UNIVERSITY OF AMSTERDAM

MASTER THESIS

Campaigning in the US presidential election based on a network model

Author:

NICK WORTEL

Student ID's:

11688866

Supervisor:

FABIAN GREIMEL

December 14, 2021

UNIVERSITEIT VAN AMSTERDAM

Abstract

This thesis analyses the optimal states to campaign in for an American presidential election based on a simulated network model. The advantage of simulation is that it can handle the non-linearities of the winner takes all system. This method is not yet performed in earlier literature for election prediction. Using the elections from 2004 to 2020 the network model can explain observed campaign events better than a model without a network component. The simulation also shows there are some states that candidates can visit to improve their winning chance in the general election.

Statement of Originality

This document is written by Student Nick Wortel who declares to take full responsibility for the contents of this document.

I declare that the text and the work presented in this document are original and that no sources other than those mentioned in the text and its references have been used in creating it.

The Faculty of Economics and Business is responsible solely for the supervision of completion of the work, not for the contents.

1 Introduction

In the elections of the United States campaigning plays a large role. Candidates sometimes announce their candidacy only months after the previous election. With this almost permanent campaign, it is important for candidates to know where to campaign. One could argue, candidates should campaign in the states that are the closest in the polls. But this disregards human interaction and behaviour.

This thesis uses a new way to look at this dilemma from a network perspective and finds which states should be prioritized by candidates to make the most of their limited visits during a campaign. These states are not contested in isolation, but are also connected to other traditional swing states.

To find those states simulation is used. Such methods have shown their strength in media—like Silver (2020) for the elections of 2016 and 2020— but have not yet been explored in scientific literature. The field of election forensic uses some similar methods to try and detect election fraud, like Tunmibi and Olatokun (2021) for the Nigerian election with a Monte Carlo simulation. However, it is not used to analyse campaign strategies. The advantage of simulation is that it is better able to capture the non-linearities of the winner-takes-all system than a simple regression and thus is better suited for analysing shifts in the electorate.

Campaign events have an influence on the voters in the county or state they are held as shown by Barton, Castillo, and Petrie (2014). So to look at the effectiveness of campaigning in a certain county or state a group of voters will be added to the candidate campaigning there. The increase in wins for that candidate will be analysed for all counties. The inclusion of a network

aspect gives a ripple effect outside the county where the event took place.

For such a simulation, the baseline winning probabilities need to be gathered by a model. For this several different tested and proven factors will be used. These include economic factors (such as unemployment and poverty), demographic factors (like racial composition and religious orientation) and characteristics of the candidates (such as their home state).

To take into account network effects, *partisan exposure* from Keuzenkamp (2021) is added. The *partisan exposure* measures what fraction of friendship links are with voters of a given party. It uses the Social Connectedness Index (SCI) of Bailey et al. (2018), which measures the probability of two people in different counties being friends on Facebook. With around 72% of people in the US using Facebook (Pew Research Center, 2021) SCI is a great estimator for connections between people.

The structure of this thesis will be as follows. A literature review on factors that affect the election. Then a look at the data and the methods used in this thesis. Followed by the results of the regression and the influence measure from the simulation. Finally, the conclusion where some final remarks on this paper, as well as some proposals for possible future research, will be made.

2 Factors that affect elections

2.1 Election factors

The literature on election prediction is vast. There are many factors that influence voting behaviour in a county. In this thesis, some of the most

prominent factors will be used in the regression. These can be categorised into three groups: economic, demographic and political factors.

The perception of the economy is important for people's voting behaviour (Lewis-Beck and Paldam, 2000), thus factors that people notice in their day-to-day life will be used, for example unemployment rate and poverty. The demographic factors are focus on ethnicity, age decomposition and religion (Ambrosius, 2016; Sioh, 2018; Ansolabehere, Rodden, and Snyder, 2006; Levernier and Barilla, 2006; Junn, 2017; Bishin and Klofstad, 2012). Finally the political factors are turnout, campaign events and home states of the candidates (Mixon and Tyrone, 2004; Kahane, 2009; Barton, Castillo, and Petrie, 2014).

2.2 Exposure

72 percent of Americans use Facebook (Pew Research Center, 2021). Bailey et al. (2018) used this data to create a network measure called SCI, which is a measure of the probability a person in county A is friends with someone in county B. The SCI was used in the paper by Keuzenkamp (2021) to estimate the fraction of friends with a certain party affiliation a person in county i has with the following formula:

$$Exposure_{i,t}^{Republican} = \frac{\Sigma_{j \neq i}^{N} SCI_{i,j} \cdot \#Republican votes_{j,t}}{\Sigma_{j \neq i}^{N} SCI_{i,j} \cdot Population_{j,t}}$$

Keuzenkamp (2021) showed that the exposure measure has positive correlation on the vote share of that party in a county. In this thesis the exposure will be slightly different, it will include the county itself and it will be lagged

by one time period.

