SIMULATING ELECTIONS IN THE CONTEXT OF VOTER BEHAVIOR AND ELECTION RULES

Michael E. Hammer[a] and Hamdi Kavak[a]

[a]Department of Computational and Data Sciences, George Mason University, USA
{mhammer4, hkavak}@gmu.edu

ABSTRACT

Electoral systems manage complex social phenomena, including representation, consent, social unrest, political stability, economic development, and many others. One enduring problem for scholars and policymakers is determining the optimal design of single-member legislative districts in order to maximize representation and minimize social conflict. This paper contributes to the general theory and practice of election simulations by demonstrating a method for identifying voter ideological ideal points in a microsimulation model using real election data. This model is calibrated with a sample of voters in five state-level legislative districts using real election results from the 2020 U.S. general election. The paper demonstrates that real election results can be replicated by modeling emergent agent preferences using simple agent characteristics and election rules. The value of this replication is then in using the synthetic electorate to test changes in voter behavior and election results under different election rules.

Keywords: elections, agent-based modeling, political science.

1 INTRODUCTION

Electoral systems are made of a set of institutions wherein social phenomena -- including representation, consent, social unrest, political stability, economic development, and many others -- are interconnected. By definition, the single-member legislative districts as used in the United States (referred to at different times as a plurality, majoritarian, or Westminster system) require the election of one member of Congress per one congressional district. Understanding the rules under which boundaries are set and elections are conducted are essential for understanding the social and political processes that emerge in a majoritarian system. Additional constraints on voting and redistricting have been a major focus of U.S. state and federal legislators. Voting access bills, in many cases, are designed to restrict voting via mail or requiring additional voter identification, and the proliferation of non-competitive legislative districts have led to various reform efforts by both voters and legislators.

Modeling and simulation can be used to assess the feasibility of a variety of rules and reforms, and thereby contributing to defining a fair legislative district. While "fair" legislative districts in U.S. elections may be desirable, scholars have struggled to provide a method for choosing such districts from a vast number of potential choices and, indeed, even on how to define "fairness" in practice. Additionally, it is difficult, with only empirical data, to model the interaction between boundary rules and choice rules. In other words, legislative boundaries (boundary rules) and primary election rules (choice rules) are often studied separately, and changes to these rules are often implemented separately in the real world.

This paper argues that agent-based modeling can provide a method for redistricting based on emergent voter preferences through simulated scenarios in which boundary rules and choice rules interact. We present a model as the first step of a comprehensive computational framework to study elections in the U.S. Section 2 of this paper provides the theoretical and conceptual background of the U.S. election

Proc. of the 2024 Annual Simulation Conference (ANNSIM'24), May 20-23, 2024, American University, DC, USA
C. Ruiz-Martin, B. Oakes and R. Cárdenas, eds.

system and how it was studied in the past. Section 3 introduces the model and its main components. The results are presented in Section 4, while the paper is concluded in Section 5 with discussion.

2 BACKGROUND

2.1 U.S. Electoral System

Understanding the rules under which boundaries are set and elections are conducted are essential for understanding the social and political processes that emerge in a majoritarian system. The discussion below is grouped into the loose categories of boundary rules and choice rules, following the language of the Institutional Analysis and Design Framework [1]. These rules collectively govern the legislative process in majoritarian regimes.

2.1.1 Boundary Rules

The dilemma of drawing fair legislative districts arises because of the possibility of offering one party an advantage in district boundaries across an entire state or region. The question of fairness arises in any scenario in which the partition of a state into legislative boundaries creates the opportunity for one party to gain an electoral advantage over another. These district boundary choices subsequently influence voters' choices in a given election and, collectively, determine the composition of state and national legislative bodies. The district creation process is delegated to the states by the U.S. Constitution but is subject to exogenous constraints through federal statute and judicial rulings. Before attempting to improve the redistricting process, it is imperative to understand several foundational rules governing it.

District Size. U.S. Supreme Court rulings in the 1960s established a rule of redistricting generally referred to as "one-person-one-vote" [4]-[8]. This foundational principle is implemented via the constitutional mechanism of guaranteeing roughly equal numbers of persons in congressional districts. This principle requires the reallocation of congressional districts and the redrawing of district boundaries (**i.e.,** redistricting) every ten years in order to guarantee as closely as possible the number of districts by state. The decennial time period is required by the U.S. Constitution and is the original purpose for administering the U.S. national census every decade. The one-person-one-vote rule applies to state legislative districts as well as to the U.S. Congress. In practice, this principle requires states to achieve the smallest possible variation in congressional and state legislative district size.

