Erasmus University Rotterdam - Erasmus School of Economics

Bachelor Thesis International Bachelor Economics and Business Economics

Did the Green Party decide the United States presidential election of 2016?

Author: Jurian van der Bijl (Student number: 415265)

Thesis supervisor: A.J. Hummel

Second assessor: B.S.Y. Crutzen

Final version handed in on September 17, 2017

Abstract

This paper models what voters do in the absence of their preferred party. In this context, some hypotheses are formulated with respect to the role of Green Party voters during the 2016 US presidential elections. To test these theoretical predictions, voting data at county level is analyzed with a Fixed Effects OLS model. Utilizing state differences in party's ballot access, state borders are used as a source of exogenous variation. Provided that voter preferences are single peaked and that the Democratic Party and Green Party are substitutes, we find that the presidential elections could have seen a different outcome had the Green Party not been on the ballot in just three states.

1 Introduction

In November 2016, the 58th US presidential elections resulted in an outcome that not many pundits had expected. The Republican Trump-Pence ticket won the presidency, acquiring 304 of the 538 electoral votes, and took office in January 2017. At the same time, the Democratic Clinton-Kaine duo won the popular vote. However, as the composition of the Electoral College is determined by a winner-takes-it-all procedure for each state, the Democratic ticket did not bring home the bacon.

In the US political spectrum - which historically has been a bipartisan system - there is a small role for some other political parties. The two most important ones are the culturally progressive but economically conservative Libertarian Party and the progressive-liberal Green Party. The Libertarian candidate, Gary Johnson, received 3.28% of the popular vote, whereas the Green Party's Jill Stein won 1.07%. Naturally, both the Republican Party and the Democratic Party ran in all states and the district of Washington DC. The same was true for the Libertarian Party, whose ballot access percentage was at 100% as well. The focus of this paper will be on the Green Party, which accessed the ballot of only 89% of voters.

The 2016 election was quite close. In Michigan, Wisconsin and Pennsylvania, the difference between the voter shares of Trump and Clinton was respectively 0.3, 0.8 and 1.2 percentage points. At the time, these states were worth 46 electoral votes in total. Hence, if all three states had seen a slightly different result, the outcome of the election would have been flipped.

In the context of close election results like this, it is very interesting to evaluate the role of the small parties in the voting decisions of citizens. In Michigan, for instance, the Green Party attained 1.1% of the vote. As this party appeals to a similar kind of voters as the Democratic Party does, the role of the Green Party might have been crucially decisive in the election's end result. This notion prompts to the question what would have happened when the Green Party would not have had ballot access in these three states.

The main issues we aim to answer with respect to the Green Party in this thesis are the following. The first goal is establishing what rational voter theory predicts that voters do in the absence of their most favorable party. After this, we empirically test these predictions on the 2016 voting data. This is done using the assumption that political preferences are clustered around a geographical determinant. Under this condition, Green Party ballot presence in only one of two bordering states, gives rise to variation, which we argue is exogenous for two small areas on either side of the border.

We find strong and significant negative effects of Green Party ballot presence on the voter shares of the Democratic Party. The results indicate that the Green Party had a decisive role in the of the election, rendering the Electoral College vote difference between Trump and Clinton close to arbitrary.

The paper is organized as follows. Section 2 provides a review of the relevant academic literature. Section 3 introduces the elements of our voting model and justifies its assumptions. Furthermore, it makes some predictions based on this model. In section 4, the data and the empirical strategy are accounted for. Section 5 and 6 respectively present the results and establish the results' robustness. The 7th section concludes.

2 Review of theoretical literature

2.1 Rational behavior

In the fields of rational choice theory and more recently positive political theory the *fundamentum absolutum inconcussum* is the assumption that the preferences which people have are generalisible and adhere to some logically sound range of propositions. In the Downsian literature, this assumption of rationality implies that people view elections as a means of electing government and selecting policy efficiently, but (mostly for post-Downsian theorists) are not only affected by materialistic motives. Thus, it is important to note that theorists who use the concept of rationality do not use this in the psychological sense (i.e. that people unemotionally asses their risks and opportunities to determine their course of action), but in a logical and generalisible way. If science is about observing regularities, rationality is fundamental to generalize behavior. In accordance with the academic literature on choice theory, we will assume voters to be rational in a similar fashion.

2.2 Fundamental models of rational decision making

When analyzing problems that traditionally belong in the field of political science, such as voting behavior or elections, simple economic models can provide valuable insights. Most of the theoretical work on public choice theory consists of adaptions on relatively straightforward models of decision making. When trying to understand the motives and options of voters, a study of these public choice models and their assumptions is beneficial in order to understand the actions of voters in a given election.

One of the most elementary models which describes majoritarian decision making and which basic intuition has always been the core of voter theory, is the median voter model. Downs (1957b), building on the work by Hotelling (1929), Smithies (1941) and Black (1948), modeled the policy positions of political candidates or parties on a one-dimensional Euclidean space. Stating that "parties formulate policies to win elections, rather than win elections in order to formulate policies", Downs (1957a, p. 28) argues that electoral competitors would design their policy platforms as to maximize the share of votes, culminating in Downs' median voter theorem.

If it holds that all voting positions fit a one-dimensional continuum and that voters have single-peaked preferences Downs spatial model has an equilibrium. In order to maximize the utility of the median voter, the party platforms will converge to the center of the spectrum. Hence, in equilibrium both parties will have selected a similar platform and will receive an equal share of the votes. Note that here it does not matter which party is ultimately elected into office: given that the newly elected officeholder comes through on his campaign promises the median voter will receive exactly what he wanted in the first place. This is what is known as the 'strong form of the median voter theorem'. The weak form of this theorem predicts that the median voter will always choose his preferred policy - but not necessarily get it (Congleton, 2003).

The model from which Downs deduced his median voter theorem assumes among other things that voters are able to analyze candidates' policy platforms into detail, mapping them as points on a line. The preferences that voters then form are assumed to be single peaked, meaning that the utility function of each voter has a maximum somewhere on that line and moves away from that point on either side with a downward slope. Moreover, the distribution of the voters and their preferences is modeled to fit a normal distribution (uni-modal) and all those eligible of voting do so (Downs, 1957a; Rowley, 1984).

The spatial representation of elections as described above is based on some assumptions on the behavior of voters. The strongest assumption with respect to realism is the supposition that all voters do indeed vote. In a world without mandatory voting it is a fact that a share of the electorate abstains. What is more, Downs makes no serious effort of explaining why individuals who he assumes to be rational and utility maximizing even vote in the first place. Using the work by Riker & Ordeshook and Davis, Hinich & Ordeshook (1968, 1973; 1970), who tried to mathematically approach the majoritarian electoral process, we establish a framework that allows us to understand the components of a voter's actions.

