that ZIP code splitting (or the reverse) during the 2010-2012 redistricting cycle had a significant impact on individual respondents.

TABLE 5.5: Difference-in-Difference Estimates for the Effect of ZIP Code Splitting on CCES Responses

	(1) Correct Rep Party ID	(2) Approval of Rep	(3) Approval of Congress	(4) Political Activity
ZIP Code Always Split	-0.0243	0.00915	0.0302	0.0153
	(-0.75)	(0.08)	(0.30)	(0.36)
ZIP Code Unified to Split	0.0107	-0.0293	0.0154	0.0307
	(0.27)	(-0.20)	(0.13)	(0.60)
ZIP Code Split to Unified	0.00319	-0.0367	-0.0988	-0.00906
	(0.08)	(-0.27)	(-0.85)	(-0.19)
Post-Redistricting (2012)	0.0295***	0.0965***	-0.0348***	0.0403***
	(9.39)	(10.39)	(-3.79)	(9.95)
ZIP Code Unified to Split x 2012	-0.00940	0.0275	-0.0213	0.00414
	(-1.09)	(1.09)	(-0.85)	(0.37)
ZIP Code Split to Unified x 2012	-0.00864	-0.0317	0.0287	0.00610
	(-1.00)	(-1.22)	(1.14)	(0.55)
Constant	0.599***	2.940	2.375***	0.868***
	(2.71)	(1.56)	(3.77)	(3.04)
Observations	37,460	34,490	36,078	37,333
State Fixed Effect	Yes	Yes	Yes	Yes
Individual Fixed Effect	Yes	Yes	Yes	Yes
R^2	0.014	0.011	0.003	0.011

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

5.2 Changes in District-Level Voting Patterns

In the second part of our analysis, we examine the relationship between geographic splitting and voting patterns at the district level. Unlike in the previous part, we do not look at the number of times a region is split; rather, our independent variable of interest in the following models is how many regions a district splits apart. In this district-level results section, we examine two primary outcomes of interest: (1) district voter turnout and (2) electoral winning margins.

We summarize our results for the county models in Table 5.6. With regards to the demographic features, we observe that older and wealthier districts with a higher percentage of White or Black residents tend to have higher voter turnout rates. Districts with a higher female percentage tend to have lower rates of voter turnout. In column 2, we observe that districts with a higher percentage of Black or female voters have larger winning percentages, but wealthier and Whiter districts have lower winning percentages.[1] Our results do not indicate a statistically significant coefficient for our county splitting terms (neither the linear nor quadratic term). This undermines the school of thought that advocates for preserving county boundaries during redistricting.

TABLE 5.6: Regression Results for District-Level Congressional Election Engagement on County Splitting

	(1) Voter Turnout	(2) Winning Percent
Number of Counties Split by District	-0.00105	-0.000436
	(-0.85)	(-0.20)
Number of Counties Split by District Squared	0.000192	-0.0000502
	(1.62)	(-0.24)
Median Age	0.00610***	-0.00118
	(14.29)	(-1.57)
Percent Black	0.210***	0.118***
	(15.50)	(5.00)
Percent White	0.271***	-0.129***
	(26.56)	(-7.25)
Percent Female	-0.823***	0.533**
	(-5.59)	(2.07)
Per Capita Income (in terms of 10,000 2018 USD)	0.0288***	-0.0112***
	(21.95)	(-4.90)
Constant	0.148**	0.576***
	(2.14)	(4.78)
Observations	3,915	3,914
Year Fixed Effect	Yes	Yes
R^2	0.631	0.149

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

[1] The heightened winning margins among Blacker but not Whiter districts may be a byproduct of majority-minority districts, where certain districts are drawn to give people of color a majority in a particular race. These districts tend to overwhelmingly vote in one party's favor.

We next turn to the results for the same models but with regards to ZIP code splitting. These results are summarized in Table 5.7. We observe similar demographics trends as in the county model. Older and wealthier districts tend to have higher levels of voter turnout; districts with a larger female population again have lower levels of voter turnout. With regards to the winning percentage, Blacker and more female districts have larger winning percentages, and Whiter and wealthier elections have lower winning percentages. All of the demographic trends are consistent across both the county and ZIP code model, both in significance and direction.

TABLE 5.7: Regression Results for District-Level Congressional Election Engagement on ZIP Code Splitting

	(1) Voter Turnout	(2) Winning Percent
Number of ZIP Codes Split by District	-0.00103***	0.000160
	(-5.14)	(0.46)
Number of ZIP Codes Split by District Squared	0.0000109***	0.000000966
	(4.25)	(0.21)
Median Age	0.00595***	-0.00123
	(13.92)	(-1.64)
Percent Black	0.231***	0.104***
	(16.85)	(4.35)
Percent White	0.287***	-0.141***
	(27.25)	(-7.66)
Percent Female	-0.943***	0.602**
	(-6.38)	(2.32)
Per Capita Income (in terms of 10,000 2018 USD)	0.0282***	-0.0107***
	(21.52)	(-4.65)
Constant	0.218***	0.544***
	(3.13)	(4.46)
Observations	3,915	3,914
Year Fixed Effect	Yes	Yes
R^2	0.634	0.150

