Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Jacob Hicks April 14, 2020

Candidate Ideology and Electoral Outcomes in the 2018 U.S. House of Representatives Election: Evidence from Primary Elections and Twitter Ideal-Point Data

by

Jacob Edward Hicks

Dr. Alan Abramowitz
Adviser

Department of Political Science

Dr. Alan Abramowitz

Adviser

Dr. Andra Gillespie

Committee Member

Dean Michael Elliottt

Committee Member

2020

Candidate Ideology and Electoral Outcomes in the 2018 U.S. House of Representatives Election: Evidence from Primary Elections and Twitter Ideal-Point Data

By

Jacob Edward Hicks

Dr. Alan Abramowitz

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Political Science

2020

Abstract

Candidate Ideology and Electoral Outcomes in the 2018 U.S. House of Representatives Election: Evidence from Primary Elections and Twitter Ideal-Point Data

By Jacob Edward Hicks

How does candidate ideology affect electoral success? This seemingly simple question has led to countless books, articles, and other publications. This thesis explores the use of Twitter-based ideology measures developed by Dr. Pablo Barbera. Using a unique dataset comprising 1,212 candidates in the 2018 U.S. House of Representatives Election, I conduct an analysis of electoral outcomes on a scale not previously possible. My results suggest that when compared to the ideology of the primary electorate in which they compete, candidate ideology is not a reliable predictor of success in primary elections. The use and validation of Dr. Barbera's measure of candidate ideology is a significant contribution and further analysis in this area will allow for expanded study of primary elections. Furthermore, this data is a promising avenue for future research into the impact of candidate ideology on electoral outcomes.

Candidate Ideology and Electoral Outcomes in the 2018 U.S. House of Representatives Election: Evidence from Primary Elections and Twitter Ideal-Point Data

By

Jacob Hicks

Dr. Alan Abramowitz

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Political Science

2020

Acknowledgements

I would like to thank my thesis adviser, Dr. Alan Abramowitz, for his support, wisdom, and mentorship throughout this project. This research would never have been possible without his efforts. Additionally, I am grateful to Dr. Beth Reingold and Dr. Jeffrey Staton for their input and feedback, which was instrumental in shaping this project. I would also like to thank Dean Michael Elliott and Dr. Andra Gillespie for serving on my committee and providing me with their expertise. Finally, I would also like to thank Max Krauskopf and Ben Palmer for their support and guidance in completing this project.

Contents

List of Figures

List of Tables

1 Introduction and Research Question

As the United States approaches the 2020 election, there will be no shortage of discussion about polarization and its implications. Despite the ubiquity of polarization as a concept in our national political discourse, there are a number of key questions that remain the subject of debate. Scholars continue to debate the true source of polarization in the United States while also questioning the extent of this polarization. Although it can seem conceptually simple, it is important to establish a uniform definition of polarization, such as that used by Nolan McCarty, which defines polarization as "increasing support for extreme political views relative to centrist or moderate views." (McCarty 2019, 12)

At the intersection of these hotly debated questions is the influence of primary elections. They have been widely proposed as one potential source of polarization in the United States Congress. (Brady, Han, and Pope 2007; Hill and Tausanovitch 2018) At the same time, the influence of the primary electorate has significant ramifications for the extent of polarization in the United States. The impact of the primary election could potentially provide evidence that a polarized electorate is driving trends in elite polarization. Despite their usefulness, effective analysis of primary elections remains elusive due to a lack of widely applicable measures of candidate ideology. In this paper I attempt to address these puzzles by exploring the validity of a new data source developed by NYU's Dr. Pablo Barberá as an avenue to determine whether ideologically extreme candidates achieve greater ideological success in primary elections.

2 Literature Review

Central to any understanding of the impact of primary elections is the ongoing debate around the extent of polarization within the electorate. There are two primary schools of thought within this area. One, proposed by Morris Fiorina, argues that voters are generally moderate and disinterested in politics, but that over time they have become better "sorted" into political parties that most accurately reflect their views. Others, like Alan Abramowitz, directly challenge this assertion and suggest that many voters, especially those who are more inclined to become politically active, are indeed becoming more polarized. (A. I. Abramowitz and Saunders 2008; Fiorina, Abrams, and Pope 2005)

Fiorina and others make the argument that voters are not becoming more ideological based on a variety of compelling statistics and pieces of evidence. For example, he notes that voters in "red" and "blue" states do not diverge significantly on surveys of cultural opinion and public opinion. For example, on issues like LGBTQ+ rights and the role of women, public opinion is converging, not diverging. Other issues, like abortion, have seen overall public opinion remain relatively consistent. In order to explain both these results and the ongoing polarization of officeholders, Fiorina presents the hypothesis that voters are simply aligning themselves with parties that most accurately reflect their ideology. (Fiorina, Abrams, and Pope 2005)

Alternatively, scholars like Abramowitz and Saunders argue that this analysis ignores evidence that suggests a shift is also occurring in the ideology of the public at large. They argue that their study of issue surveys provides crucial evidence that voters indeed are becoming polarized. Support for this argument is derived from a number of sources, particularly the American National Election Survey (ANES). Their

analysis of the ANES suggests that the number of voters in the ideological "center" has decreased from 49 percent in 1972 to 32 percent in 2012. They also find that political issue questions on the ANES can reveal ideological polarization when analyzed for their consistency. They observe this in 24 percent of respondents from 1982-1990 and in 33 percent of respondents from 2002-2004.