Number of Candidates. Majoritarian legislative districts create a special type of political conflict. A majoritarian electoral system is one in which a candidate wins with the largest share of votes, even if this largest share is not the majority [9]. However, such systems typically have two major candidates competing within a given race. Despite potential issues that this may create for voter engagement and representation, the two-party feature of majoritarian systems that dominate, especially in the U.S., reduces the computation of voter choice to a binary choice. In practice, a binary option in a free and fair election, by definition, means that a majority of voters have chosen the winning candidate.

2.1.2 Choice Rules

The rules of the general election in a given majoritarian system are very clear: voters cast a vote for only one candidate, and the candidate with the most votes wins. There are some variations between systems, even between American states, but this general framework is sufficient for this paper.

However, candidate selection rules in majoritarian systems define the choices voters will have in the general election. Before the progressive era reforms in the early 20th century, candidates were selected by party elites and put forward to the general election in the United Kingdom and the United States. Almost all American states now use a primary election, allowing voters to directly choose the candidates for each party to be put forward.

However, a wide variety of choice rules are implemented in these primaries. As a result, those concerned with election reform often focus on reforming the primary election choice rules. Most American states use either a closed primary or open primary system. These systems share one feature: a voter may vote for only one choice of one party's slate of candidates. What then makes a primary "closed" or "open" is defined by whether or not a voter is required to be a registered member of that party to participate in that party's primary election.

2.2 Mathematical and Statistical Models of Voting Reform

Mathematical, machine learning, and statistical models can successfully produce a vast ensemble of redistricting plans that comply with both the foundational and reformist constraints on redistricting.

The computational simulation of legislative districts was first undertaken over fifty years ago. Thoresen and Littschwager [15] developed an early computational simulation of congressional districts. This effort, despite the computational limitations of the time, produced one-hundred-fifty redistricting plans that conformed to the one-person-one-vote rule. More recent models have changed focus to developing a range of plans that comply with decision-making parameters that require redistricting plans to meet one or more "fairness" criteria. One criterion is the seats-votes (S/V) curve and the variant inverted seats-votes curve [16], [17] where the S/V graphs "plot the predicted number of seats S versus the statewide percentage of votes won V for all districts combined" [18, p.348]. Despite debate on how to properly calculate the S/V metric, the underlying concept is straightforward: if there are ten Congressional elections in State *i*, and the state has 60% of voters who lean Republican and 40% who lean Democrat, the S/V graph visually represents the difference between the predicted number of Republican seats (6) and the real election outcomes.

[18]-[20] produce simulated redistricting plans for Pennsylvania, Florida and Maryland, respectively, that use similar criteria for fairness. Nagle [18] and Chen and Rodden [19] similarly analyze the fairness of plans implemented in those states. Cho and Liu [20] use a proposed "unfair" plan as the baseline against which to measure a minimum degree of partisan efficiency. These simulations and partisan bias measures have developed a wider range of possible solutions and of measuring partisan bias but have done little more in terms of determining an optimal redistricting plan.

Additionally, *gap efficiency* was first proposed as a metric by Stephanopolous and McGhee [21]. This concept argues that the fairness of districts should be measured by "the difference between the parties' wasted votes divided by the total number of votes cast in the election" [*ibid*, p.851]. "Wasted vote" is defined as both "lost" and "surplus" votes [*ibid*]; consequently, gap efficiency calls upon boundary makers to achieve the voter preference gap as consistently as possible across districts rather than by creating "safe seats" that are lopsidedly in favor of one party but that still achieve correlation to the statewide voter preference.

Other scholars have focused on how to measure plans against various fairness criteria. DeFord et al. [22] attempt to resolve the stubbornness of the redistricting problem by Monte Carlo sampling of an ensemble of redistricting plans rather than continuing the search for a mathematically optimal solution. Even putting aside optimization, however, DeFord *et al*. note that it is difficult, if not impossible, to find a plausible redistricting solution that meets fairness measures without violating one of the three foundational constraints of equality, compactness, and one-person-one-vote.

In this body of research, as well as the actual work of redistricting, voters are generally analyzed by voting precinct. Precinct blocks are moved between districts to test the effect of different partitions of the full electorate. Agent-based modeling offers an alternative method for considering voters in the redistricting process.