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

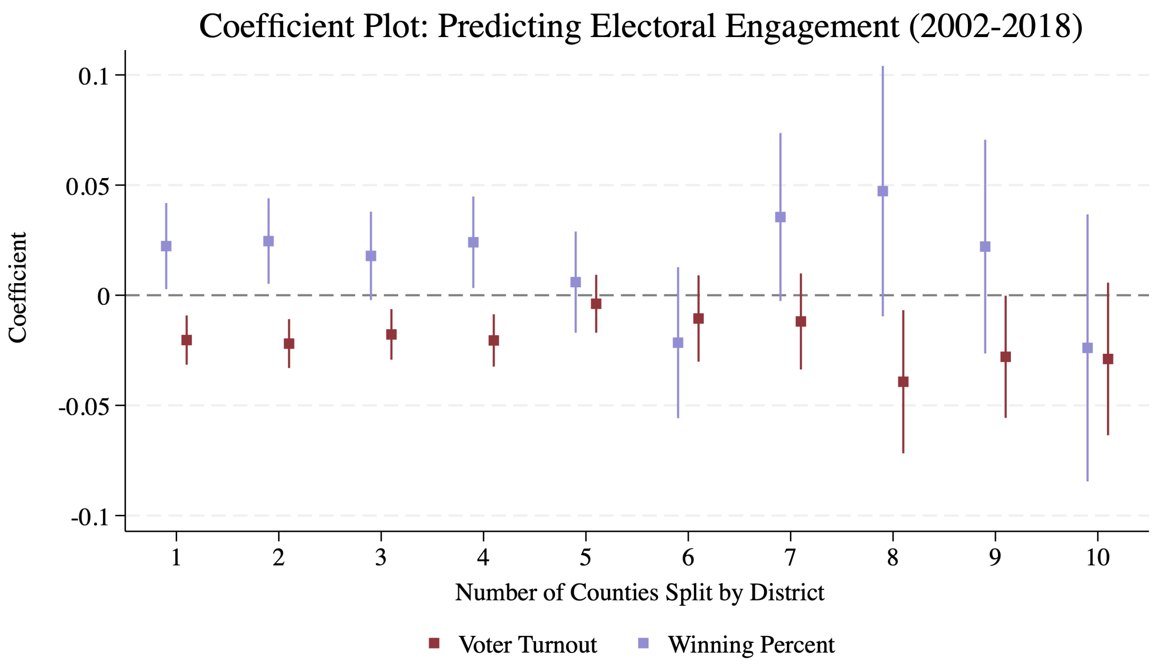
Unlike the county model, we do observe some significant results for our ZIP code terms. The first model in Table 5.7 predicts a significant quadratic relationship for the number of ZIP codes a district splits through. The significantly negative linear term and positive quadratic term suggests that more ZIP code splitting reduces voter turnout at low levels of geographic splitting; however, this relationship eventually rebounds at higher levels of ZIP code splitting. This non-linear relationship is further explored in Section 5.2.1 of this paper. Given that the ZIP code splitting was significantly predictive of congressional election turnout, this would bolster the school of thought that advocates for not splitting ZIP codes in redistricting. Turning to the second outcome (results for winning percentage), we do not observe a significant coefficient on either the linear nor quadratic terms. Across both the county and ZIP code model, we observe that the R^2 values for predicting winning percentages are much lower than the R^2 values for predicting voter turnout.

5.2.1 Further Exploring Potential Non-Linear Relationships

Given the non-linear relationship we previously observed in Figure 4.5 and Appendix A, we further explore whether the extent to which a district splits counties or ZIP codes influences electoral results across certain ranges. That is, we explore the potential non-linear electoral returns to county and ZIP code splitting. Here, we compare the electoral results (voter turnout and winning margin) of districts comprised of a set number of ZIP code or county splits to a baseline group. For the county level of analysis, we compare districts to baseline districts that split no counties. Given the larger range of ZIP code splitting, we group districts in five-unit increments. Our baseline group for the ZIP code level of analysis is therefore districts that split only 0-4 ZIP codes.

Figure 5.2 illustrates the electoral engagement of districts as a function of their county splitting relative to districts who split no counties, controlling for all aforementioned demographic features. The y-axis describes the predicted coefficient of Equation 4.4 when comparing just districts of a given count split count to those that split no counties. We observe generally significant relationships at low levels of county splitting, but not at high levels of county splitting. At low levels of county splitting, districts experience lower levels of voter turnout compared to our baseline (districts that split