In order to explain these results, they propose that the "secular realignment" of American voters occurred in the wake of the civil rights movement and changing political conditions in the American South. They argue that conflicting interests between southern whites, northern whites, and other groups within the Democratic party meant that their success in the South was unsustainable. As Democratic politicians supported the civil rights movement in particular, they alienated southern whites, who were eventually "captured" by the Republican Party. As these electoral coalitions realigned the parties developed a much more comprehensive set of policy positions and ideologies. (A. I. Abramowitz and Saunders 2008)

The distinction between these two theories may seem trivial, leading to the reasonable question of whether this difference between sorting and mass polarization is significant if the end result is the same. The underlying logic of both theories would seem to suggest that ideologically extreme candidates might be rewarded. However, the direction of causality is an essential distinction between the arguments presented by Fiorina and Abramowitz. Fiorina's theory implies that sorting has occurred in response to action by political elites, while Abramowitz's theory would contend that the changes in elite polarization were driven by changes in the ideological makeup of the electorate.

Although some aspects of polarization are subject to debate, there is overwhelming consensus that elected officeholders are deeply polarized. Roll-call voting

data analysis developed by Poole and Rosenthal makes this trend abundantly clear. In Congress, there is practically no overlap between Republicans and Democrats, a significant change from only a few decades ago. Whether measuring inter-party differences or intra-party differences, the evidence for Congressional polarization is overwhelming. Furthermore, their data indicate that the polarization of officeholders shows little sign of abating. (Poole and Rosenthal 2001) In the words of Abramowitz and Saunders, "ideological differences are probably greater now than at any time in the past half century." (A. I. Abramowitz and Saunders 2008, 542)

In developing this data, Poole and Rosenthal utilize their measure known as DW-NOMINATE, or "dynamic, weighted, nominal three-step estimation." Through the use of roll-call votes, they create a one-dimensional linear model of ideological positions on which all Congressional officeholders can be placed. This is traditionally represented using a "left" vs. "right" dimension, which has been the most consistent dimension throughout American history. This metric is highly accurate, widely respected, and has been cited hundreds of times. DW-NOMINATE is a remarkable achievement that has spawned a vast number of derivative works. (Poole and Rosenthal 2001)

However, DW-NOMINATE is not without flaws. Most notable is its inability to account for candidates who have not had a chance to establish a voting record. The candidates whose campaigns are not ultimately successful cannot be measured using roll-call voting data. Simply put, "since most primary challengers lose, we do not have any systematic measures of their ideological positions." (Brady, Han, and Pope 2007, 90) This is a significant number of cases that cannot be studied or analyzed using this method, whose analysis has the potential to reveal additional insights into the relationship between polarization, ideology, and electoral outcomes.

This effect is particularly pronounced in primary elections, where alternative measures may be unavailable. In the absence of roll call data, some scholars have attempted to use other metrics such as organization endorsements. However, these are often not available for primary elections. Thus, the aforementioned effect of the lack of systemic measures is even more pronounced. Even though primary elections are a promising avenue for exploration, "a lack of quantitative data on candidate ideology makes [the claim that the success of extreme candidates in primary elections drives Congressional polarization] difficult to test." (King, Orlando, and Sparks 2015, 406)

One alternative measure was developed by Stanford's Adam Bonica, using campaign finance data. He draws on publicly available campaign finance information and data analytics to create an entirely new way of developing a measure of ideological positioning. Bonica draws over 100 million data points from finance records at all levels of government and in order to create a spatial model of candidate donations, compares the donors and recipients of campaign spending, with the underlying assumption that donors are ideologically consistent. (Bonica 2014)

The ability to use CFScores to understand candidate positioning and electoral success is particularly promising. The work of Ansolabehere et al. that was discussed previously is also supported by Bonica's analysis. He finds that voters do indeed reward candidates who position themselves towards the center, although this is relatively small in scale. In 1996 his analysis suggests that a move towards the center of two standard deviations will correspond with a four percent increase in vote share for Democratic candidates. This analysis of CFScores also supports Ansolabehere et al's finding that the responsiveness of voters to the positions of congressional candidates has decreased over time (Bonica 2014).