3   Data and methodology

The focus will be on the elections of 2004 to 2020. First, I will explain the data and its sources, then I will show the basis of the panel regression, and examine the exposure measure. Third, I will explain the parameters of the simulation and how it works. Finally, I show how the influence is obtained.

3.1   Data

Most data that is used in this thesis comes from the Census Bureau. These include: (youth)poverty, median income, education attainment (U.S. Census Bureau, 2021b), unemployment rate, sex, age and racial composition (U.S. Census Bureau, 2021a). The rural continuum is published by The US Department of Agriculture (2020). The values of the S&P500 and the NASDAQ are from Yahoo! Finance (2021). The religious memberships are from The Association of Religion Data Archives (2002). The election results are from MIT Data and Lab (2018). Voter turnout is calculated by dividing the total votes in a county by all inhabitants older than 18. The values of the SCI are also publicly available at Mihaila (2021). The campaign visits were web scraped from Appleman (2021). Some of these variables are only calculated once every decade. To solve this problem simple linear inter- and extrapolation are used to fill the gap. This is done for religion, attained level of education and the rural continuum.

3.2   Regression

Because of the structure of the data, a panel model will be used. This will be done for five elections (2004 until 2020) and 3098 counties. So the regression equation looks as follows:

$$Voteshare_{i,t}^{Republican} = \beta_0 + \beta_1 Control_{i,t} + \beta_2 Exposure_{i,t}^{Republican} + \epsilon_{i,t}$$

Here i is for the county and t for the election year. *Control* consist out of all aforementioned variables and β_1 is thus a vector of length 39. Alaska uses election districts that are different from the counties and thus only has a limited data set of 3 counties with approximately 40 percent of the population. This means that it can be used for later simulation, but shifts here will be over weighted. Another important thing about panel regression is that values that stay constant through time need to be excluded and their influence will be incorporated through the use of a constant. So variables like the census region are excluded.

3.3   Exposure measure over the years

Because the SCI stays constant the difference in the exposure measure is caused by population and voting behaviour. In figure 1 the difference between Democratic and Republican exposure is shown.
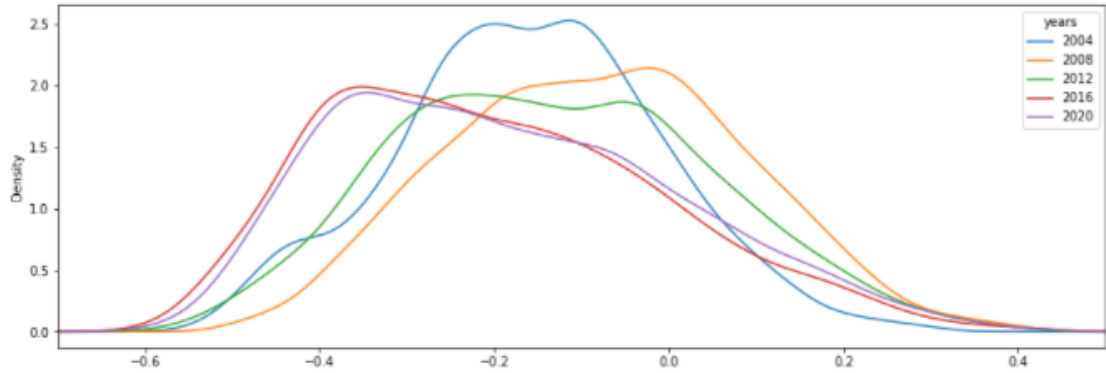
Figure 1: Kernel density estimate of Democratic exposure minus Republican exposure

Here just like in Keuzenkamp (2021) the Republicans have a higher exposure than Democrats. With the exception being 2008 the year that Obama was elected for his first term and what also is considered a blue wave election. This is because Democrats are more present in large, densely populated counties which leads to their exposure being slightly lower than expected. This can be seen when the exposure is divided by the county's population, the distribution becomes centered closely around zero. Interestingly, the difference between 2016 and 2020 is very small but this can be explained by states like Wisconsin and Michigan, which both went to Trump with only 20,000 and 10,000 votes respectively in 2016 and also with small margins to Biden in 2020.