3 Theory

3.1 Voter preferences and candidate positions

Who a citizen's preferred candidate is, depends on his preferences and his evaluation of the platforms candidates are running on. In order to arrive at general outcome, Downs assumes that a voter is interested in a single issue. A citizen then compares how the ideological positions of the candidates relate to his own stance on the matter (Downs, 1957a). However, elections are often about multiple domains of policy. For instance, a candidate might have a conservative stance towards issues such as taxation and gun ownership rights, but a more liberal viewpoint on capital punishment and same-sex marriage[1]. This variety of stances makes for a very differentiated political spectrum which cannot be captured on a one dimensional continuum. To circumvent this problem, we expand the framework by Riker and Ordeshook (1968) and represent the position of one's preferred position by υ_i. This is a set of i's preferred positions for all convictions i looks for in a candidate. This ideal position υ_i can be denoted by a vector containing i's viewpoints for the n relevant dimensions as in equation (1). Consequently, citizens compare this preferred set of political convictions to the position that each candidate has in his perception. This perception of candidate j's position θ_j is assumed to be equal for all civilians and is captured by a vector as displayed in equation (2). As a result of this simplification, we can capture preferences and positions on a one dimensional left to right scale while still being able to maintain the notion that for instance a 'left-leaning' party can have some 'right-leaning' stances.

[1] This was especially the case for Gary Johnson (Libertarian Party) and Jill Stein (Green Party) in the 2016 US presidential elections. For some issues their platforms were explicitly to the left and for other issues their stances stood to the right of the platform by the major party (GOP for Johnson & Democrats for Stein) that their respective bases are associated with ideologically.

$$v_i = \begin{Bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{Bmatrix} \quad (1) \qquad \theta_j = \begin{Bmatrix} \theta_j 1 \\ \theta_j 2 \\ \vdots \\ \theta_j n \end{Bmatrix} \quad (2)$$

In this context, given his convictions v_i, i reaches a utility level of $U(v_i, \theta_j)$ when his perception of a candidate is θ_j. In the case that a candidate is an exact match to the convictions of the citizen, the utility level should be at the highest possible value, denoted by γ. In every instance for which it holds that the positions are not identical the utility level is less than the maximum point. To summarize this, it holds that

$$U(v_i, \theta_j) \begin{cases} = \gamma & \text{if } \theta_j = v_i \\ < \gamma & \text{if } \theta_j \neq v_i. \end{cases}$$

As the utility of a voter only reaches γ for one unique θ, the utility function has one maximum and is concave but not necessarily symmetric, as can be seen in Figure 1. Thus, voter preferences are assumed to be single peaked, meaning that the further a point is from v_i, the less it is preferred by a voter (i.e. individuals' utility is downward sloping from their point of preference in this political spectrum).

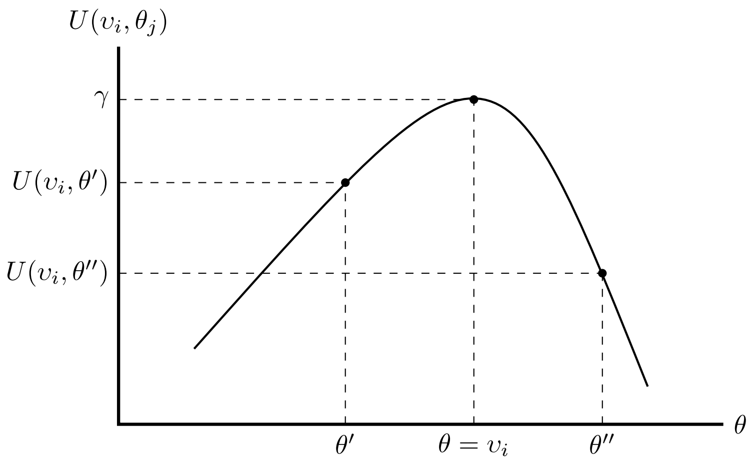

Figure 1: The utility function of a citizen with single peaked voting preferences.

In this model, a voter will choose the candidate that provides him the highest utility level. However, if there is no candidate for which it holds that $\theta = v_i$, a voter's utility cannot reach level γ. Hence, it might very well be that his cost of voting outweighs the utility gain provided by choosing a candidate. If this is the case, a citizen might abstain.

3.2 A descriptive model of voting

For simplicity we assume political parties to fit a one-dimensional continuum denoting left to right preferences. Also, as described before, we presume that preferences are single peaked. As in the work by Downs (1957a, 1957b), we state that the decision whether to vote or not depends on the corresponding expected utilities derived from voting and abstaining. In a Downsian world where rationality can be read as economic efficiency it is only rational to vote as long as the expected utility derived from casting a vote exceeds the expected utility of not doing so. To evaluate this for an individual i whose preferred party is p, Riker and Ordeshook (1968) specify expected utility V_i as

$$V_i = P_i \cdot W_i^p - C_i \tag{3}$$

where P_i represents the probability at which individual i estimates that his vote will affect the election outcome and where W_p^i is the personal gain that i experiences when p wins the election; C_i portrays an individual's cost of voting. Also, we denote the set of eligible voters of an election by N of whom a set of Y actually voted.

In this world it holds that i votes as long as it holds that $V_i > 0$, contrariwise it holds that i will abstain when $V_i \leq 0$. Since national elections often have a large number of eligible voters and voter turnouts are usually not extremely small, the Bayesian probability of one's vote making a serious impact is fairly low[2]. Hence, the utility denoted by W_i will be dampened out. It thus follows that in order for i to vote it should hold that W_i is a very large number ($W_i > \frac{C_i}{P_i}$). Since it is unlikely that a political victory will cause large individual reallocation, this this condition is unlikely to hold. Hence, under the specification given by (3), most of the citizens that vote will do so while expecting negative utility by voting.

However, Riker and Ordeshook (1968) want their equivalent of V_i to be descriptive for the actions of all individual voters and state that for each voter

[2]Following an estimate by Riker and Ordeshook (1968), would translate P_i into being as low as 10^{-8} for presidential elections in the United States.

i it must be the case that

$$\text{if } i \in N \text{ and } i \in Y, \text{ then } V_i > 0 \text{ and if } i \in N \text{ and } i \notin Y, \text{ then } V_i \leq 0. \quad (4)$$

This constraint on V_i is quite obvious, but not trivial as it allows for non-materialistic factors beyond the Downsian definition of rationality. For (4) to hold, there must be another factor which drives citizens in their decision to vote. An example of this could be the satisfaction of voting itself as in complying with a state's democratic tradition or a principled vote on a particular platform. We can thus rewrite (3) into

$$V_i = P_i \cdot W_i^p + D_i - C_i \quad (5)$$

where D_i displays the satisfaction that i derives from voting for his preferred party. Then, it follows that in order for i to vote it should either hold that W_i is a very large number or that D_i is sufficiently larger than C_i. Riker and Ordeshook (1968) go on to conclude that $P_i \cdot W_i^p - C_i$ is usually negative, which means that D_i is conclusive in a citizen's decision whether or not to vote.