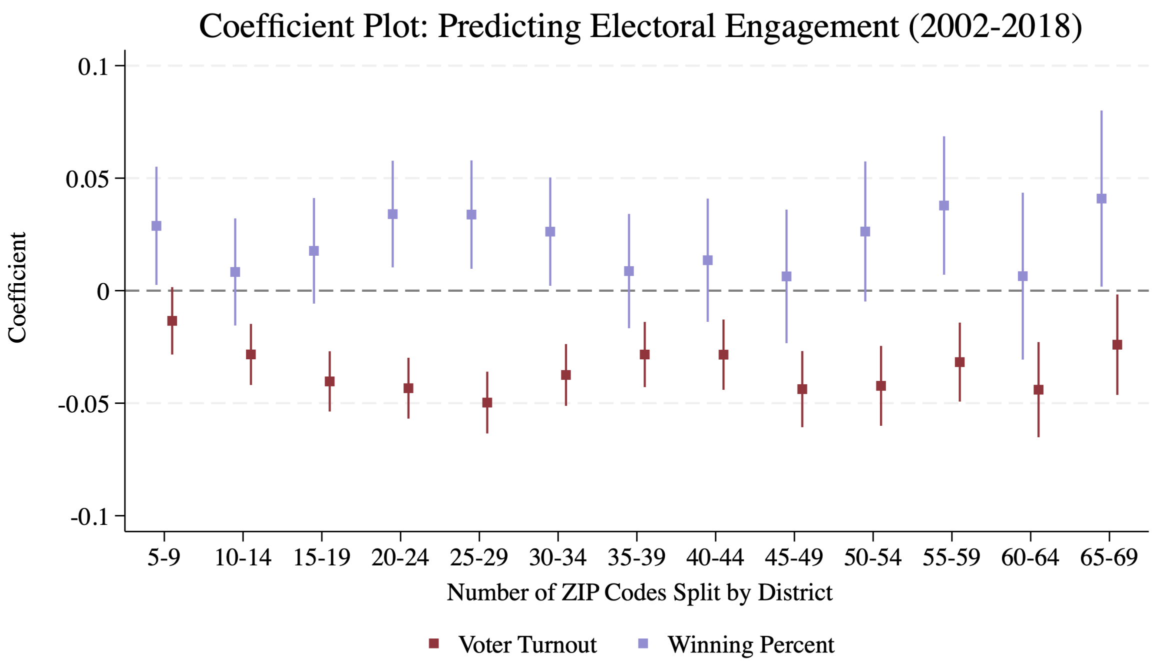
FIGURE 5.2: Regression Result Summaries Comparing Electoral Engagement in Districts of Varying County Split Counts to Districts that Split No Counties



no counties whatsoever). Inversely, we observe that districts with low levels of county splitting have significantly higher margins of victory as compared to districts with no splitting whatsoever. This is an indication that county splitting (at least at low levels) can reduce voter turnout and make elections less competitive. We theorize that the insignificant results at higher levels of county splitting may in part be due to the smaller sample size of districts at such high levels of county splitting.

Figure 5.3 demonstrates an analogous plot with regards to how many ZIP codes a district splits. As previously mentioned, we group districts into ranges of five ZIP code splits. Figure 5.3 therefore plots the predicted regression coefficients of Equation 4.5 when comparing districts of a given ZIP code split count to a district that only splits 0-4 ZIP codes. Compared to districts with only 0-4 ZIP code splits, we observe that almost all districts with ZIP code splits experience a significantly lower voter turnout. With regards to a district's winning percentage, generally districts in the middle tier of the ZIP code split distribution (20-34 splits) have significantly higher winning percentages

FIGURE 5.3: Regression Result Summaries Comparing Electoral Engagement in Districts of Varying ZIP Code Split Counts to Districts that Split 0-4 ZIP Codes



and therefore less competitive elections.

The difference in the shape of these relationships between the ZIP code and county model is interesting. We theorize that the county model is significant at lower levels of splitting because counties are much more populous than ZIP codes. Because ZIP codes are less populous, splitting just a handful of ZIP codes may not affect enough constituents to influence district-level outcomes. By sheer size, splitting just one populous county affects more citizens than splitting several ZIP codes. This may explain why more ZIP code splitting is necessary for districts to demonstrate altered winning margins. Across both the county and ZIP code models, splitting reduces competitiveness (by increasing winning margins) only across select ranges; however, ZIP code splitting is unambiguously worse for voter turnout. Voter turnout is reduced for districts that split ZIP codes across the board when compared to our baseline group.

6 Discussion and Limitations

This section discusses the potential implications and interpretations of the previously outlined results. From there, we explore some of the potential limitations of the aforementioned work.

The goal of this work is to compare the relative democratic harms of geographic splitting with respect to counties and ZIP codes during congressional redistricting. Our results are mixed. The first part of our analysis uses survey responses of individual constituents as the primary outcome of interest. We begin with a purely correlational approach, akin to Curiel and Steelman (2018) and Bowen (2014). We find that ZIP code splitting had a more significantly negative relationship with respondents' ability to identify their representative's party affiliation and their reported approval of said representative. Meanwhile, we demonstrate that county splitting is more significantly correlated with an increase in approval ratings of Congress and a decrease in respondents' involvement in political activity. Given just the correlational results, ZIP code splitting and county splitting are each more relevant for different outcomes of interest.