However, the possibility of discovering a single optimal district for a given redistricting problem is generally argued to be a mathematically impossible task. Keung, Mixon, and Villar [13] provide mathematical proof to demonstrate that choosing a fair map from a range of legal redistricting maps is

NP-hard by definition. A similar argument is made by Alexeev and Mixon [14]. If these assessments are correct, a single optimal district cannot be identified mathematically. Consequently, computational models have instead achieved the goal of identifying a range of redistricting plans that comply with the foundational constraints. The issue, however, is that existing redistricting simulations have produced thousands or even millions of redistricting plan scenarios that can achieve the required constraints.

2.3 Agent-Based Modeling of Voting Behavior

Agent-based modeling is an alternative for identifying voter behavior under alternative boundary and choice rules. As shown above, the nature of the redistricting problem lends itself to analyses that approach the problem through graph partition, boundary optimization, and cellular automata methods. Currently, no agent-based models of the U.S.'s single-member legislative districts analyze these plans against district fairness measures. However, recent efforts to develop agent-based models of competition between political parties and voter behavior provide theoretical insights useful to redistricting studies.

Laver and Sergenti [23] developed an agent-based model of political party competition for voters. They model the competition of political parties for voters based on the mathematical distance between the voter's randomly-assigned ideal point and the party's ideological ideal point in a multi-party simulation. The party competition within the model examines voter choice under a variety of simulated scenarios in which parties adjust their boundaries to fit voter preferences. Throughout this paper, as in Laver and Sergenti, *ideal point* is used to refer to a given voter's bundle of desired policy outcomes.

Baltz et al. [24] and Melbane et al. [25] implement an agent-based model to identify strategic voting behavior for the purpose of detecting election fraud. Their work captures strategic voting behavior through ABMs to differentiate it more clearly from statistical anomaly detection models. Their models illustrate that individual voting preferences and behaviors can be modeled in a way that differentiates the outcomes of the model from statistical methods that use aggregate voting data. This insight is directly relevant because aggregate historical voting data is used for both real-world redistricting plans and for the computational modeling of fairness measures.

Silver [26] developed an ABM of electoral college voting by congressional district. This model applies to U.S. presidential voting rather than to legislative voting, but it explores the important issue of how agents engage in strategic decisions when voting is made in a more compact election district versus a larger (statewide) district. That includes the boundary of the agent's actions as a key factor in decision-making, which is directly relevant to this paper.

Kononovicius [27] uses an agent-based "herding" model to recreate the empirical distribution of three Lithuanian parliamentary (multi-party) elections. The agents in this model are recruited by each party with a linear "recruitment function." The agents respond to this function deterministically but do so based on a stochastic "switching probability" and an "idiosyncratic behavior term" that defines that particular agent's likelihood to switch parties. Baltz [28] models the results of agent voting behavior under different election regimes in the UK, Canada, New Zealand, and the United States.

ABMs have obvious strengths related to the potential for modeling the information sharing, cognitive structure, and actions of voters. A more difficult challenge is to determine the usefulness of an agent-level model as a tool for a political activity that is regularly conducted with aggregate data focused only on historical voting patterns.

2.4 Spatial Modeling of Political Ideology

Majoritarian legislative districts create a special type of political conflict: a candidate wins with the largest share of votes, even if this largest share is not the majority [9; this phenomenon is why the phrase "plurality" is sometimes used in place of "majoritarian"]. Despite the real-world issues that this may create for voter engagement and representation, the two-party feature of majoritarian systems dominate especially in the U.S. and it reduces the computation of voting to a binary choice. In practice, a binary choice in a free and fair election, by definition, means that a majority of voters have chosen the winning candidate.

A voting model requires quantifying the ideological distance between voters and the two candidates. Consequently, spatial voting models have been a feature of the political science landscape since the mid-20th century. One influential spatial model of election behavior is Downs' Median Voter Theory (MVT hereafter) [29]. MVT argues that in a majoritarian, single representative district, the victorious candidate will be the one whose ideology most closely represents the median voter [30].

The NOMINATE scoring system is one method used for over three decades for scoring legislator ideal points [31], [32]. This system assigns a scaled ideology score to each (current and past) federal legislator according to their roll call votes in Congress. These votes are factored into two dimensions, social and economic, based on the content of the vote in question. The NOMINATE method treats legislators as voters, with each roll call vote on legislation as a measure of a conservative or liberal vote on that given issue. Roll call votes are then collectively used to identify the legislator's two-dimensional ideal points using a Markov chain method [31]. As a result, each congressperson is assigned an ordered pair of graph coordinates, allowing for all representatives to be plotted in a continuous Euclidean space according to their two-dimensional ideal points on social and economic issues. NOMINATE scores have been assigned for every member of Congress in U.S. history based on their roll call votes [33].