Bonica's work shows that there is space for alternative and competing measures of candidate ideology. These new methods allow for the measurement of candidates who could not have been previously measured. Although the use of CFscores has been explored, an increasingly promising avenue is the use of data from social media. The accessibility of Twitter, in particular, allows for low-cost, large-scale analysis of candidate positioning. Within the emerging field of social media analysis within political science, there are a variety of methods that can be used. In this particular case the use of network analysis is highly promising, as it is quite similar to Bonica's CFscores in terms of methodology. Network analysis relies on calculating the differences between a massive matrix of data points representing individual candidates, and in this case Twitter metadata will replace the use of campaign finance data. (Bonica 2014)

Network analysis has been effectively demonstrated by King, Orlando, and Sparks with regards to U.S. Senate elections. In "Ideological Extremity and Success in Primary Elections: Drawing Inferences From the Twitter Network" they develop a matrix of follower relationships between candidates on Twitter. Using this matrix, they are able to develop an ideological dimension that they are then able to compare with electoral outcomes. They also suggest that Twitter data may also be more effective for the study of primary elections compared to campaign finance data. King, Orlando, and Sparks note that financial decisions in primary elections are motivated by additional factors besides ideology, such as party relationships, incumbency, and competitiveness. In their study of Senate primary elections, they find conclusive evidence not only that Twitter can be used as a valid metric of ideology, but that candidate extremity leads to increased vote share in primary elections. (King, Orlando, and Sparks 2015)

Their work provides a significant indicator that Twitter data is a promising avenue for the study of primary elections and they present their work as a significantly step forward for this sort of analysis. However, such an analysis is incomplete and there is significant room for expansion. An expansion of similar analysis to the U.S. House of Representatives would allow for a significantly larger sample, exposing additional avenues for exploration and statistical analysis. Their work has shown that this metric was valid in one particular Senate election year, but an expansion to the House of Representatives would provide significant additional evidence of the measure's validity. The work of Pablo Barberá will allow for this expansion, using open-source Twitter analysis software. Barberá has published a preliminary analysis of a similar network analysis-based method for determining the ideology of candidates. Furthermore, this method has been tested in general elections and shown to be highly correlated with DW-NOMINATE, a metric known to be highly accurate. Nevertheless, the expansion to primary elections in the House of Representatives has not yet been conducted (Barberá 2015). Throughout all of these analyses, there is consistent evidence that candidate ideology has a significant effect on electoral outcomes and vote shares. Although the method for determining candidate positioning may be very new, the underlying theoretical model is not. In addition to supporting the validity of their new measures, each of these pieces of literature provides additional support for the continued relevance of the median voter theorem.

3 Theory

As proposed by Anthony Downs, the median voter theorem draws on economic models and rational choice theory to construct a spatial model of electoral be-

havior. The median voter theorem is conceptually simple, suggesting in its most basic form that given two-party competition candidates will converge on the ideal point of the median voter in order to capture the greatest number of votes (Downs 1957). This simplistic version of the median voter theorem is not empirically supported, as candidates clearly do not converge in the manner that would be predicted by this model. For example, American congressional candidates often have distinct policy positions (Poole and Rosenthal 2001). Although convergence has not been empirically supported, the underlying concept behind the median voter theorem - that voters select candidates based on their ideological preferences - can provide a theoretical foundation for understanding the impact of candidate ideology on electoral outcomes.

For many years there was “very little direct examination of this hypothesized effect,” as Ansolabehere et al. note. However, in recent years new methods and datasets have allowed for the study of this “core prediction of the Downsian model” (Ansolabehere, Snyder, and Stewart 2001). A significant volume of literature has now been developed based upon this theory, suggesting that with additional factors this model may be highly useful for understanding electoral incentives. One very straightforward evolution of the median voter theorem suggests that the “ideal point” located at the “median voter” varies between primary elections and general elections. Evidence from Brady, Han, and Pope indicates that candidates proactively position themselves away from the ideal point within their district and towards the ideal point for their primary electorate. Their evidence, based on an analysis of primary and general election results from 1956 to 1998, suggests that primary voters indeed “favor more ideologically extreme candidates”.

Nevertheless, there is research that supports this particular evolution of the median voter theorem. Evidence for this theory from Westley, Calcagno, and Ault argues that the different ideal point in primary elections is due to the fact that primary voters are more ideologically extreme. The addition of primary elections to our understanding of the median voter theorem thus provides a compelling explanation for the failure of candidates to converge in the "center" as predicted by Downs' original model (Westley, Calcagno, and Ault 2004). Historical analysis also shows that sorting has occurred within primary electorates. A recent paper by Hill and Tausanovitch suggests that the difference in the ideal points between Democratic and Republican primary elections is six times greater in 2012 than it was in 1958. They argue that this is, at least in part, due to the fact that "liberal" voters have become more likely to participate in Democratic primaries and "conservative" voters have become more likely to participate in Republican primaries (Hill and Tausanovitch 2018).

There is also evidence that ideologically extreme candidates are less likely to lose primary elections, which has the potential to force candidates to the ideological extreme in order to prevent a primary challenge. (Brady, Han, and Pope 2007) Just as there is evidence that candidates are rewarded for positioning themselves towards the ideological extreme in primary elections, there is evidence of a "small but persistent benefit" for candidates who position themselves according to the ideal point of their district in general elections, supporting the overall theory of conflicting incentives. As might be expected as polarization increases within the electorate, they find that the responsiveness or change in ideological positioning of candidates to overall district-level preferences has decreased over time (Ansolabehere, Snyder, and Stewart 2001).