We do observe the limitations with such correlational approach. Namely, the aforementioned concerns with endogeneity (that mapmakers select areas to be split based upon constituents' preexisting behaviors). As such, we turn to the difference-in-difference estimates. Neither county nor ZIP code splitting was found to have a significant impact on constituents' approval ratings or likelihood to engage in political activity. In previous work by Curiel and Steelman (2018) and Bowen (2014), researchers had not looked into these outcomes as potential variables of interest. However, both sets of researchers did examine respondents' likelihood to correctly identify their congressional representatives' party affiliation based on similar survey results. Our results found a significant relationship between county splitting and the likelihood of a correct identification: splitting a county (negative treatment) has a negative impact and reunifying a county (positive treatment) has a positive impact. These results are consistent with the county work of Bowen (2014) who similarly found a significant relationship between county splitting and constituents' ability to identify their representative's party. We did not find any significant impact with regards to ZIP code splitting, unlike Curiel and Steelman (2018). This discrepancy in results is potentially due to two factors: (1) Curiel and Steelman (2018) studied a broader timeframe and (2) Curiel and Steelman (2018) employed a purely correlational approach, not a difference-in-difference model. With regards to individual constituent survey behavior across all outcomes of interest, we find evidence that county splitting likely has significant detrimental impacts, but ZIP code splitting does not.

Our research then explored the potential implications of geographic splitting with regards to district-level electoral results, namely voter turnout and electoral competitiveness (measured via winning percentage). With regards to electoral competitiveness, county splitting is associated with increased margins of victory only at a lower range, while ZIP code splitting is associated with an increase in the mid-tier range. Our models find that more geographic splitting with regards to ZIP codes reduces voter turnout across the entire range; however, county splitting is only associated with a reduction in voter turnout across a select range. ZIP code splitting is unambiguously worse for electoral turnout, but the results with regards to electoral competitiveness are more nuanced.

Returning to the motivation of this paper: what geographic boundary (counties or ZIP codes) should mapmakers prioritize in preserving more when redistricting? Our research suggests that the answer is not black-and-white. The results depend on what democratic objectives take precedence. If mapmakers and the courts focus on the harm done to citizens via the electoral process itself, then ZIP code splitting appears to have more significantly detrimental impacts to voter turnout and competitiveness than county splitting. If policymakers focus on the harm done to citizens via their engagement and knowledge, county splitting has a significantly detrimental effect on constituent knowledge of their representative, but ZIP code splitting does not. Both of these approaches are in line with the Roberts Court's approach to redistricting where they focus on the harm done to the citizenry, not political parties. However, it is not the purpose of this paper to address which outcomes policymakers should value most in the redistricting process: electoral engagement or constituent knowledge. That is a decision left up to policymakers' preferences and is therefore outside of the scope of this paper.

Our approach and modeling are not without its limitations. For one, we had to impute the lion's share of respondents' ZIP code information based upon their provided county data when ZIP code data was limited. As a result, we test the robustness of their imputation by rerunning our difference-in-difference models (Equation 4.3) on just the select sample of respondents who provided ZIP code data in both timeframes (reducing our sample from over 18,000 respondents to just 3,829). The majority of our regressions remain unchanged (the ZIP code models with or without imputation are insignificant). However, two of our models do change when we exclude imputed ZIP code values. These results are summarized in Appendix C. With regards to respondents' ability to identify their representatives, our results with the selected sample indicate a significantly negative effect of ZIP codes going from split to unified (positive treatment) on respondents' ability to identify their representatives. In other words, respondents whose ZIP codes became unified from 2010 to 2012 had a significantly lower chance of a correct party identification. This result only reinforces our previous conclusions where we asserted that county splitting is more detrimental to constituent knowledge than ZIP code splitting. These results indicate that unifying ZIP codes is harmful (not beneficial as hypothesized) to citizen's knowledge.

Our second model that changes is the one where we predict respondents' likelihood to engage in political activity. Here, some ZIP code covariates are now significant.

Our restricted model (with only respondents who did not require imputation for their ZIP codes) predicts a significant increase in respondents' likelihood to engage in political activity if their ZIP code went from split to unified from 2010 to 2012 (positive treatment). The results were previously insignificant with the full sample (including imputed ZIP codes). This now-significant result bolsters the claim that ZIP code splitting is more influential than county splitting for individual constituents' political involvement. We do note that with this restricted model with only non-imputed ZIP codes, we only observe the positive impact of ZIP code unification. We do not observe a negative impact to ZIP code splitting. Additionally, the results we detail here for ZIP code splitting among non-inputted values are only significant at the 5% level, not at the 1% level. Our county level results detailed above are more significant (significant at the 1% level).