3.3 Abstention: indifference and alienation

When all citizens vote, for instance in a system of compulsory voting, every voter maximizes his utility with respect to θ. However, if voting is not compulsory citizens might opt to stay at home. The literature recognizes two reasons for civilians to abstain from a vote. First of all, it could be the case that all candidates are very close to each other on the political spectrum. If this is true, it does not really matter to a citizen who gets elected into office since all candidates will yield him roughly the same utility payoff. This is abstention out of *indifference* indifference between the candidates. Using (3), where V_i displays the expected utility of voting, we now define W_i^p as ω, such that

$$V_i = P_i \cdot \omega - C_i \quad (6)$$

where ω is the difference in utility U between the candidate with U at a level closest to γ and the candidate with a level of U that is the furthest away from γ. Thus, it holds that

$$\omega = |U_c(\upsilon_i, \theta) - U_f(\upsilon_i, \theta')| \quad (7)$$

with the difference in utilities in its absolute value. This shows that when the candidates' platforms resemble each other to a greater extent, the additional benefit a citizen gets from voting for either of the candidates becomes smaller. Recalling (2), this does not necessarily mean that candidates are identical, since it could also be the case that two politicians have a different stance on all issues, but still end up generating an equal amount of utility U. This makes sense intuitively. If all politicians propose something that a citizen values similarly, the citizen will not care which politician will be elected into office and might abstain.

Another reason for abstention mentioned in the literature is *alienation*. This means that if there is a citizen who disagrees greatly with his 'most preferred politician', the citizen might decide not to vote at all. In order to model this, we should recall that in equation (5), D_i was defined as the satisfaction that i derives from voting for his preferred party. Given the definition of alienation, we regard this level of satisfaction as dependent on the difference between the optimal utility level γ and the actual utility level U. A citizen with a set of beliefs μ_i knows that he will yield utility level γ if these beliefs are equal to the positions of a politician j with an array of positions θ_j. Consequently, a citizen will get less satisfaction D_i by voting for a politician that gives him a utility level short of γ. Hence, D_i decreases at an increasing rate when the difference between γ and $U(\mu_i, \theta)$ becomes larger. We specify this as

$$D_i = \gamma - (\gamma - U(\mu_i, \theta))^2 \tag{8}$$

meaning that we can rewrite equation (5) as

$$V_i = P_i \cdot |U_c(\upsilon_i, \theta) - U_f(\upsilon_i, \theta')| + [\gamma - (\gamma - U(\mu_i, \theta))^2] - C_i \tag{9}$$

In the context of US presidential elections, the political convictions of the candidates on the ballot are usually not equable. Hence, ω will tend to be greater than zero. However, it is important to stress that ω is still conditional on P_i. Since many citizens participate in the United States' first-past-the-post electoral system, P_i will dampen out ω anyhow. As a result of this, abstention because of indifference plays a limited role when applying this model to presidential elections in the US.

In order to evaluate an election, we need to know how the preferred positions μ_i of all citizens are distributed. If we randomly select citizen i from the population, there is a certain probability that i has a preferred set of political convictions μ_i'. A characterization of the preferred positions of all citizens is given by a plot of the density of each μ, $f(\mu)$. For the US presidential elections, we assume this density to be bimodal for two reasons. First and foremost, this specification corresponds best to the two broad ideologies corresponding two the historical two party system. There is a concentration of voters with a similar set of convictions to the left and to the right of the center of the political spectrum. This notion is apparent in both the theoretical and empirical literature (Davis et al., 1970; Lewis, 2001). The second argument for not assuming a uni-modal distribution is that, in contrast with Downs' (1957b) prediction, political platforms do not converge to the center of the system. Alienation, as modeled above, combined with a bimodal distribution of preferences would be an explanation of the observed non-convergence, provided that turnout reacts sufficiently to changes in a candidate's strategy. Figure 2 is an example of this bimodal density function.

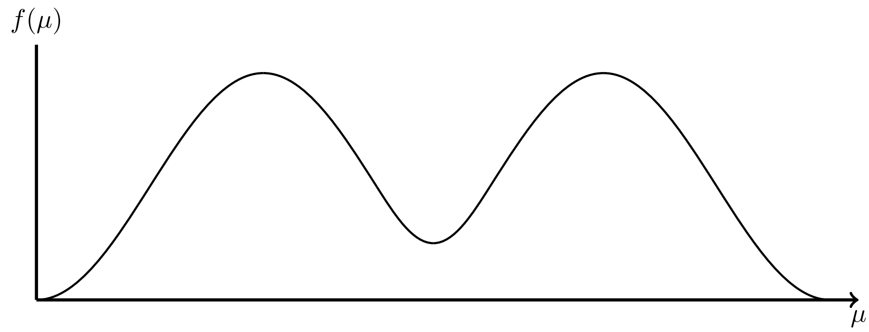

Figure 2: A bimodal distribution of civilians' preferences μ.

3.4 Predictions

Using the insights of the descriptive model, we can make some predictions on how the presence or absence of the Green Party on the ballot impacts the outcome of the election. First of all, recall that each citizen that does not abstain casts his vote towards his preferred party. If we would plot a density function for the preferences of all individuals that do vote, it could look like Figure 3. The dashed vertical axes are the positions θ that the four involved political parties take on the political spectrum, whereas the filled areas are the proportions of voters for who regard the indicated party as their preferred one. To give an example, for individuals associated to the green area it will hold that $U(\mu_i, \theta_{GP}) > U(\mu_i, \theta_{Dem})$, meaning that the Green Party is their preferred party. Note that the utility functions were assumed to not necessarily be symmetric, which means that preferred party identification is not by definition assigned to the shortest distance between μ and θ.

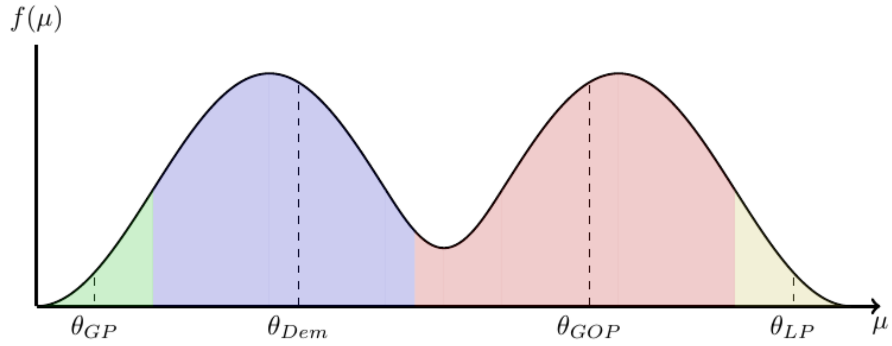

Figure 3: A bimodal distribution of all non-abstaining civilians' preferences μ and their preferred party, with party stances θ.

Consider the voters whose preferred party is the Green Party. If the Green Party would not be on the ballot, the Democratic Party becomes the preferred party for these voters. For some of these citizens it will hold that V_i is still positive, for others it will have become negative. Put differently, if the Green Party is on the ballot it will receive votes from people who would otherwise abstain as well as from people who would otherwise have cast their vote for the Democratic Party. To test whether these predictions hold, the following hypotheses will be evaluated:

Hypothesis 1: *The elections' turnout will be higher when the Green Party is on the ballot (abstainment effect).*

Hypothesis 2: *The Democratic Party will get a smaller share of the vote when the Green Party is on the ballot (preference effect).*

Since both parties compete for the left wing part of a similar base (i.e. the distance between θ_{GP} and θ_{Dem} is not that large), we suspect that the *abstainment effect* is smaller than the *preference effect*. We thus expect that:

Hypothesis 3: *The decrease in the Democratic Party's share of the vote is bigger than the decrease in the share of abstentions when the Green Party is on the ballot.*

4 Data and empirical strategy

We analyze county-level voting data for the US presidential elections of 2016. We consider nine separate dependent variables: (1-4) the percentages of the vote that the four participating parties get in a county, $\frac{\text{Votes for party } p}{\text{Total votes}}$, (5-8) the percentages of the vote that the four participating parties get in a county including abstainers, $\frac{\text{Votes for party } p}{\text{Total eligible votes}}$ and (9) the percentage of abstainers. The voting numbers come from two news organizations - Townhall and The Guardian - who published the data in a closed environment on their websites. The number of eligible voters in a county and county demographics were provided by the United States Census Bureau (USCB), which is responsible for producing economic and sociological data on the composition of the American populace.