One additional reason for the reduction in responsiveness to overall district preferences can be found in the decreasing competitiveness of general elections. This can be seen through district-level results from presidential elections. In the 1976 presidential election, 187 House districts were decided by a margin of less than five percent. By the 2012 presidential election, only 47 districts were won by less than five percent. This significant shift in such a short period of time further underscores the increased significance of the primary election in driving the success of ideologically extreme candidates (A. Abramowitz 2018). One immediate and important question that this raises is whether this trend is simply the result of partisan gerrymandering. However, evidence that candidate extremity can lead to increased vote share in primary elections can also be seen in United States Senate elections, where partisan gerrymandering is not possible. Furthermore, analyses of district competitiveness show that despite the best efforts of many state redistricting commissions, much of this decline in district competitiveness has not immediately followed redistricting. Something larger than partisan gerrymandering must be at play here (A. Abramowitz 2018).

Overall, the median voter theorem clearly suggests that there is significant incentive for candidates to position themselves closer to the ideal point of the primary electorate in order to increase their vote share. Furthermore, this conclusion has been supported by multiple scholars across a variety of analyses. This understanding of electoral incentives not only supports the usefulness of the median voter theorem but has significant implications for the understanding of how candidate ideology affects outcomes in different elections.

3.1 Model

The combination of this supporting literature and existing theory allows for the development of a compelling but simple model for understanding impact of primary elections on candidate ideology and the polarization of the United States House of Representatives more generally. Based upon the primary implications of the median voter theorem, namely that candidates will position themselves according to the preferences of their constituencies, the impact of candidate ideology on electoral success can be understood.

I propose that the model that drives the phenomena behind all of these empirical observations is a simple evolution of the basic median voter theorem: *Given that electoral success is predicted to be determined by a candidate's distance from the ideal point of the median voter, candidates fail to converge at the median voter in a given election due to a conflict between the median voter of the primary electorate and the median voter of the general electorate. The relatively extreme median voter of the primary electorate ultimately rewards extremity.*

3.2 Hypothesis

If this model is accurate, a number of key implications follow that will inform an empirical evaluation of its accuracy. In each case, electoral success is defined simply as the vote share received by each candidate. Candidate ideology is a quantitative measure of a candidate's position on a left-right political spectrum, as determined by an analysis of their Twitter network. Finally, ideological distance is a candidate's distance from the ideal point of voters in a given election.

In a primary election, vote share will be impacted by a candidate's proximity to the ideal point of the primary electorate. If the model is accurate and this implication is upheld, it is expected that in a large-scale analysis – when controlling for factors like electoral structure – there will be a statistically significant and negative relationship between a candidate's ideological distance and their vote share. Note that due to the direction of the measures being used (where negative values are more conservative) we should expect a positive effect among Democrats and a negative effect among Republicans.

4 Methods

In order to conduct a comprehensive analysis of these hypothesized effects, I will build upon previous examinations of electoral success. Previous examinations have been conducted on primary and general elections in the United States Senate, and on primary and general elections in the United States House of Representatives. The effect of candidate ideology on electoral success in general elections in both the House and the Senate have been analyzed using a number of methods to capture ideology, most notably DW-NOMINATE roll call voting data. The effects have also been studied using Bonica's CfScores and Twitter data. However, this is not the case for primary elections (Ansolabehere, Snyder, and Stewart 2001).

4.1 U.S. House of Representatives

For previously discussed reasons, DW-NOMINATE is not a viable measure of candidate ideology if one seeks to compare primary candidates. Furthermore, only

recently has the use of Bonica's CfScores been attempted for primary elections. The use of Twitter data has also only recently been attempted, and for previously discussed reasons is a very promising avenue for potential research. Across all of these methods of measuring ideology, the United States House of Representatives remains a promising unit of analysis. Most immediately and most importantly, the House allows for the largest possible sample size in federal elections in the United States, 435 general elections and 870 major primary elections. In this case, I define major primary elections as those conducted by either the Republican or Democratic party. This sample size will allow for much greater confidence in any statistical tests. Furthermore, this unit of analysis will allow for the largest possible comparison in a single year. One of the most notable limitations of Twitter data is its limited history. Twitter does not maintain publicly-available history of an account's content or network. Given this limitation and the inability to use historical comparisons as a method of establishing causation or maximizing the internal validity of the analysis, it is crucial to ensure the sample size is as large as possible.

A key electoral component that will be necessary to control for is the structure of the primary election. Primaries can take several forms depending on the electoral laws of a given state, limiting the validity of any comparisons as the . Based upon preliminary research, I will separate primary elections into two categories: traditional and jungle. A jungle primary occurs when all candidates compete against one another irregardless of their party affiliation. It will not be possible to directly compare these with traditional primaries, so they will be separated out and discarded. Furthermore, there is evidence that open primaries do not have a significant impact on the ideological

extremity of candidates. For this reason, I will not control for open or closed primaries (Rogowski and Langella 2015).