3.4 Simulation

After the model is fitted and the predicted vote shares for every county are calculated. The question arises, what effect will a shift in voting behaviour in one county have on the general election results. With the winner takes all

system, if a candidate already has more than half of the votes in a state it will not make a difference at all, because he will still receive the same amount of electoral votes from the state. To make the model more dynamic and thus be able to look at the effects of additional voters in a county, a simulation is used. The simulation consists out of 4 main steps.

1. The first step is calculating all predicted vote shares in the counties. This is done using the panel model.

2. The second step is calculating the expected state-level vote share with the following equation for state j:

$$Statevotesshare_j = \Sigma_{i \in j} \widehat{Voteshare}_i^{Republican} * eligiblevoters_i$$

So the value of state vote share is a prediction of the popular vote in a state.

3. The third step is using the calculated state vote share and using this value, which is between 0 and 1, as the mean of a normal distribution. This is done for all 50 states plus D.C., so there are 51 normal distributions with all different means but the same variance.

4. The final step is drawing a value from all of those 51 distributions. If a value is higher than 0.5 the state is won by the candidate, if it is lower it is lost. The electoral votes of the states that were won are summed up and if this is higher than 269 a candidate wins if it is lower he or she loses.

To make the results robust the last step of the simulation must be repeated several times. The variance needs to be estimated, this will be done by grid search. The variance needs to be as high as possible for the model to be dynamic, but not too high that it will cause the simulation to be completely random and not converge.

With this simulation even though a state has been called by the regression for a given candidate a change in voter behaviour will have consequences for the win chances, especially for close races. Note that this simulation does not take into account electoral systems that states where there are electoral votes for districts and the state as a whole, for example Maine and Nebraska. These will be seen as winner takes all system just like all other states.

3.5 Influence measure

To quantify the importance of campaigning in a certain county the influence measure is introduced. First, the baseline is calculated by running the simulation N times with the estimated county's popular votes. The number of simulated elections where the candidate obtains more than 269 electoral votes are counted up and divided by N to generate a national win chance. To measure the effects of campaigning in a county j county j's votes will be increased by δ_j.

$$\Delta \widehat{Voteshare}_{j,t}^{Republican} = \frac{\delta_j}{Population_{j,t}}$$

In the case of exposure, this delta will cause a ripple effect to other counties based on the social connectedness between them. This ripple effect for county

i when there is a shift in county j can be incorporated as follows:

$$\Delta \widehat{Voteshare}_{i,t}^{Republican} = \widehat{\beta}_2 \frac{SCI_{i,j} \cdot \delta_j}{SCI_{i,j} \cdot Population_{j,t}}$$

The new vote shares are used in the simulation and this simulation is repeated N times. The new win chance will be calculated the same as before. This is repeated for all 3098 counties. The influence of a county is the new win chance minus the baseline.

4 Results

In this section, the results of the panel regression will be shown. After this, I will show the effectiveness of the simulation. Finally, I will thoroughly examine the influence measure for both Democrats and Republicans.

4.1 Regression

Note that the vote share used in the regression is not the fraction of total votes but the fraction of total Democratic and Republican votes. The exposure here is the exposure of the party that is used as the dependent variable. Table 1 shows the result of the panel regression with the left column being for the Democrats and the right one for the Republicans and the standard deviation between brackets.