The main variable of interest is a dummy variable indicating whether or not Jill Stein (Green Party) was an option on the ballot in a given county in the 2016 presidential election. In the United States, parties that want to participate passively in an election have to register for the ballot in each state separately. It is possible to register for the ballot in some states, while refraining to do so in other states. As of the nineteenth century, the two main political parties in the United States - the Republican (GOP) and Democratic Party (Dem) - have been on the ballot in each state continuously. In the 2016 election, the Libertarian Party (LP) and the Green Party (GP) were their only viable contenders. The Libertarian Party was on each state's ballot, whereas the Green Party participated in all but six states: Georgia, Indiana, Nevada, North Carolina, Oklahoma and South Dakota.

Numerous political scientists have shown that political preferences correlate with geographical patterns. It is not clear whether geography determines one's political stances or the other way around, but it is conventional wisdom that a distribution of preferences is usually clustered around a geographical determinant (Rodden, 2010; Chen, Rodden, et al., 2013). Moreover, the Democratic party's electorate is present in more densely populated areas, whereas GOP voters can be found in rural areas. In Figure 4 it can be seen that counties who voted for Trump are usually surrounded by a number of counties with a similar election outcome. This geographic clustering is partly driven by a high homogeneity in the demographic, which historically comprises blue collar whites (USCB, 2010).

This geographical dispersal of political stances means we can credibly assume

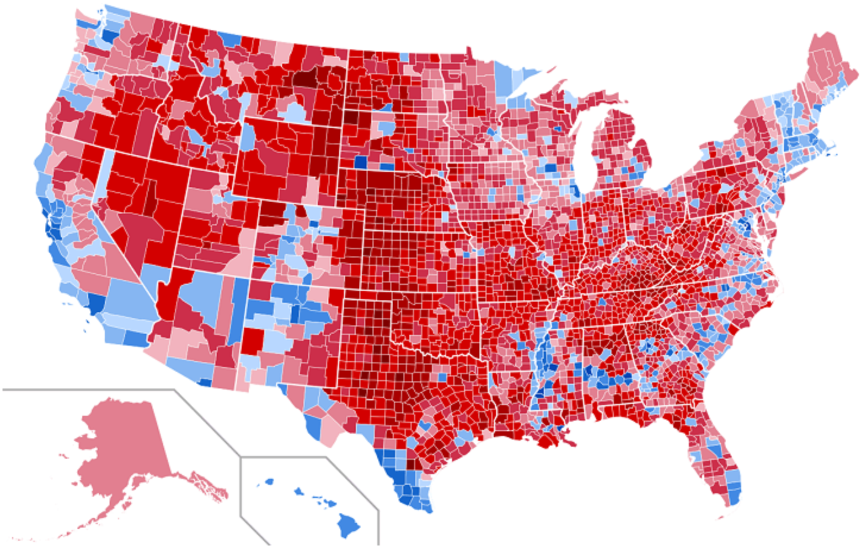

Figure 4: Share of votes on a county level for the 2016 US Presidential Election. Blue counties went to Hillary Clinton; red counties went to Donald Trump. The darker the shade, the larger the share by which the county was won.
Source: The Guardian

that the distribution of political preferences in two small neighboring areas is similar. Hence, if we take two counties on both sides of a state border, there will be limited variation in political stances. If the Green Party participates in the first but not in the second county, political preferences are similar, whereas the ballot is different. Since ballot registration is determined on the state level, this variation is random with respect to both counties (exogenous variation).

In order to accurately capture the effect of the variable of interest on the dependent variables, the ballots should be the same in all considered counties, irrespective of the independent variable of interest. As a result of this, 9 state borders with 185 counties can be taken into account. See Table 5 and Figure 6 in Appendix A for the relevant borders and their geographic location.

The considered counties on each border are characterized by similar demographic features. The populace in these rural areas is predominantly white (∼85%); most of them are dominated by GOP voters. Some of the counties taken into account have demographic outliers. The populace of these areas differs greatly from the other counties in its vicinity for mostly historical reasons.

For instance, the population of Sioux County in North Dakota exists for 80.6% of Native Americans, compared to ∼1% in the rest of the sample. Observations like this one contradict the assumed exogeneity of the variation and bias the estimated effects. Hence, nine outliers are dropped from the dataset. See Table 6 in Appendix A for an overview of these counties and their specific reason for exclusion.

Table 1 presents the main descriptive statistics of our paper specified for countries with and without Green Party ballot access. As the hypotheses suggest, the Democratic Party on average has a lower share of the vote in counties where the Green Party is a competitor. In contrast to the theoretical predictions however, a higher share of the electorate votes for the Republican Party or abstains in counties where the Green Party is an option on the voting sheet.

It is important to stress that if two bordering counties have similar demographic characteristics this does not mean that all counties in the sample share those very traits. To give an example, if some counties clustered in an area in Texas show equivalent demographic attributes, it is unlikely that these characteristics extend one-to-one to some neighboring counties in North Dakota. It might be the case that in some counties where Jill Stein was running, a higher share of the population was already inclined to abstain or vote for Trump. Hence, when analyzing the data, it is important to differentiate between different geographical areas. The measure for this geographical specification in this paper is the state border that each county belongs to.

More alarming is the fact that most counties in the sample vary with respect to their size (as in geographical size, e.g. acres), their amount of inhabitants and most importantly the ratio between these two. It is conventional wisdom that rural areas harbor predominantly conservative views and since all counties in the sample are located in rural areas, it is no surprise that Trump won all of them. However, these rural counties might be different to the degree in which they are urbanized[3]. If this is true, the assumption that preferences in two bordering counties are alike need not be true if there are differences in the degree of urbanization. The only measure at our disposal to control for this is a county's population. This control is not flawless, since it only distinguishes large and small population counties, disregarding their geographical size (where

[3]Both the New York Times and The Washington Post published an interesting analysis on how the degree of county urbanization affected the 2014 Iowa Senate election, see https://tinyurl.com/ycrvgv8z & https://tinyurl.com/yc25yc3c

the ratio between population and size would be a good estimate of the degree of urbanization).

Table 1: Descriptive statistics: the average percentages the four parties got in each state, the average percentage of eligible voters that abstained and county population averages, by Green Party ballot access.