This sample would then be restricted to primary elections with at least two viable candidates. In the case of primary elections a viable candidate is defined as a candidate who has raised over ten thousand dollars. This requirement stems from the fact that the underlying median voter theorem assumes at least two candidates competing for votes. Furthermore, in a non-competitive or uncontested election, one candidate is likely to receive an overwhelming majority of votes, introducing a significant competing determinant of vote share. With these restrictions in place the estimated sample size, based on preliminary data from the FEC and BallotPedia, is approximately 300 candidates.

4.2 Measuring Electoral Success

The dependent variable across all cases is electoral success, which I define as a given candidate's vote share. Drawing on primary election data from the MIT Election Data + Science Lab, the vote share received by each of the candidates in the 2018 primary elections can be easily collected (MIT Election Data and Science Lab 2017). The next, and far more difficult task, is to match each of these candidate-level results with an ideology score calculated from their Twitter networks.

4.3 Measuring Candidate Ideology

In order to determine the ideology of each candidate in the 2018 primary elections, Twitter analysis developed by Pablo Barberá will be used. This method draws upon existing methods of determining ideology based on finance data (Bonica's CfS-

cores) or roll-call voting data (DW-NOMINATE). Instead of using a set of ideal points in the hundreds (DW-NOMINATE) or thousands (CfScores), Barberá's method includes the estimation of ideal points for hundreds of thousands of Twitter users. Barbera then creates a spatial model where ideology is a latent variable, which is determined by examining which political accounts are followed by a user (Barberá 2015).

There is a variety of evidence developed by Barbera to support the validity of these measures. The most apparent is the comparison between the resulting ideology measure and the commonly understand of ideology for prominent politicians. For example, Bernie Sanders is shown to be far more extreme than most Senators by this measure. In this regard, the measure is proven to be quite accurate. To further support these results, Barbera compares the results of this measure with existing measures such as DW-NOMINATE, which are show to be very similar. Finally, Barbera compares several thousand ideology scores of known voters with their campaign donation data. Here too, Twitter is shown to be an accurate measure of ideology (Barberá 2015). Based upon all of this evidence, I believe there is sufficient reason to assume that Barbera's ideology measure can be used in this manner.

4.4 Measuring Ideological Distance

Establishing a uniform measure of the ideology of voters in primary elections is essential to evaluating the implications of this model. The most common measure of the ideal point for voters in a given district, the previous presidential election, cannot reveal the ideal point for voters in the primary electorate of either party. Although it is widely accepted that primary electorates are more extreme than those of general elections, this difference may not be uniform across regions, states, and congressional districts.

Previous research has simply assumed that primaries will typically reward candidate extremity, but this is a relatively crude measure that ignores the actual implications of the median voter theorem. The true determinant is not simply ideological extremity, but ideological distance. A candidate may be rewarded for being extreme, but they may also be punished for being too extreme. By measuring their distance *in either direction* from the ideal point, a more compelling understanding of the impact of ideology on electoral success can be developed.

In order to measure district ideal points, data from campaign data firm Catalist will be used. Their database of over 200 million voters includes an estimate of voter ideology. Using a subscription provided by the Robert W. Woodruff Library, a 1 percent representative sample can be collected for voters in a given election. For the purposes of this analysis, I will collect the ideology measure of individuals who are registered to vote in either the Democratic primary or Republican primary within the Congressional districts of interest. From these samples I will find the median voter in each election, yielding a median ideology for each of the two elections being studied.

In order to find the ideological distance for each candidate, the difference between their ideology measure and the ideology measure of the electorate must be calculated. These measures have been compared in similar ways before in a variety of different studies and there are several reasons to believe that this is an appropriate usage of the data (Gilens and Page 2014; Ansolabehere and Brian Schaffner 2015). First, both measures are derived from a relative comparison to the entire population within each dataset. Barbera's measure compares ideology using hundreds of thousands of Twitter users and Catalist's measure compares ideology using millions of US voters. Furthermore, both measures are normally distributed and their overall distributions are

nearly identical. Finally, exact comparisons for individual cases can be drawn from the individual ideology measures of the candidates themselves within both the Catalist and Barbera data. This allows for the appropriate scaling model to be determined and has been used to compare Catalist data with DW-NOMINATE data for state legislators (Raja and B.F. Schaffner 2015). In order to create an overall measure of the difference between the candidate an district ideology measures, I will regress the candidate ideology scores on the district ideology scores within each party. For each candidate, the residual resulting from this regression will serve as the distance between their ideology and that of their district. This is a standard way of conducting this comparison and has been used in previous papers, such as Brady, Han, and Pope 2007.

4.5 Controls

Because an observational design is being used, it will be essential to identify and control for four key variables: *jungle primaries, incumbency, spending, and the total number of candidates.* Jungle primaries will be indicated by a categorical variable for candidates competing in Washington state or California. Incumbency will also be identified as a categorical variable, based upon the results of the 2016 congressional election. Fundraising data will be gathered from the FEC, where a ratio of the total amount raised by a candidate compared to the total raised by all competitors within their district in the 2018 primary election will be used. There is compelling evidence that fundraising is a determinant of electoral success in primary elections (Bonica 2017). Because the relative vote share decreases as the number of primary candidates increases, the a count of the total number of candidates in a primary will also be used as a control.