Table 1

	Democratic	std err	Republican	std error
const	-0.267191***	(0.024697)	0.154685***	(0.024481)
% 0-19 years old	-0.091755***	(0.024098)	0.130108***	(0.023976)
% 20-29 years old	0.172465***	(0.025959)	-0.304633***	(0.025756)
% 30-44 years old	0.412029***	(0.026317)	-0.277962***	(0.026417)
% 45-64 years old	-0.020826	(0.031604)	-0.157593***	(0.031454)
% bachelor	0.003416***	(0.000133)	-0.003498***	(0.000133)
% high school	0.001114***	(0.000123)	-0.001763***	(0.000122)
% less than high school	0.001065***	(0.000121)	-0.000964***	(0.000121)
% Amish	0.000053**	(0.000025)	-0.000060**	(0.000025)
% Baptist	0.000062***	(0.000019)	-0.000056***	(0.000019)
% Catholic	0.000045***	(0.000004)	-0.000044***	(0.000004)
% Evangelicals	0.000014***	(0.000005)	-0.000042***	(0.000005)
% Jewish	-0.000328***	(0.000069)	0.000510***	(0.000069)
% Salvation army	0.000269***	(0.000100)	-0.000193*	(0.000099)
% Southern batptis	-0.000083***	(0.000006)	0.000082***	(0.000006)
% United methodist	0.000035***	(0.000010)	-0.000079***	(0.000010)
% Muslim	0.000084***	(0.000031)	-0.000102***	(0.000031)
% white female	-0.141043***	(0.024328)	0.047945**	(0.024212)
% white male	-0.345624***	(0.026108)	0.435557***	(0.025921)
% black female	0.178701***	(0.029641)	-0.179120***	(0.029531)
% black male	-0.401032***	(0.033313)	0.329305***	(0.033232)
% Hispanic female	0.255481***	(0.038247)	-0.281432***	(0.038099)
% Hispanic male	-0.535632***	(0.037024)	0.520171***	(0.036923)
% Asian male	0.030753	(0.049409)	0.023404	(0.049249)
Unemployment rate	0.001234***	(0.000264)	-0.001596***	(0.000262)
SP500	-0.000059***	(0.000012)	-0.000048***	(0.000012)
Nasdaq	0.000029***	(0.000003)	0.000001	(0.000003)
Inflation	0.052167***	(0.001868)	-0.067920***	(0.001878)
Median income	-0.000002***	(0.000000)	-0.227536***	(0.029426)
Youth poverty	-0.161294***	(0.019975)	0.083574***	(0.019926)
Poverty	0.203285***	(0.029524)	0.000002***	(0.000000)
Homestate Democratic candidate	0.013931***	(0.003175)	-0.004660	(0.003168)
Homestate Democratic vp	0.002889	(0.003068)	0.000699	(0.003059)
Homestate Republican candidate	-0.017438***	(0.002670)	0.024263***	(0.002661)
Homestate Republican vp	0.004910	(0.003130)	-0.006635**	(0.003119)
Metro	-0.010504***	(0.001287)	0.008340***	(0.001283)
Metro adjacent	-0.000886	(0.001090)	0.001176	(0.001086)
Campaign visits	0.000324*	(0.000195)	0.000170	(0.000253)
Exposure	1.359807***	(0.007981)	1.337363***	(0.007768)
Third party vote	0.251889***	(0.025102)	-0.200961***	(0.025031)
Voter turnout	0.131459***	(0.008600)	-0.083329***	(0.008617)
R-squared and Adj. R-squared	0.890436	0.890145	0.891027	0.890738

Both regressions perform well with an R-squared around 0.89 and most variables have a p-value lower than 0.01. Most variables have the same sign as in the literature. There are a few exceptions these include, age 0 to 19, metro (adjacent) and bachelor those variables have signs opposite to the literature. The lagged exposure has values of 1.360 and 1.337 for democrats and republicans respectively, both are also significant with the p-value below 0.0000001. So the value of exposure is close to the results of Keuzenkamp (2021) where they were above one.

4.2   Parameter selection

The simulation has three important parameters that need to be selected, these are N the number of times the simulation is repeated, sigma the variance of the distribution in selecting a state winner and delta the shift in votes. For N it is important that it is high enough that it will converge, but low enough to be computable in a reasonable time. To find the value of N for which the general win chance has converged, the simulation was performed one million times. By taking the mean of an increasing number of observations out those million a line was drawn. For N smaller than 50,000 the line is quite volatile, but after 150,000 iterations the percentage has stabilised. With N set to 150,000 one iteration takes around 17 seconds and is thus still manageable.

To find the optimal sigma a grid search is used. Lower values of sigma lead to overestimation for the candidate that was predicted to win by the model. Higher values of sigma push the average electoral votes to 269 and thus favours the candidate that would lose based on the regression only.

For all years the predicted value of electoral votes is the same as the real electoral votes for sigma set to 0.03. If sigma is changed with a value of 0.01 the prediction is then 2 electoral votes off.

Finally, the value of δ_j needs to be chosen. This value needs to be reasonable but sufficient to compensate for the randomness of the simulation. A campaign event can vary wildly in scale, from a town hall with a few hundred people to a full baseball stadium with 30,000 visitors. So, to take both into account and assume that a candidate can stay more than one day, δ is chosen to be 50,000.