A second major limitation is that our survey data is only available in two-time frames: 2010 and 2012. This presents a significant limitation for several reasons. First, it is possible that two years is not enough time for voters to internalize the negative impacts of geographic splitting. It is possible that the impacts of geographic splitting are not realized by constituents until several election cycles after redistricting finalizes. Second, we are comparing voter behavior in a midterm year (2010) to a presidential year (2012). While we control for this in our regression, the impact of a presidential election may morph the relationship between geographic splitting and voter behavior. Gerrymandering and redistricting only affect the boundaries for congressional elections, not presidential ones. Thus, redistricting and its impacts may be less salient during a presidential year where congressional election coverage is overshadowed by a presidential race. This limited sample size is in part the motivation for the second part of our analysis where we look at district-level electoral outcomes from 2002 to 2018.

7 Conclusion

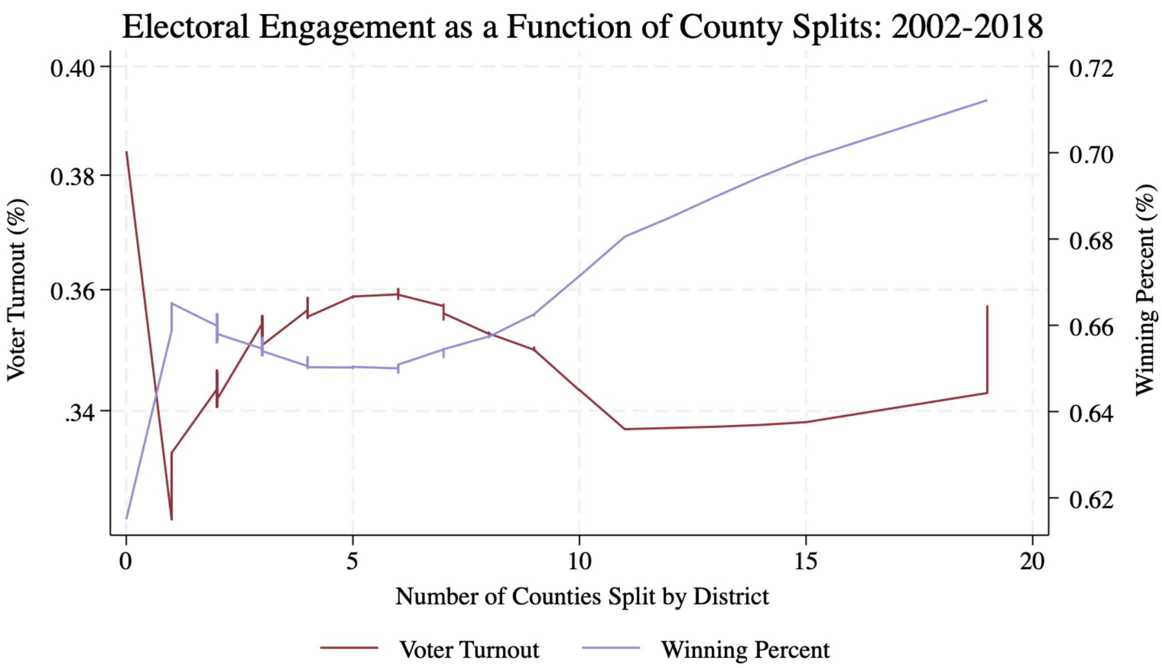
The Supreme Court's current jurisprudence on gerrymandering necessitates a change in academia's approach to the subject. The Court continues to deny legal challenges that rely on previous metrics that quantify the extent to which congressional maps yield proportional results. Instead, the Court emphasizes the need for a metric that captures harm done to constituents, not political parties as a whole. In turn, researchers like Bowen (2014) and Curiel and Steelman (2018) suggest that congressional maps avoid splitting natural geographic boundaries to the smallest extent possible. Our research expands upon their previous work by (1) employing a difference-in-difference strategy instead of a purely correlational approach, and (2) by comparing the differential impact of ZIP code splitting versus county splitting. The question of whether ZIP code or county splitting is more deleterious is an especially important one, given that maps cannot simultaneously preserve ZIP codes and counties due to their intersecting boundaries.

Our results indicate that the relevance of ZIP code versus county splitting for individual constituents depends on what primary outcome of interest is more important to policymakers. County splitting is associated with a significant reduction in constituents' knowledge of their representative, but it is not significantly detrimental to electoral turnout. On the other hand, ZIP code splitting is associated with a significant reduction in electoral turnout, but we find no significant impact on constituent knowledge. It is up to policymakers to debate which impacts society should value more. The extent to which we value constituent knowledge versus electoral turnout motivates whether county splitting or ZIP code splitting is more detrimental. Future research may incorporate tenants of political economy to embed policymakers' preferences into a complete model that compares the relative strengths and weakness to preserving counties and ZIP codes. Alternatively, future work may expand upon our difference-in-difference approach to examine how voters' knowledge and sentiments fluctuate over a period of more than two elections as a function of their geographic splitting status.