4.6 Analysis

In order to conduct this analysis, a multiple regression model will be used, estimating the impact of ideological distance on electoral success primary elections for each candidate. This relatively simple model will be used on all viable candidates within our sample in all traditional primary elections for the United States House of Representatives in 2018. In this case a quantitative model is essential. A wide variety of factors, many of them unique to a given candidate, can impact an election. Thus, it is important to establish a sample size that is as large as possible and use measures that are as widely applicable as possible. This will allow for the testing of both empirical predictions, and if the regression analysis shows a negative relationship between ideological distance and electoral outcomes as predicted, it will provide support for the proposed theoretical model.

4.7 Comparisons and Establishing Viability

Because one of the most important contributions of this paper is the establishment of Barbera's model as a valid measure of ideology, a comparison to other measures will also be conducted. Given their widely-accepted validity, a comparison will be conducted where possible with DW-NOMINATE scores. If the measure developed by Barbera is valid as expected, then there should be a great deal of correlation between the different measures of ideology.

4.8 Data Collection

The full dataset for this paper comprised 1,212 candidates. The initial data and list of candidates was based on the Federal Elections Commission's *Federal Elections 2018*, a comprehensive list of all House primary and general elections in 2018. Candidates who were deemed viable (by receiving at least 15 percent of the primary vote) were selected, resulting in the full list of 1,212 candidates. Ideology data based on twitter profiles was then compiled and added to the initial data. Twitter profiles were first identified using data scraped from Ballotpedia.com, which maintains information on many congressional candidates. Only a few hundred profiles, primarily incumbents and high-profile candidates were identified. Further automated collection of candidates was done by automatically searching candidate names combined with the keyword "congress" using the Twitter API. A manual search was conducted for any candidates who were not identified using these methods. This resulted in a final list of 1,101 candidates who were present on Twitter. Of these candidates, many did not follow enough users for the estimated ideology scores to be valid. For example, a republican candidate who only follows three "elites" would be skewed significantly to the left if one of them happened to be President Barack Obama. Given this, only candidates who followed at least 200 users were included in the final analysis. Additionally, to ensure that the estimates were reliable they were calculated for each candidate four times and then compared. Any candidate whose total variance was 0.1 or more was also excluded from the final analysis. After removing these potentially unreliable estimates a final count of 813 candidates was reached.

A series of data was then added to allow for control variables and comparison with DW-NOMINATE scores. Financial data was scraped from the Federal Elections

Commission's database of candidates and candidate committees. Primary-only spending was not immediately identifiable from this database, so the data was extracted from candidate committee reports. Primary and authorized committees for each were identified using the FEC API. The Pre-Primary report for each candidate was then collected and the total disbursements were added to the dataset. Some manual collection was required for candidates who had changed their information or the designation of their committees. Replication code is available upon request. The total spending in each race was then identified and a spending ratio was calculated for each candidate to be used as a control variable. Further controls were added for election types, incumbency, and the number of candidates in each race. Each of these has been identified in previous literature as having the potential to significantly impact the vote share received by each candidate. Finally, DW-NOMINATE First-Dimension estimates were collected from data available on Voteview.com for any incumbent candidates.

5 Results

In order to further establish the use of Barbera's Twitter scores as a valid measure of candidate ideology, they were regressed and plotted against the First-Dimension NOMINATE scores when available. Figure 1 demonstrates a significant association between these values and estimates that the Twitter-based ideology scores can account for 83 percent of the variation in the NOMINATE scores. Furthermore a Pearson Correlation test resulted in a statistically significant correlation of 0.912. Combined, this validates the use of Twitter data as a valid measure of candidate ideology when compared to NOMINATE scores.