Mean	Counties without GP	Counties with GP
% Dem (Clinton)	22.93	19.79
% GOP (Trump)	73.27	76.59
% GP (Stein)	NA	0.71
% LP (Johnson)	3.80	2.91
% Abstained	46.20	47.06
Population	23,020	30,944
Population median	14,700	14,676

Linear regression models using Ordinary Least Squares (OLS) and Fixed Effects (FE) were estimated to investigate the effect of the Green Party's ballot presence. To prevent inconsistent standard errors, heteroskedasticity robust standard errors were opted[4]. The most extensive version of the model distinguishes between the state borders by means of state border Fixed Effects and controls for population differences, such that

$$Y_{cs} = \alpha_s + \beta_1 X_{1cs} + \beta_2 X_{2cs} + \epsilon_c. \tag{10}$$

In this model, Y_{cs} depicts the share of votes a party obtains in county c in border area s and X_{1cs} represents the variable of interest - Green Party ballot presence. This is controlled for the population level, depicted by X_{2cs}. α_s is the state border specific constant, which demeans observations from area specific fixed effects. Moreover, ϵ_c is the error term, which captures all variation not explained by the incorporated variables.

[4]See the section dealing with robustness for additional measurements with Weighted Least Squares (WLS) and Feasible Generalized Least Squares (FGLS).

5 Results

The interpretation of the estimated regression models is straightforward. The explanatory variable of interest, the option to vote for the Green Party ticket, has a significant effect on the votes cast for the Democratic Party, when expressed as a share of all voters as well as all eligible voters. Observe models (3) and (6) in Table 2, which distinguish between the general geographic areas the counties are located in and which control for population levels in a county. The Clinton ticket gets a 3.35 percentage points lower share of the vote when the Green Party participates, in addition the a share of eligible votes won by Clinton becomes 1.79 percentage points smaller. This is in accordance with Hypothesis 2, which predicted that Green Party participation would shrink the Democrat's vote share.

What is more, Hypothesis 1 predicted that when Jill Stein would be a ballot option, a lower share of the electorate would abstain. The underlying idea is that there are voters who ideologically align with the Green Party and do not find their believes sufficiently reflected in the platforms of other parties and decide to abstain. No evidence of this effect can be found in the OLS regressions, as displayed in the models in Table 3. None of the explanatory variable's coefficients is statistically significant. The first two specifications of the model result in a positive effect that is somewhat economically efficient with a change of 0.86 and 0.32 percentage points respectively. The coefficient of the

Table 2: The results of the OLS regressions estimating the impact of Green Party ballot presence on the share of votes and share of eligible votes of the Democratic Party.

Dependent variable:	Share of votes (Dem)			Share of eligible votes (Dem)		
	(1)	(2)	(3)	(4)	(5)	(6)
Green Party on ballot	-0.0315**	-0.0259**	-0.0335***	-0.0180**	-0.0142**	-0.0179***
Standard Error	(0.0137)	(0.0126)	(0.0119)	(0.0079)	(0.0070)	(0.0067)
Fixed Effects	No	Yes	Yes	No	Yes	Yes
Population control	No	No	Yes	No	No	Yes
Number of observations	175	175	175	175	175	175
R-squared	0.030	0.232	0.326	0.029	0.298	0.366

* p<0.1, ** p<0.05, *** p<0.01

third model changes sign and is thus small that it cannot be considered to be of impact. The variable that controls for population differences[5] dampens out the effect of the explanatory variable of interest. This indicates that the main driver of abstentions is a county's population size (as a proxy of urbanization). The notion that more urban areas have relatively lower election turnouts is recognized widely in the literature, see for example the work by Timpone (1998); Leighley and Nagler (2013) & Cancela and Geys (2016). These findings clearly contradict Hypothesis 1, but are strictly speaking in line with Hypothesis 3, which predicted that the decline in the Democrat's voter share would be greater than the decline in abstentions.

Table 3: The results of the OLS regressions estimating the impact of Green Party ballot presence on the share of eligible voters who abstain.

Dependent variable:	Share of votes abstained		
	(1)	(2)	(3)
Green Party on ballot	0.0086	0.0032	-0.0006
Standard Error	(0.0141)	(0.0124)	(0.0123)
Fixed Effects	No	Yes	Yes
Population control	No	No	Yes
Number of observations	175	175	175
R-squared	0.002	0.276	0.299

* p<0.1, ** p<0.05, *** p<0.01

The story told by the regressions seems obvious. When the Green Party is an electoral candidate, Clinton's voter share drops, whereas abstentions are constant. However, when examining the explanatory variable when the Republicans or Libertarian voter shares are taken into account, something happens in addition to theoretical predictions. As shown in Tables 7 and 8 in Appendix B, the Republican Party's voter shares increase, whereas the shares of the Libertarian party become smaller. Since the data is an aggregate taken at county level, these numbers are difficult to interpret.

However, it should be stressed that the net effect on the combined Republican and Libertarian share of votes and eligible votes does not offset the change

[5]For every 10,000 county inhabitants, the abstention rate increases by 0.4 percentage points. This coefficient is significant with a p-value of 0.02 (standard error: 0.00017).

in the Democrat's shares. When the Green Party runs, it wins 0.7% of the vote. This is equal to the loss of the the Democratic Party, net of the changes in outcome for the Libertarians and Republicans (see Table 4).

It could be the case that the one dimensional representation of voting preferences is not accurate. If this is true, there could be a group of voters that does not identify with the two major parties, and likes to vote on the candidates in the periphery[6]. In the absence of the Green Party, these citizens would vote for the Libertarians. In the presence of the Green Party, a subgroup would vote for the Greens' candidate. Nonetheless, it would take data at the individual level to back this notion up.

Another explanation that should be mentioned is that Green Party ballot presence affects voters' evaluation of the Republican Party's candidate, surging utility levels $U(\mu_i, \theta_{GOP})$. Even though this could technically be inferred from the regressions, we deem the previously mentioned explanations as more realistic. Moreover, the literature does not back up the notion that the presence of political parties impacts voter preferences in this context.

Table 4: The results of the OLS regressions estimating the impact of Green Party ballot presence on the share of votes and share of eligible votes of the Green Party itself (which in essence is an estimate of in what vote shares participating or not will result in).

Dependent variable:	Share of votes (GP)			Share of eligible votes (GP)		
	(1)	(2)	(3)	(4)	(5)	(6)
Green Party on ballot	0.0071***	0.0070***	0.0070***	0.0038***	0.0037***	0.0037***
Standard Error	(0.0005)	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
Fixed Effects	No	Yes	Yes	No	Yes	Yes
Population control	No	No	Yes	No	No	Yes
Number of observations	175	175	175	175	175	175
R-squared	0.552	0.718	0.718	0.516	0.708	0.708

* p<0.1, ** p<0.05, *** p<0.01

[6]Which is a reasonable assumption, given the notion that both the Green Party as the Libertarian Party have stances that overlap and that are regarded as contradictory to the stances their respective bases are associated with. See also the footnote in the section on voter preferences.

6 Robustness

A serious threat to the validity of our results is heteroskedasticity, which biases the regression estimates. The problem of heteroskedasticity is that the variance of the error terms differs across observations. The reason that this is apparent in our dataset is straightforward. Recall that aggregate party preferences correlate with the level of urbanization. However, there is no clean indicator of county urbanization at hand. As explained in the data section, the population levels which were used are not more than a proxy of this.