If the last two decades are any indication of trends to come, gerrymandering will continue to be a prevalent tool wielded by both major American political parties. The prospect of gerrymandering runs antithetical to core democrat tenants of majority rule and accountability. Nonetheless, American jurisprudence requires an objective and implementable standard for determining when a congressional map becomes unconstitutional. The extent to which a congressional district splits natural geographic boundaries offers a promising launchpad for adjudicating whether a given map harms the democratic process and individual citizens. The regular redrawing of arbitrary congressional lines on top of preexisting geographic boundaries provides an ideal natural experiment for such econometric research.

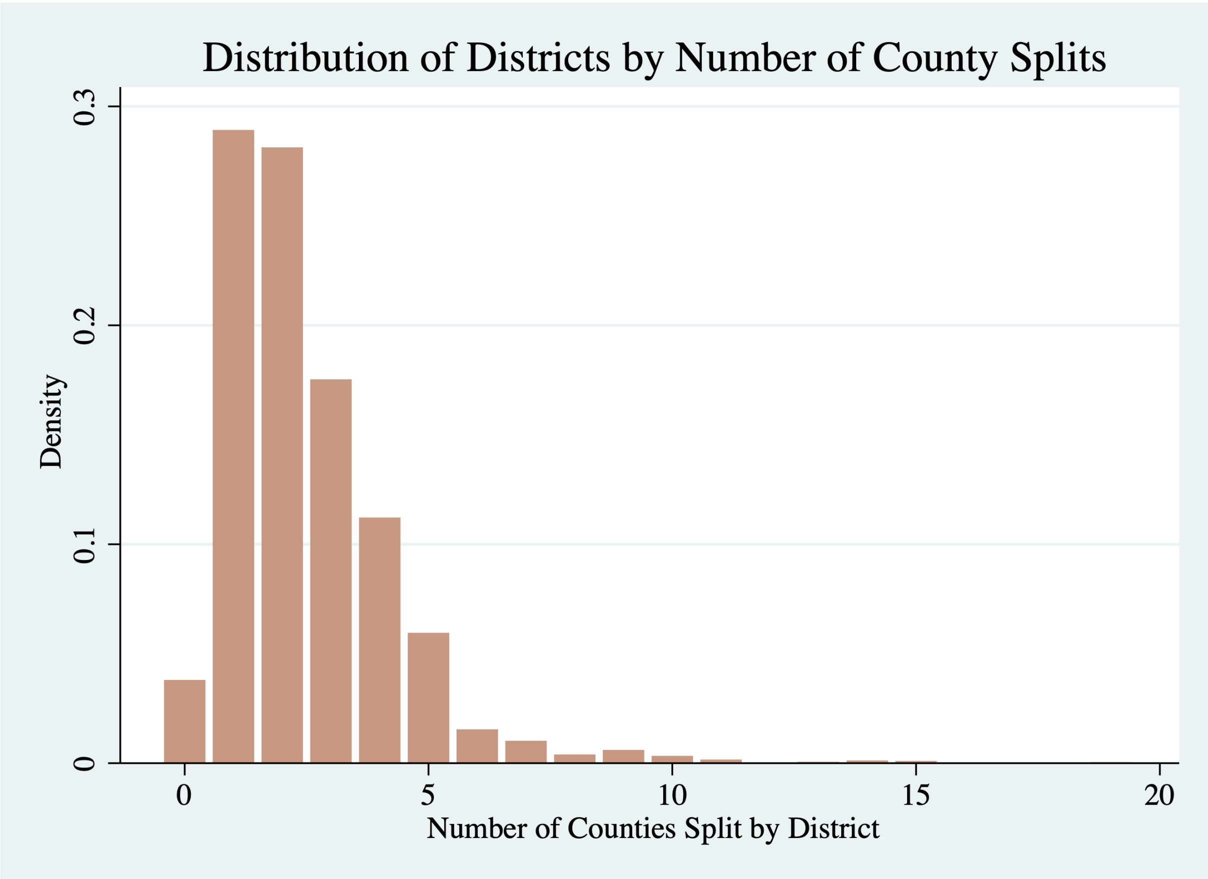
A Appendix A

FIGURE A.1: Smoothed Scatter Plot of Electoral Engagement as a Function of the Number of Counties Split by District



B Appendix B

FIGURE B.1: Histogram of Number of Counties Split by District



C Appendix C

TABLE C.1: Difference-in-Difference Estimates for the Effect of ZIP Code Splitting on CCES Responses, Excluding Imputed ZIP Code Entries

	(1) Correct Rep Party ID	(2) Political Activity
ZIP Code Was Always Split from 2010 to 2012	-0.0234	0.0149
	(-0.61)	(0.35)
ZIP Code Went from Split to Unified from 2010 to 2012	0.0243	-0.0316
	(0.53)	(-0.63)
ZIP Code Went from Unified to Split from 2010 to 2012	0.00700	0.0442
	(0.15)	(0.84)
Responses Collected in 2012 Survey	0.0437***	0.0452***
	(5.39)	(5.02)
ZIP Code Went from Split to Unified x 2012	-0.0495**	0.0516**
	(-2.16)	(2.04)
ZIP Code Went from Unified to Split x 2012	0.00321	-0.0261
	(0.13)	(-0.96)
Constant	0.564**	0.854***
	(2.17)	(2.97)
Observations	7,525	7,525
Individual Fixed Effect	Yes	Yes
Year Fixed Effect	Yes	Yes
R^2	0.038	0.030

t statistics in parentheses
* $p < .10$, ** $p < .05$, *** $p < .01$

References

Ansolabehere, S., Brady, D., & Fiorina, M. (1992). The vanishing marginals and electoral responsiveness. *British Journal of Political Science*, *22*(1), 21-38. https://doi.org/10.1017/S0007123400000338.