Measurement Correlation with DW-NOMINATE

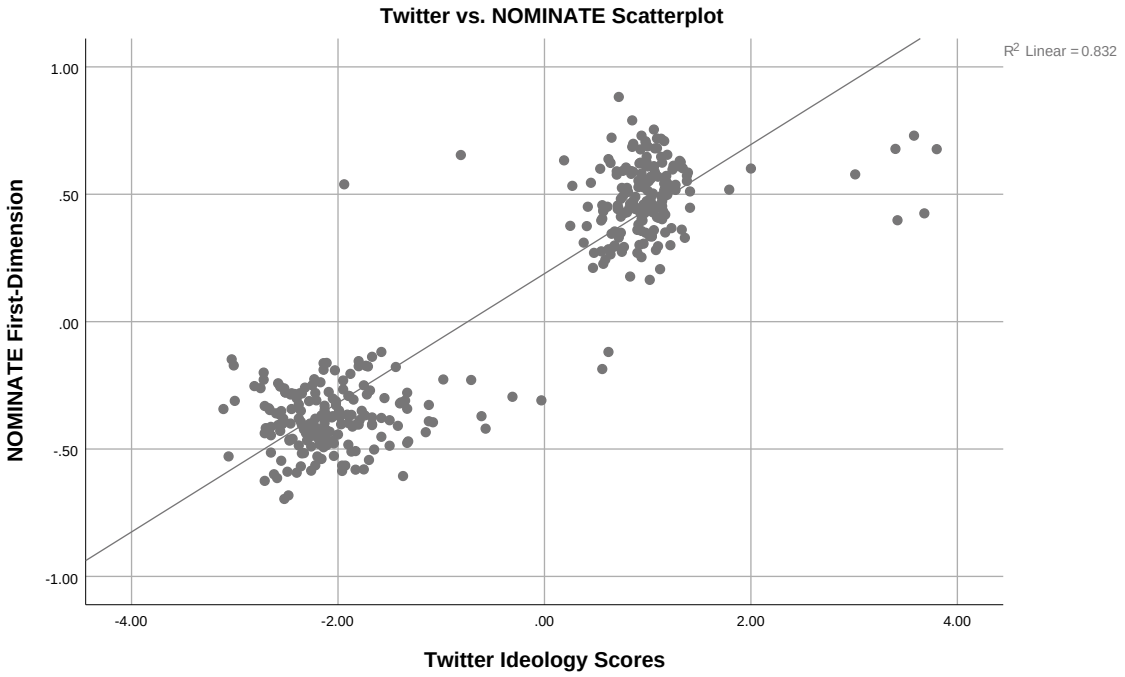

Figure 1: Measurement Correlation with DW-NOMINATE

Table 1 Presents the results of a linear regression model of the distance between candidate ideology and the ideology of their respective primary electorate among Republicans in 2018. The presence of a ”jungle primary” system, the number of candidates, the spending ratio, and incumbency are all shown to have significant effects in the direction predicted. However, our key variable - the distance from their primary constituency - is shown to lack statistical significance. Although the direction is correct, with a p-value of 0.2 we cannot conclude that the difference in candidate ideology is a significant predictor of success in primary elections.

Linear Regression Model (Republicans)

	B	Standard Error	Sig.
Constant	36.791	2.159	.000
Ideological Distance	-1.035	.811	.204
Jungle Primary	-20.802	3.136	.000
Number of Candidates	-2.025	.369	.000
Disbursement Ratio	14.352	2.603	.000
Incumbent	28.661	2.161	.000

Table 1: Linear Regression Model (Republicans)

Table 2 presents the results of a linear regression model of the distance between candidate ideology and the ideology of their respective primary electorate among Democrats in 2018. Here, the same trend continues. The results are in the predicted direction for all variables, and all controls are highly significant. Although the direction is as predicted in this case, the effect is even less significant than among Republicans in 2018.

Linear Regression Model (Democrats)

	B	Standard Error	Sig.
Constant	40.588	2.081	.000
Ideological Distance	0.49	.713	.945
Jungle Primary	-21.775	2.562	.000
Number of Candidates	-3.369	.412	.000
Disbursement Ratio	25.611	2.226	.000
Incumbent	26.277	2.155	.000

Table 2: Linear Regression Model (Democrats)

6 Discussion

With these results, it is clear that the model cannot definitively show that candidate ideology is a significant predictor of electoral outcomes in primary elections. However, there is still significant value in this method of analysis. Having shown that Barbera's method of calculating ideology from Twitter scores is strongly correlated with DW-NOMINATE, a widely-accepted measure, it is likely that this method of determining candidate ideology is accurate. Existing research has shown that Twitter can be a valid and accurate way of measuring candidate ideology. These results would seem to indicate that while previous research has shown that primaries reward ideology, this effect is due to the relatively extreme ideology of primary electorates. Although it is not possible to prove a negative, these results suggest that when compared to the ideology of the primary electorate in which they compete, candidate ideology is simply not a reliable predictor of success in primary elections.

Many reasonable explanations for these results exist. The simplest, and perhaps most compelling, is that primary voters are simply not capable of making informed decisions about the relative ideology of candidates when voting in primaries. However, it must be acknowledged that this paper incorporates several novel uses of uncommon data sources, and only covers a single midterm election. There is compelling evidence that these measures are reliable, but there is significant risk in using novel sources of data. Further examination and expansion of this analysis into the 2020 election would be a significant step towards validating the results of this research.

7 Conclusion

How does candidate ideology affect electoral success? This seemingly simple question has led to countless books, articles, and other publications. This paper contributes a novel method of analyzing primary elections that allows for a scale not previously possible. The use and validation of Dr. Barbera's measure of candidate ideology is a significant contribution and it is my hope that further analysis in this area will allow for expanded study of primary elections which previously could not be studied. Ultimately, although the results of my analysis preclude a greater conclusion about the impact of polarized primary electorates, I believe that the collection of this data is a promising avenue for future research into the impact of candidate ideology on electoral outcomes.

Bibliography

Abramowitz, A. (2018). *The Great Alignment: Race, Party Transformation, and the Rise of Donald Trump*. Yale University Press. ISBN: 978-0-300-20713-2. URL: https://books.google.com/books?id=ix9dDwAAQBAJ.