The vast majority of the counties in the sample have a population that ranges from 2,000 to 20,000 individuals. Be that as it may, this does not by definition indicate that 2,000 person counties are more rural than 20,000 person counties. Since we do not know the geographic size of these counties, it could be that some of these counties are relatively small spaces in relatively densely populated areas. This issue applies with lesser extent to the more populous counties. Since the variability in geographic size is not too large, we know that

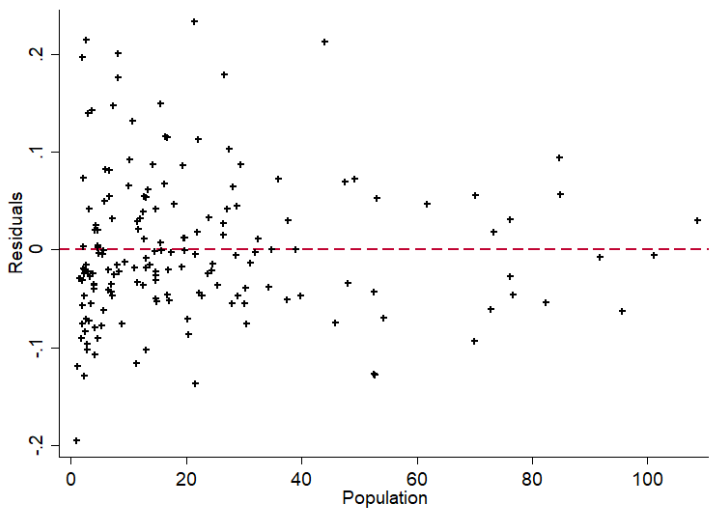

Figure 5: The residuals of Clinton's voter share when regressed on the Green Party ballot access, plot against population numbers (expressed in thousands); for a more clear representation four observations with a population over 120,000 were disregarded in the scatter plot. For a plot including these observations, see Figure 7 in Appendix C.

counties with a populace of 40,000 and over are likely to be relatively urbanized: the ratio of populace to size is high anyhow. Because of this certainty, there is less variance in the voter shares in these counties: they tend to be more Democratic leaning.

This effect can be seen in Figure 5. The larger the population becomes, the smaller the variance in residuals. The heteroskedasticity robust standard errors[7] used this paper's regressions improve upon regular OLS estimates, but are still biased (White, 1980). To further improve the estimates we weigh the observations so that counties with a higher populace - which have a smaller variance in their standard errors - are assigned higher gravity. To accomplish this we transform the model and apply Weighted Least Squares.

To do this, we first of all specify a model for the variance in the error term, conditional on a county's population level,

$$\text{Var}(\epsilon_c|P_c) = \sigma^2 \cdot \frac{1}{P_c}, \tag{11}$$

such that the conditional variance decreases when population increases and where σ^2 is unknown. As it is difficult to pin down the exact relationship, this is an estimate which should be regarded with some caution. Using the specification under (11), the dependent and independent data are transformed in order to run the full OLS regression with weights. This is done by giving observations with a higher population (of which the residuals are less spread out) more emphasis. Given that the specification of the model for the variance in the error term is correct, the transformed model has homoskedastic error terms.

The result of the WLS regression differs slightly from the full model estimate obtained before. The variable of interest has a negative coefficient similar to the fixed effects model's estimate net of GOP and LP effects, but is statistically insignificant. Nevertheless, we do not regard this result as decisive as the used specification of the variance in the error term is but a rough approximation.

In order to model the variance in the error term more accurately and correct for it using weights, we apply Feasible Generalized Least Squares. Taking again into account the full FE model from Equation (10), we specify the variance in the error term, conditional on a county's population level, as the unknown value σ^2 times the exponent of the linear combination of all the independent variables

[7]Heteroscedasticity-consistent standard errors as in White standard errors.

used in the full model, such that

$$\mathrm{Var}(\epsilon_c|P_c) = \sigma^2 \cdot \exp(\delta_0 + \delta_1 X_{1cs} + \delta_2 X_{2cs} + \gamma_2 D_{2s} + \ldots + \gamma_9 D_{9s} + \eta_c). \quad (12)$$

Since the variance cannot be negative, the exponent of the linear model is taken. In this linear model, X_{1cs} indicates Green Party ballot presence; X_{2cs} stands for county population levels (otherwise specified as P_c). The variables D_{1s} to D_{9s} are dummy variables indicating the state border areas[8]. The use of categorical variables renders this fixed effects model suitable for FGLS.

Using (10), we calculate the residuals $\hat{\epsilon}_c$ of the full FE model. Consequently, we estimate $\hat{g}_c$ by regressing the conditional variance on all independent variables[9]. This is formally specified as:

$$\ln \epsilon_c^2 = \delta_0 + \delta_1 X_{1cs} + \delta_2 X_{2cs} + \gamma_2 D_{2s} + \ldots + \gamma_9 D_{9s} + \eta_c, \quad (13)$$

the fitted values of which we specify as

$$\hat{g}_c = \hat{\delta}_0 + \hat{\delta}_1 X_{1cs} + \hat{\delta}_2 X_{2cs} + \hat{\gamma}_2 D_{2s} + \ldots + \hat{\gamma}_9 D_{9s}. \quad (14)$$

Then, we define $\hat{h}_c$ as

$$\hat{h}_c = \exp(\hat{g}_c). \quad (15)$$

Thereupon, this condition is used to transform the full FE model (10) into

$$\frac{y_c}{\sqrt{\hat{h}_c}} = \frac{\alpha_s + \beta_1 X_{1cs} + \beta_2 X_{2cs} + \epsilon_c}{\sqrt{\hat{h}_c}}. \quad (16)$$

Estimating (16) with regular OLS results in the FGLS estimate of (10), which in theory fixes the problem of heteroskedasticity as observations with a closer grouping of estimates are given more weight. Indeed, a Breusch-Pagan test applied on the FGLS model displays a slight improvement compared to the original model. However, the test indicates that the transformed model still suffers from a violation of normality in the distribution of its residuals.

The fixed effects FGLS regression displays that the Green Party ballot presence is associated with a 0.46 percentage point lower share of votes for the Democratic Party. This coefficient is not statistically significant (C: -0.0046; SE: 0.0082). However, since the sample is not very large (175 analyzed observations), the FGLS estimate is likely to be biased. In a small sample like this, the

[8] $\gamma_1 D_{1s}$ is left out of the model to prevent multicollinearity.

[9] Note that $\mathrm{Var}(\epsilon_c|P_c)$ is equal to $\mathrm{E}[\epsilon_c^2|P_c] - (\mathrm{E}[\epsilon_c|P_c])^2$ of which the second term is assumed to be 0 (zero conditional mean assumption).

FGLS distribution is likely to not be accurately centered around the parameters of the true population. Be that as it may, the sign and magnitude of the coefficient point in the same direction as the original FE model net of the GOP and LP effects.

7 Conclusion

Using and extending upon traditional voting models from the field of public choice theory, we made some predictions on how Green Party ballot presence would affect the aggregate of a county's voters. First of all, we predicted that having an extra ballot option would mobilize more citizens to come out and cast their vote. Furthermore, we predicted that the Green Party would cannibalize the votes normally cast towards the Democratic Party. Finally, it was stated that the second effect would be greater than the first.