Ansolabehere, Stephen; Schaffner, Brian, 2014, "2010 - 2012 CCES Panel Study," https://doi.org/10.7910/DVN/24416, Harvard Dataverse, V4, UNF:6:VJa/Am9qCmzeEW8dH72wUQ==[fileUNF].

Arrington, T. S. (2010). Redistricting in the U.S.: A review of scholarship and plan for future research. *The Forum: A Journal of Applied Research in Contemporary Politics*, *8*(2), 14 15. https://doi.org/10.2202/1540-8884.1351.

Bernstein, M., & Duchin, M. (2017). A formula goes to court: Partisan gerrymandering and the efficiency gap. *Notices of the American Mathematical Society*, *64*(9), 1020-1024. https://doi.org/10.1090/noti1573.

Bowen, D. C., & Clark, C. J. (2014). Revisiting descriptive representation in congress: Assessing the effect of race on the Constituent–Legislator relationship. *Political Research Quarterly*, *67*(3), 695-707. https://doi.org/10.1177/1065912914531658.

Bowen, D. C. (2014). Boundaries, redistricting criteria, and representation in the U.S. house of representatives. *American Politics Research*, *42*(5), 856-895. https://doi.org/10.1177/1532673X13519127.

Cain, B. E., Ferejohn, J. A., & Fiorina, M. P. (1987). The personal vote: Constituency service and electoral independence.

Congressional Research Service. (2021, November 15). *Congressional Redistricting Criteria and Considerations*. CRS Insight. https://crsreports.congress.gov/product/pdf/IN/IN11618.

Cover, B. P. (2018). Quantifying partisan gerrymandering: An evaluation of the efficiency gap proposal. *Stanford Law Review*, *70*(4), 1131-1233.

Curiel, J. A., & Steelman, T. (2018). Redistricting out representation: Democratic harms in splitting zip codes. *Election Law Journal*, *17*(4), 328-353. https://doi.org/10.1089/elj.2018.0528.

Curiel, J. A., & Steelman, T. (2020). A response to "Tests for unconstitutional partisan gerrymandering in a post-Gill world" in a Post-Rucho world. *Election Law Journal*, *19*(1), 101-109. https://doi.org/10.1089/elj.2019.0612.

DeFord, D. R., Eubank, N., & Rodden, J. (2022). Partisan dislocation: A precinct-level measure of representation and gerrymandering. *Political Analysis*, *30*(3), 403-425. https://doi.org/10.1017/pan.2021.13.

Engstrom, R. L. (2020). Partisan gerrymandering: Weeds in the political thicket. *Social Science Quarterly*, *101*(1), 23-36. https://doi.org/10.1111/ssqu.12743.

Gill v. Whitford, 585 U.S. ___ (2018). https://www.oyez.org/cases/2017/16-1161.

Grofman, B., & Cervas, J. (2021). ZIP codes as geographic bases of representation. *Election Law Journal*, *20*(2), 215-223. https://doi.org/10.1089/elj.2020.0687.

Hyer, V. (2021, April 26). *2020 Census Apportionment Results Delivered to the President*. The United States Census Bureau. https://www.census.gov/newsroom/pressreleases/2021/2020-census-apportionment-results.html.

Jackson, R. A. (2000). Differential Influences on Participation in Midterm Versus Presidential Elections. *The Social Science Journal*, *37*(3), 385. https://link-gale-com.proxy.lib.duke.edu/apps/doc/A67320157/EAIM?u=duke_perkins&sid=summon&xid=19326409.

Jeong, D., & Shenoy, A. (2022). The targeting and impact of partisan gerrymander ing: Evidence from a legislative discontinuity. *The Review of Economics and Statistics, 1-47*. https://doi.org/10.1162/rest_a_01204.

Jones, D., Silveus, N., & Urban, C. (2023). Partisan gerrymandering and turnout. *The Journal of Law & Economics*, *66*(3), 557-579. https://doi.org/10.1086/725767.

Kulachai, W., Lerdtomornsakul, U., & Homyamyen, P. (2023). Factors influencing voting decision: A comprehensive literature review. *Social Sciences (Basel)*, *12*(9), 469. https://doi.org/10.3390/socsci12090469.

Kromkowski, Charles A. and John A. Kromkowski. "Why 435? A Question of Political Arithmetic." *Polity* 24, no. 1 (1991): 129-145.