4.3 Simulation

Let now $\sigma = 0.03$ and $N = 150,000$. Figure 2 shows how well the simulation works for the 2020 election.
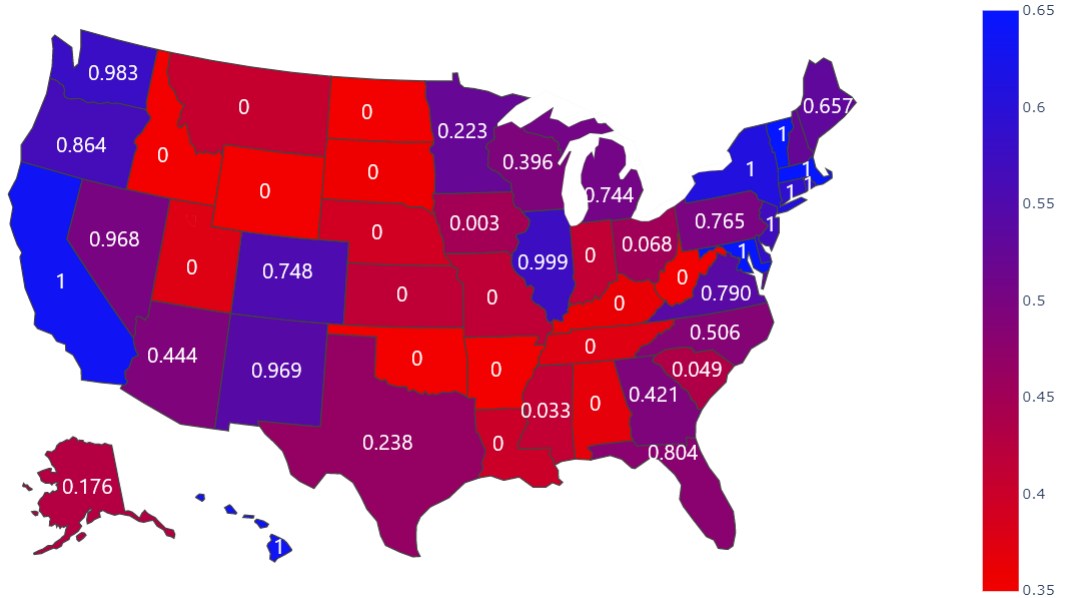
Figure 2: The real and simulated election results

In figure 2 the colours of the states are the observed vote shares for Biden. This means that low values and thus red coloured states were won by Trump and blue coloured by Biden. The purple states are contested states. The values shown on the states are the percentage of wins for Biden for that state in the simulation. New Hampshire, Delaware and D.C. are not visible on the map. Their values are 0.676, 1 and 1 respectively, all three were won by Biden; New Hampshire with 7 percentage points, Delaware and D.C. with margins of larger than 15 percentage points difference.

The first thing to notice is that 26 states and D.C. are 1 or 0 these states can be seen as solid blue or red and thus not of great importance on the campaign. This aligns well with the states that were won by a difference of more than twenty percentage points either way. Of the 50 states and D.C., the simulation was correct for 45 states. The six that were not correct are Arizona, Montana, Wisconsin, North Carolina, Georgia and Florida. Most of those states were very close with the difference between Trump and Biden being less than 1 percentage point in real life. The model shows these states to be close as well with their win percentages not more than 10 percentage points off from the 50 percentage threshold. The two exceptions to this are Florida and Minnesota. Florida was won by Biden more than 80 percent of the time in the model, but in real life Trump won with a 3.3 point difference. Minnesota was in the simulation only won 22 percent of the time by Biden, while in the real election Biden won by 7 percentage points.

So in general the simulation performs well. It calls the right winner in most states. The ones it predicts incorrectly are indicated as close races by the model and that narrowly went the other way in real life.

4.4 Influence measure

The influence measure is calculated for all 3098 counties and for the election of 2020. The mean of the influence is 0.006652 and 0.010759 for Democrats and Republicans respectively. Both have a standard deviation around 0.0067. The results are shown on the map in figure 3 and 4.
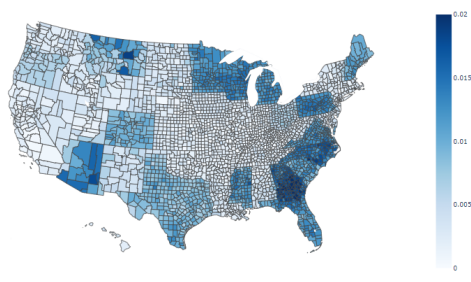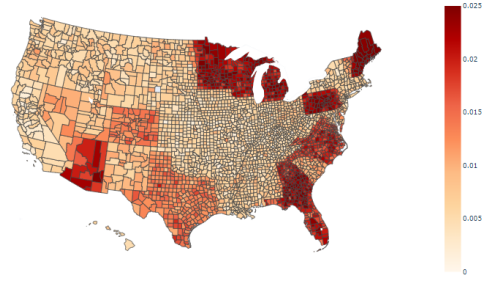

Figure 3: Influence for Democrats

Figure 4: Influence for Republican

Most states are similar for both parties, which is to be expected because they are all swing states and or closely connected to swing states. For both parties, Georgia has the highest influence and this is also in line with the extremely close results of 2020. Other states that are highly influential are Minnesota, Wisconsin, Michigan, Pennsylvania, Virginia, North Carolina, Florida and New Mexico and to a lesser extent Colorado and Texas. Maine and New Hampshire seem to be more influential for the Republicans. For Democrats, Montana and Mississippi show potential. These results show that the model finds most traditional swing states to be influential in this model as well.