Abramowitz, Alan I. and Kyle L. Saunders (2008). "Is Polarization a Myth?" In: *The Journal of Politics* 70.2, pp. 542–555. ISSN: 00223816, 14682508. DOI: 10.1017/s0022381608080493. URL: http://www.jstor.org/stable/10.1017/s0022381608080493.

Ansolabehere, Stephen and Brian Schaffner (2015). "Beyond the Core and Periphery: A New Look at Voter Participation Across Elections". In:

Ansolabehere, Stephen, James M. Snyder, and Charles Stewart (2001). "Candidate Positioning in U.S. House Elections". In: *American Journal of Political Science* 45.1, pp. 136–159. ISSN: 00925853, 15405907. DOI: 10.2307/2669364. URL: http://www.jstor.org/stable/2669364.

Barberá, Pablo (2015). "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data". In: *Political Analysis* 23.1, pp. 76–91. ISSN: 1047-1987. DOI: 10.1093/pan/mpu011. URL: https://www.cambridge.org/core/article/birds-of-the-same-feather-tweet-together-bayesian-ideal-point-estimation-using-twitter-data/91E37205F69AEA32EF27F12563DC2A0A.

Bonica, Adam (2014). "Mapping the Ideological Marketplace". In: *American Journal of Political Science* 58.2, pp. 367–386. DOI: 10.1111/ajps.12062. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12062.

— (2017). "Professional Networks, Early Fundraising, and Electoral Success". In: *Election Law Journal: Rules, Politics, and Policy* 16.1, pp. 153–171. DOI: 10.1089/elj.2016.0413. URL: https://doi.org/10.1089/elj.2016.0413.

Brady, David W., Hahrie Han, and Jeremy C. Pope (2007). “Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?” In: *Legislative Studies Quarterly* 32.1, pp. 79–105. ISSN: 03629805. URL: http://www.jstor.org/stable/40263411.

Downs, Anthony (1957). “An Economic Theory of Political Action in a Democracy”. In: *Journal of Political Economy* 65.2, pp. 135–150. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1827369.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy C. Pope (Nov. 19, 2005). *Culture War? The Myth of a Polarized America*. 2 edition. New York: Longman. 256 pp. ISBN: 978-0-321-36606-1.

Gilens, Martin and Benjamin I. Page (2014). “Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens”. In: *Perspectives on Politics* 12.3, pp. 564–581. ISSN: 1537-5927. DOI: 10.1017/S1537592714001595. URL: https://www.cambridge.org/core/article/testing-theories-of-american-politics-elites-interest-groups-and-average-citizens/62327F513959D0A304D4893B382B992B.

Hill, Seth J. and Chris Tausanovitch (July 1, 2018). “Southern realignment, party sorting, and the polarization of American primary electorates, 1958–2012”. In: *Public Choice* 176.1, pp. 107–132. ISSN: 1573-7101. DOI: 10.1007/s11127-017-0478-0. URL: https://doi.org/10.1007/s11127-017-0478-0.

King, Aaron S., Frank J. Orlando, and David B. Sparks (Aug. 13, 2015). “Ideological Extremity and Success in Primary Elections: Drawing Inferences From the Twitter Network”. In: *Social Science Computer Review* 34.4, pp. 395–415. ISSN: 0894-4393. DOI: 10.1177/0894439315595483. URL: https://doi.org/10.1177/0894439315595483 (visited on 10/16/2019).

McCarty, N. (2019). *Polarization: What Everyone Needs to Know®*. What Everyone Needs to Know Series. Oxford University Press, Incorporated. ISBN: 978-0-19-086778-2. URL: https://books.google.com/books?id=TguXDwAAQBAJ.

MIT Election Data and Science Lab (2017). *U.S. House 1976–2018*. Ed. by MIT Election Data and Science Lab. DOI: 10.7910/DVN/IG0UN2. URL: https://doi.org/10.7910/DVN/IG0UN2.

Poole, Keith T. and Howard Rosenthal (2001). "D-Nominate after 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting". In: *Legislative Studies Quarterly* 26.1, pp. 5–29. ISSN: 03629805. DOI: 10.2307/440401. URL: http://www.jstor.org/stable/440401.

Raja, R.J.L. and B.F. Schaffner (2015). *Campaign Finance and Political Polarization: When Purists Prevail*. University of Michigan Press. ISBN: 978-0-472-05299-8. URL: https://books.google.com/books?id=KZDJCgAAQBAJ.

Rogowski, Jon C. and Stephanie Langella (2015). "Primary Systems and Candidate Ideology: Evidence From Federal and State Legislative Elections". In: *American Politics Research* 43.5, pp. 846–871. DOI: 10.1177/1532673X14555177. URL: https://doi.org/10.1177/1532673X14555177.

Westley, Christopher, Peter T. Calcagno, and Richard Ault (2004). "Primary Election Systems and Candidate Deviation". In: *Eastern Economic Journal* 30.3, pp. 365–376. ISSN: 00945056, 19394632. URL: http://www.jstor.org/stable/40326400.