Only the second hypothesis was backed up firmly by the analysis of the data. No evidence was found for an effect between voters having access to the Green Party and abstention rates. Hence, the third hypothesis is strictly true, but does not convey additional information.

The Fixed Effects OLS model displayed an economically and statistically significant negative effect of Green Party presence on Democratic voter shares. Since heteroskedasticity is a problem in the sample, some measures to insure robustness were established. These measures did not irrefutably substantiate the effect found, most likely due to a not optimally large sample size. However, they connote that the outcome found is at least statistically sound. Despite this, caution is key in its interpretation.

Regressions for the other political parties provided results not predicted by theory: the share of votes cast for the Republican Party increases, whereas those for the Libertarians decrease. This may have some implications for the theoretical model. For example, it might be that the one dimensional display of party preferences is not adequate. The model operated under the assumption that preferences (utility functions) differ across individuals, but that all prefer the political party closest to one's convictions. Alienation and indifference then decided whether a citizen would actually take the effort of voting. However, the data seems to indicate that some individuals like to vote on the periphery. This would for instance mean that the Libertarian Party on the right side of the spectrum is the first best alternative for some Green Party voters. However, more research - ideally using data at the individual level - is needed to draw a firm-standing final conclusion on this. Moreover, this does not explain the surge in GOP voter shares when the Green Party is on the ballot.

A possible explanation is one that partly undermines the results of the Fixed Effects model. In this estimation, all observations were demeaned for the state border fixed effects. The underlying rationale is that two small, neighboring counties are likely to have similar demographic characteristics. However, not all counties in the sample have the same population traits. Hence, they were grouped by border. Thus, a county on the east side of the Oklahoma-Texas border is regarded to be in the same geographic location as a county on the west side of the same area. Although demographic characteristics are quite stable in rural areas like these, it might be that the used grouping was too rough, biasing the results. This notion could explain the unexpected effect for the Republican Party, and partly attenuate the effect found on the Democratic Party.

It is, however, important to recall that the effect on the Democratic voter shares net of the GOP and LP effects still is rather large. If this relationship were one-to-one - which cannot be proven due to the aggregate nature of the data - Green Party presence means winning 0.7% of the vote at the cost of the Democrats; this with border specific constants applied.

In the introduction, it was mentioned that the outcome of the 2016 presidential election would have been different if the results had been slightly different in just three states. In Michigan, Wisconsin and Pennsylvania the margin between the Republicans and Democrats was just 0.3, 0.8 and 1.2 percentage points, respectively. Clearly, these numbers are easily offset by the effect measured by the Fixed Effects model. As mentioned before however, it is unlikely that this effect is robust enough to allow for this conclusion. The effect net of the GOP and LP estimates would only sway the outcome in Michigan. The same holds for the coefficient found in the Fixed Effects Feasible Generalized Least Squares model.

Despite this, recall that there was no sign of an effect on abstentions: contrary to the theoretical expectation, it could not be established that more people do not vote when the Green Party is not on the ballot. In this context, it should be noted that the Green Party received 1.1%, 1.1% and 0.8% of the vote in the three respective states. These numbers are higher than those found in the regressions where the GP's voter share was the dependent variable, as the regressions controlled for fixed effects and these states tend to be more urban (i.e. more progressive) than the areas included in the sample.

Under the condition that the one-to-one effect implied by the previously described coefficient net of GP and LP holds, the election outcome is even closer. Given the lack of an abstention effect, Jill Stein's absence on the ballot in these states would then be enough to turn at least Michigan and Wisconsin. The margin in Pennsylvania between Trump and Clinton net of the GP voters that transfer to Clinton is 0.4%. This is a gap of approximately 24,000 Pennsylvanian votes; an amount so small that the outcome of the election is close to arbitrary.

On the road to further research, acquiring individual level voting data is key. Mapping individual voter preferences will provide more insight in the decision making which is at play when a citizen decides who to vote for. If this individual level data is not available, it would still be possible to improve upon the robustness of the findings in this paper. First of all, the incorporation a true variable for county urbanization will doubtlessly provide valuable insights. Furthermore, the Fixed Effects model could be specified to demean for smaller geographic areas. Last but not least, improvements could be made by integrating all US counties to the sample. Combined with a range of variables conveying county specific demographics, this sample would be suited for Propensity Score Matching. This more deliberate approach of differences in the traits of county populations will strengthen the assumption of exogeneity in the variation observed. Also, it would allow for more precise and county specific estimates of the effect.

8 Appendix A

Table 5: The state borders that were included in the sample.

State with Green Party	*State without Green Party*
Alabama	Georgia
Illinois	Indiana
Kansas	Oklahoma
North Dakota	South Dakota
Ohio	Indiana
Tennessee	Georgia
Tennessee	North Carolina
Texas	Oklahoma
Wyoming	South Dakota

Table 6: The counties that were excluded from the sample.

State	*County*	*Reason excluded from sample*
Alabama	Russell County	41.4% of the population is black
Georgia	Clay County	61.1% of the population is black
Georgia	Muscogee County	44.2% of the population is black
Georgia	Stewart County	54.2% of the population is black
Illinois	Cook County	County includes the suburbs of Chicago
Illinois	Will County	County includes the suburbs of Chicago
Indiana	Lake County	County includes the suburbs of Chicago
North Dakota	Sioux County	80.6% of the population is Native American
Ohio	Hamilton County	County includes the suburbs of Cincinnati