The swing states that were not identified were Iowa, Nevada and Ohio. This can be understood when looking at the connections these states have. For example, Ohio is connected mostly to Kentucky, Wyoming and Indiana.

These are all solid red states whilst Nevada is connected to California, New Mexico and Oregon; these are all solid blue states. Thus shifts in these states have a ripple effect that does not help the candidate much.

Crisp County, Georgia is the highest performing county for both parties. It is a small county with only around 22.000 inhabitants in the southern centre of the peach state. This county has 5 voting districts, one around Cordele for voted Biden with 3 percentage points while the others are more rural and voted for Trump with large margins up of to 90 percentage points. When we look at its influence we can see why it is the highest for both parties. First of all, because Georgia itself is one of the closest races of 2020, secondly it also has a lot of connectedness with Florida, which is also an important swing state.
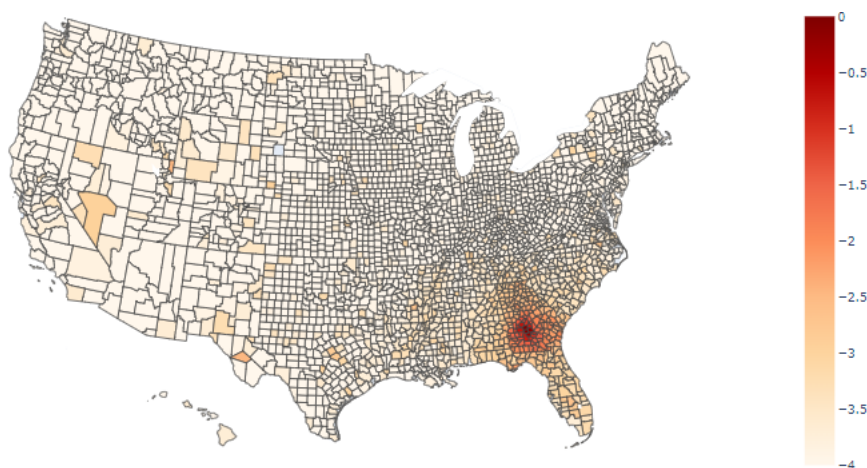

Figure 5: Influence of Crisp County, Georgia

Note that the values are the log10 so -2 means the value is 0.01. For Democrats, the lowest-performing county is Edwards County Illinois has 10,000 inhabitants and is in the rural part of the state. This performance is

17

understandable because Illinois is a very Democratic state that was won by Biden with a difference of almost 17 percentage points. For Republicans, the lowest-performing county is Monroe county Kentucky with also a population of just over 10,000, and in a dark red state that Trump won by 26 points in 2020. Both low-performing counties are in states that are not only nearly certain wins for the candidate but also both have only very little connections with counties in traditional swing states.

4.5   Influence measure and campaigning

Comparison of the campaign and the influence measure is done on a state level. The reason for this is that when a model has no network effects every county in a state has the same influence, because it only influences that state-wide voting count by δ. To calculate the state-level influence for the model with the network, the mean is taken from all counties in that state. This influence measure could be used in practice by a candidate when deciding where to campaign. If we look at the correlation of the last election for visits and influence it seems the candidates might already be using a model of this sort.

Figures 6a and 6b display all states visited by Kamala Harris, Joe and Jill Biden with blue dots. The left y axis is the number of visits divided by population and multiplied by 10,000. This 10,000 is chosen as the average size of the campaign events that can be held on a visit. The red dots are states that are in the top ten based on influence but were not visited. On the x axis, the visited states are ordered by their influence. On the right y axis is the kernel density estimate of the influence measure. This is shown

with the blue line. The vertical red dotted line is the mean of the influence measure.