Legislative Service Commission. (2021). *Article XIX, Section 2 Requirements for drawing congressional districts*. Ohio Laws & Administrative Rules. https://codes.ohio.gov/ohio-constitution/section-19.2#:~:text=(5)%20Of%20the%20eighty%2D,which%20counties%20may%20be%20split.

Li, M., Limón, G., & Kirschenbaum, J. (2021, September 16). Who Draws the Maps? *Brennan Center for Justice*. https://www.brennancenter.org/our-work/research-reports/who-draws-maps-0.

Makse, T. (2012). Defining communities of interest in redistricting through initiative voting. *Election Law Journal*, *11*(4), 503-517. https://doi.org/10.1089/elj.2011.0144.

MIT Election Data and Science Lab, 2024, "U.S. House 1976–2022", https://doi.org/10.7910/DVN/IG0UN2, Harvard Dataverse, V12, UNF:6:A6RSZvlhh8eRZ4+mvT/HRQ== [fileUNF].

National Conference of State Legislatures. (2021, July 16). *Redistricting Criteria*. NCSL. https://www.ncsl.org/redistricting-and-census/redistricting-criteria.

Redistricting Data Hub. (2023, March 20). *118th congressional districts and Block Assignment File (Nov. 2022 election, active Jan. 2023 - Jan. 2025)*. Redistricting Data Hub. https://redistrictingdatahub.org/dataset/118th-congressional-districts-nov-2022-election-active-jan-2023-jan-2025/.

Rossiter, K. M., Wong, D. W. S., & Delamater, P. L. (2018). Congressional redistricting: Keeping communities together? *The Professional Geographer, 70*(4), 609-623. https://doi.org/10.1080/00330124.2018.1443477.

Sabouni, H., & Shelton, C. (2021). State legislative redistricting: The effectiveness of traditional districting principles in the 2010 wave. *Election Law Journal, 20*(2), 198-214. https://doi.org/10.1089/elj.2019.0608.

SAS Institute Inc. (2014, April 14). *County and ZIP code boundaries not matching*. SAS Communities. https://communities.sas.com/t5/Graphics-Programming/County-and-zip-code-boundaries-not-matching/td-p/177083.

Schaffner, Brian; Ansolabehere, Stephen; Shih, Marissa, 2023, "Cooperative Election Study Common Content, 2022", https://doi.org/10.7910/DVN/PR4L8P, Harvard Dataverse, V3.

Stephanopoulos, N. O., & McGhee, E. M. (2015). Partisan gerrymandering and the efficiency gap. *The University of Chicago Law Review, 82*(2), 831-900.

Stephanopoulos, N. O., & Warshaw, C. (2020). The impact of partisan gerrymandering on political parties. *Legislative Studies Quarterly, 45*(4), 609-643. https://doi.org/10.1111/lsq.12276.

Steven Manson, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 18.0 [dataset]. Minneapolis, MN: IPUMS. 2023. http://doi.org/10.18128/D050.V18.0.

The University of Missouri. (2022). *Geocorr 2022: Geographic Correspondence Engine*. The Missouri Census Data Center. https://mcdc.missouri.edu/applications/.geocorr2022.html2.

Rakich, N. (2022, June 15). *The New National Congressional Map is Biased Toward Republicans*. FiveThirtyEight. https://fivethirtyeight.com/features/the-new-national-congressional-map-is-biased-toward-republicans/.

Rossiter, K. (2014, July 31). *Understanding Geographic Relationships: Counties, Places, Tracts and More*. The United States Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2014/07/understanding-geographic-relationships-counties-places-tracts-and-more.html.

Rush, M. (2020). The efficiency gap after *Gill v. Whitford*. *Social Science Quarterly, 101*(1), 53-67. https://doi.org/10.1111/ssqu.12742.

United States Census Bureau. (2023). *County Adjacency File*. United States Census Bureau. https://www2.census.gov/geo/docs/reference/county_adjacency.txt.

Wang, S. S. (2016). Three practical tests for gerrymandering: Application to Maryland and Wisconsin. *Election Law Journal*, *15*(4), 367-384. https://doi.org/10.1089/elj.2016.0387.

Warrington, G. S. (2018). Quantifying gerrymandering using the vote distribution. *Election Law Journal*, *17*(1), 39-57. https://doi.org/10.1089/elj.2017.0447.

Winburn, J., & Wagner, M. W. (2010). Carving voters out: Redistricting's influence on political information, turnout, and voting behavior. *Political Research Quarterly*, *63*(2), 373-386. https://doi.org/10.1177/1065912908330728.

Wines, M. (2022, February 18). *In Nashville, a Gerrymander Goes Beyond Politics to the City's Core*. The New York Times. https://www.nytimes.com/2022/02/18/us/nashville-gerrymandering-republican-democrat.html.

WP Company. (2022, February 5). *Opinion: This is gerrymandering at its worst. It doesn't have to be this way*. The Washington Post. https://www.washingtonpost.com/opinions/interactive/2022/gerrymandering-examples-north-carolina-illinois-alabama-texas-how-to-fix/.