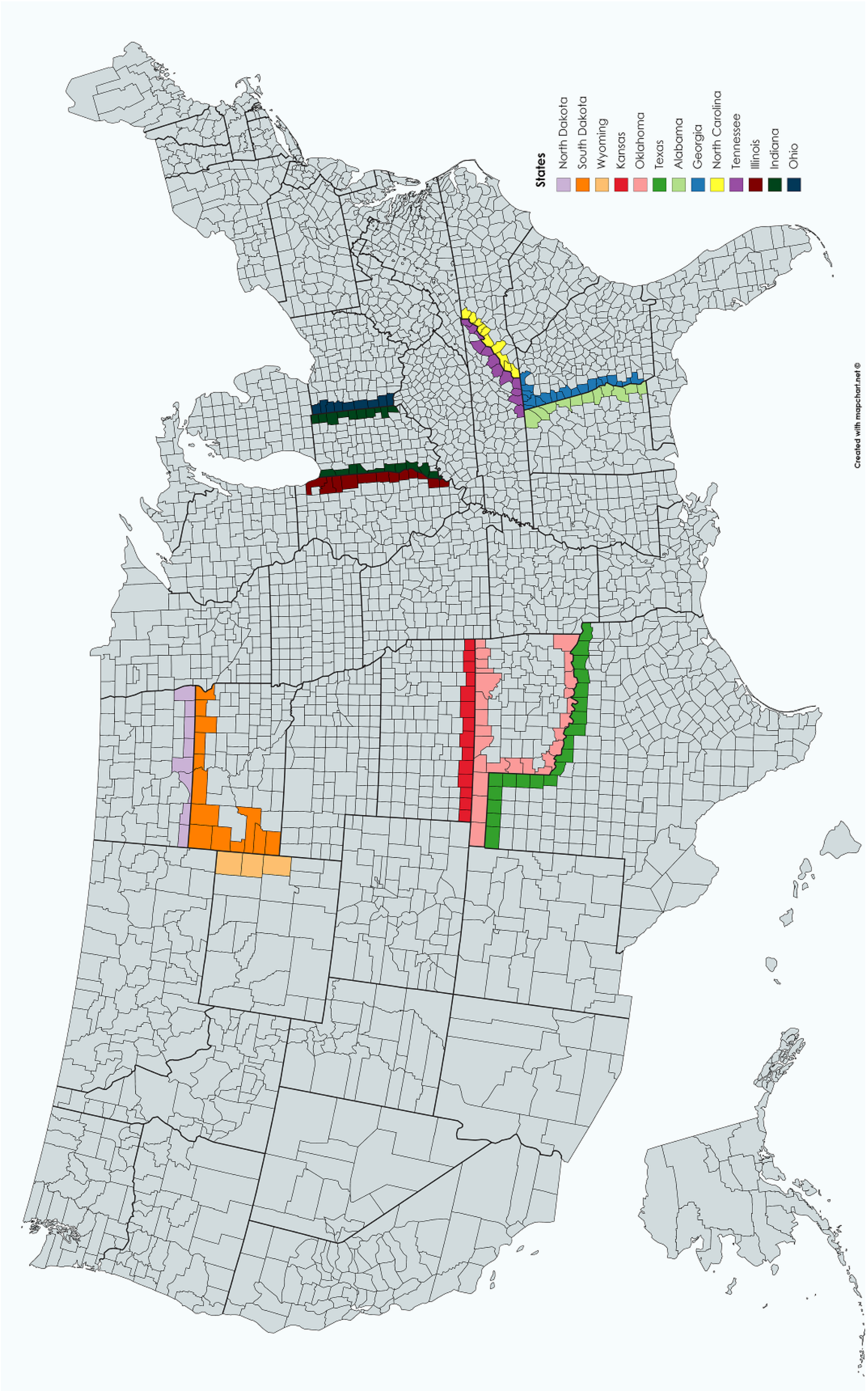

Figure 6: Overview of all US counties with the bordering counties taken into account for this research highlighted.

Source: Mapchart.net

9 Appendix B

Table 7: The results of the OLS regressions estimating the impact of Green Party ballot presence on the share of votes and share of eligible votes of the Republican Party.

Dependent variable:	Share of votes (GOP)			Share of eligible votes (GOP)		
	(1)	(2)	(3)	(4)	(5)	(6)
Green Party on ballot	0.0333**	0.0281**	0.0363***	0.0104	0.0119	0.0196
Standard Error	(0.0142)	(0.0131)	(0.0124)	(0.0134)	(0.0130)	(0.0124)
Fixed Effects	No	Yes	Yes	No	Yes	Yes
Population control	No	No	Yes	No	No	Yes
Number of observations	175	175	175	175	175	175
R-squared	0.031	0.219	0.322	0.003	0.114	0.220

* p<0.1, ** p<0.05, *** p<0.01

Table 8: The results of the OLS regressions estimating the impact of Green Party ballot presence on the share of votes and share of eligible votes of the Libertarian Party.

Dependent variable:	Share of votes (LP)			Share of eligible votes (LP)		
	(1)	(2)	(3)	(4)	(5)	(6)
Green Party on ballot	-0.0089***	-0.0092***	-0.0097***	-0.0047***	-0.0047***	-0.0049***
Standard Error	(0.0020)	(0.0014)	(0.0014)	(0.0013)	(0.0008)	(0.0008)
Fixed Effects	No	Yes	Yes	No	Yes	Yes
Population control	No	No	Yes	No	No	Yes
Number of observations	175	175	175	175	175	175
R-squared	0.107	0.567	0.590	0.075	0.614	0.623

* p<0.1, ** p<0.05, *** p<0.01

10 Appendix C

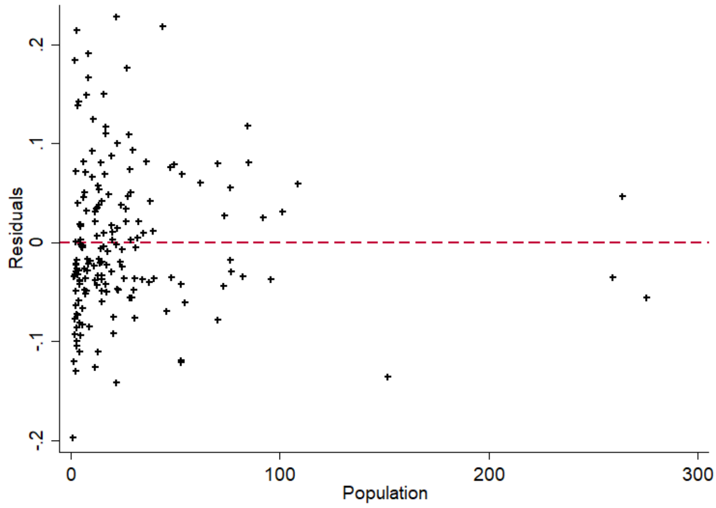

Figure 7: The residuals of Clinton's voter share when regressed on the Green Party ballot access, plot against population numbers (expressed in thousands).

References

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy*, *56*(1), 23–34.

Cancela, J., & Geys, B. (2016). Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral Studies*, *42*, 264–275.

Chen, J., Rodden, J., et al. (2013). Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, *8*(3), 239–269.

Congleton, R. D. (2003). The median voter model. *The Encyclopedia of Public Choice*, 707–712.

Davis, O. A., Hinich, M. J., & Ordeshook, P. C. (1970). An expository development of a mathematical model of the electoral process. *American Political Science Review*, *64*(02), 426–448.

Downs, A. (1957a). *An economic theory of democracy* (3rd ed.). New York: Harper & Row.

Downs, A. (1957b). An economic theory of political action in a democracy. *Journal of Political Economy*, *65*(2), 135–150.

Hotelling, H. (1929). Stability in competition. *The Econonomic Journal*, *39*(153), 41–57.

Leighley, J. E., & Nagler, J. (2013). *Who votes now?: Demographics, issues, inequality, and turnout in the united states.* Princeton University Press.

Lewis, J. B. (2001). Estimating voter preference distributions from individual-level voting data. *Political Analysis*, *9*(3), 275–297.

Riker, W. H., & Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review*, *62*(01), 25–42.

Riker, W. H., & Ordeshook, P. C. (1973). *An introduction to positive political theory* (Vol. 387). New Jersey: Prentice-Hall.

Rodden, J. (2010). The geographic distribution of political preferences. *Annual Review of Political Science*, *13*, 321–340.

Rowley, C. K. (1984). The relevance of the median voter theorem. *Journal of Institutional and Theoretical Economics*(H. 1), 104–126.

Smithies, A. (1941). Optimum location in spatial competition. *Journal of Political Economy*, *49*(3), 423–439.

Timpone, R. J. (1998). Structure, behavior, and voter turnout in the united states. *American Political Science Review*, *92*(1), 145–158.

USCB. (2010). *The united states census bureau census 2010.* (`https://www.census.gov/quickfacts/fact/table/US/PST045216`)

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.