Figure 6: The relation between influence and campaign visits from Democrats
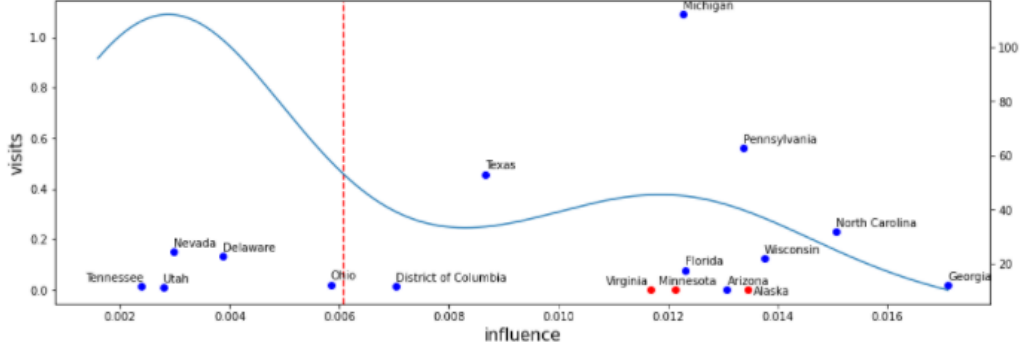
(a) With network effects
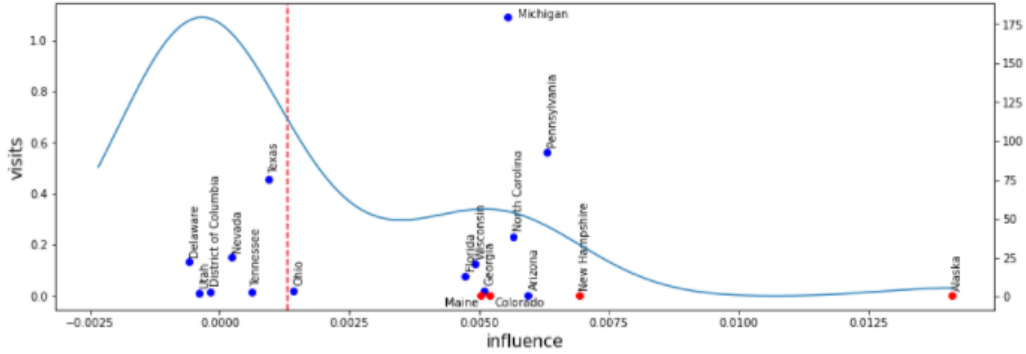
(b) Without network effects

When looking at the influence with network's top ten states, seven were visited by the candidates which include the two most visited states. The states that were in the top ten but were not visited are Alaska which as already explained is a unique case and Minnesota (9th) and Virginia (10th) which are both considered swing states. Compare this to the model without exposure where there are 4 states not visited that are in the top ten. These

are Alaska (1st), New Hampshire (2nd), Colorado (7th), Maine (9th).

When we compare the visited but low influence states we find that they are Delaware, Utah, District of Columbia, Tennessee, Nevada and Ohio. Delaware and the District of Colombia both are solid blue with Democrats winning there consecutively since 1992 and 1964 respectively. These states were most likely visited for fundraisers more than rallying the base, as well as being places where the candidates often are with Delaware being Biden's home state and the Capitol being in D.C.. Ohio and Nevada are both seen as swing states. When we look at their connectedness we can see why they perform low on the influence measure. Ohio is mostly connected to states such as Kentucky, Wyoming and Indiana; all three states lean very heavily to the Republican party. Nevada is connected to mainly pacific states and thus has the opposite problem because California, Washington, Oregon and Hawaii are solid blue states.

5   Conclusion

The goal of this thesis is to show the strength of modelling elections with network components and looking at optimal campaign locations using simulations. The model shows its strength with a high R-squared and the exposure measure being significant and influential. The effects of most features are also in line with earlier research.

The search for influential counties and states has interesting findings. It shows that many states that are already being visited by candidates perform better in a model with network measures than without. Implying that can-

didates already use a model that includes ripple effects of some sort or that conventional already wisdom incorporated these effects via trial and error of previous candidates. Either way, the model still shows some room for improvement with candidates prioritising Ohio and Nevada over states that appear to have more potential like Virginia and Minnesota. Both candidates would also gain from campaigning more in Georgia especially around Crisp County, because this will not only increase their win chance in Georgia but also in Florida. It is also important to note for future candidates that these results are for 2020, and for the run-up to the 2024 elections the priorities might shift and the regression and simulations need to be rerun. This probably won't lead to extreme shifts in the influence of states but for a close election might still make a difference.

There is still room for further research, for example it is possible to focus more on the money trails. A campaign does not consist solely out of campaign visits, radio and television advertising also plays a significant role in the exposure of candidates. These factors could be analysed in a similar manner to the methods used in this thesis.

References

Ambrosius, Joshua D (2016). "Blue City ... Red City? A Comparison of Competing Theories of Core County Outcomes in U.S. Presidential Elections, 2000-2012". eng. In: *Journal of urban affairs* 38.2, pp. 169–195. ISSN: 0735-2166.

Ansolabehere, Stephen, Jonathan Rodden, and James M Snyder (2006). "Purple America". eng. In: *The Journal of economic perspectives* 20.2, pp. 97–118. ISSN: 0895-3309.

Appleman, Eric M. (2021). *Democracy in Action.* URL: https://www.democracyinaction.us/about.html.

Bailey, Michael et al. (2018). "Social Connectedness: Measurement, Determinants, and Effects". eng. In: *The Journal of economic perspectives* 32.3, pp. 259–280. ISSN: 0895-3309.

Barton, Jared, Marco Castillo, and Ragan Petrie (2014). "What Persuades Voters? A Field Experiment on Political Campaigning". eng. In: *The Economic journal (London)* 124.574, F293–F326. ISSN: 0013-0133.

Bishin, Benjamin G and Casey A Klofstad (2012). "The Political Incorporation of Cuban Americans: Why Won't Little Havana Turn Blue?" eng. In: *Political research quarterly* 65.3, pp. 586–599. ISSN: 1065-9129.

Data, MIT Election and Science Lab (2018). *County Presidential Election Returns 2000-2020.* Version V9. DOI: 10.7910/DVN/VOQCHQ. URL: https://doi.org/10.7910/DVN/VOQCHQ.

Junn, Jane (2017). "The Trump majority: white womanhood and the making of female voters in the U.S". eng. In: *Politics, groups identities* 5.2, pp. 343–352. ISSN: 2156-5503.

Kahane, Leo H (2009). "It's the Economy, and Then Some: Modeling the Presidential Vote with State Panel Data". eng. In: *Public choice* 139.3/4, pp. 343–356. ISSN: 0048-5829.

Keuzenkamp, Joep (2021). "MSc Thesis: Polarization, Populism and Political Participation: The Effects of Social Media on American Democracy". In.

Levernier, William and Anthony G Barilla (2006). "The Effect of Region, Demographics, and Economic Characteristics on County-Level Voting Patterns in the 2000 Presidential Election". eng. In: *The Review of regional studies*. The Review of Regional Studies 36.3, pp. 427–47. ISSN: 0048-749X.

Lewis-Beck, Michael S and Martin Paldam (2000). "Economic voting: an introduction". eng. In: *Electoral studies* 19.2, pp. 113–121. ISSN: 0261-3794.

Mihaila, Dan (2021). *Facebook Social Connectedness Index*. URL: https://www.thearda.com/Archive/Files/Descriptions/RCMSCY.asp.

Mixon, Franklin G and J. Matthew Tyrone (2004). "The 'Home Grown' Presidency: empirical evidence on localism in presidential voting, 1972-2000". eng. In: *Applied economics*. Applied Economics 36.16, pp. 1745–1749. ISSN: 0003-6846.

Pew Research Center (2021). *Social Media Fact Sheet*. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/. (accessed: 10.11.2021).

Silver, Nate (2020). *2020 election forecast*. URL: https://projects.fivethirtyeight.com/2020-election-forecast/.

Sioh, Maureen (2018). "The wound of whiteness: Conceptualizing economic convergence as Trauma in the 2016 United States Presidential Election". eng. In: *Geoforum* 95, pp. 112–121. ISSN: 0016-7185.

The Association of Religion Data Archives (2002). *Religious Congregations and Membership Study, 2000 (Counties File)*. URL: `https://www.thearda.com/Archive/Files/Descriptions/RCMSCY.asp`.

The US Department of Agriculture (2020). *Rural-Urban Continuum Codes*. URL: `https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx`.

Tunmibi, Sunday and Wole Olatokun (2021). "Monte Carlo simulation of vote counts from Nigeria presidential elections". eng. In: *Cogent social sciences* 7.1. ISSN: 2331-1886.

U.S. Census Bureau (2021a). *County Intercensal Datasets: 2000-2010*. URL: `https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-counties.html`.

— (2021b). *Educational attainment tables*. URL: `https://www.census.gov/topics/education/educational-attainment/data/tables.2020.List_2016040495.html`.

Yahoo! Finance (2021). *NASDAQ Composite*. URL: `https://finance.yahoo.com/quote/%5C%5EIXIC/history?period1=1477958400&period2=1478649600&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true&guccounter=1`.