The NOMINATE method has also been applied to public officials other than members of Congress. For example, Bertelli and Grose [34] use U.S. executive branch public statements to assign NOMINATE scores to appointed executive officials. Since executive branch officials do not participate in roll call votes, public statements and proposed federal rules are used as proxies for identifying an official's ideological preferences. As with the original NOMINATE score, these scores are then graphed in two-dimensional Euclidean space, illustrated in Figure 1.

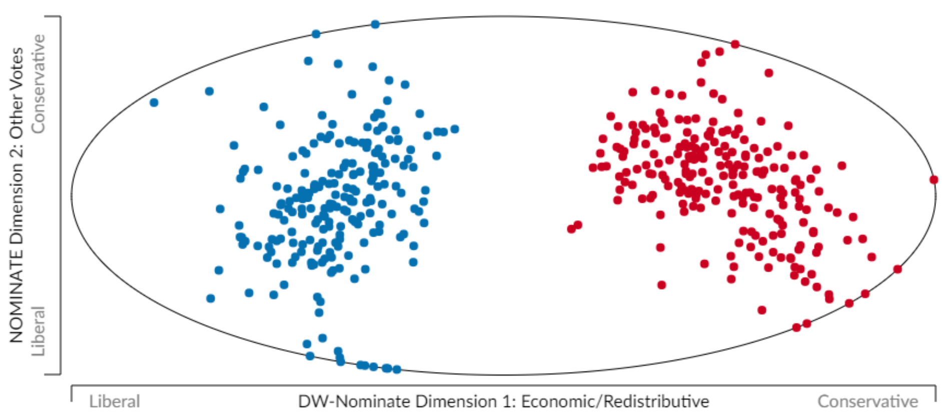

Figure 1. The NOMINATE scores of the 118th U.S. House of Representatives (Voteview 2023).

This paper argues that the NOMINATE two-dimensional ideology scoring system is a theoretically sound method for assigning agent-level policy ideal points. The method for assigning scores in this research is discussed in the next section. The goal of focusing on this particular voter attribute is to ensure a reliable data-based synthetic electorate that can provide meaningful insight into how voters would behave under a variety of institutional constraints.

3 MODEL OVERVIEW AND INITIALIZATION

This paper contributes to the theory and practice of legislative redistricting by demonstrating a method for identifying voter ideological ideal points in a microsimulation model using real election data. This model is calibrated based on voters in five state-level legislative districts using real election results from the 2020 U.S. general election. This model is intended to explore one potential method for identifying voters' preferences before then using those defined preferences to model alternative election scenarios.

3.1 Entities

The model includes two classes of agents: *voters* and *candidates*. The number of voters is determined by the number of votes cast in the 2020 election in the five state legislative districts used to calibrate the model. There are two candidates for every legislative district.

3.2 State Variables

Ideology. Voters and candidates are assigned two ideology scores based on the NOMINATE scoring system, representing ideology on a two-dimensional graph. This long-standing scoring system compresses the full range of ideology into a two-point range, with -1.0 representing the most liberal possible position and +1.0 representing the most conservative possible position. The x-axis (dimension 1) represents votes on economic policy, whereas the y-axis (dimension 2) represents "other" votes considered to be social policy. These two dimensions are represented in the model by *Dim1* and *Dim2*, respectively. The criteria for assigning NOMINATE scores for voters and candidates are discussed in-depth in section 3.4. Agents' scores are fixed throughout the candidate vote in each time-step. However, scores vary between time-steps according to the rules specified below.

Party Affiliation. The real number of party identification is determined according to the voter registration data for the five legislative districts.

District. Voters and candidates are all assigned one of five districts. The total number of voters in each district is determined according to the voter registration data for the five legislative districts. This is the only spatial position for agents in this model.

Turnout. In every election, the actual number of voters is less than the total number of voters in a given electorate. The turnout variable determines whether or not a given agent participates in the election in that given time step. The turnout rate is determined by the data and is implemented randomly in each time step by drawing a sample from all available voters and assigning them the value of turnout = 1. Only voters with that turnout indicator participate in the election. It is important to note that the empirical reality of who participates in elections is determined by several agent characteristics not included in this model. This is an important limitation to note and should be considered for future model versions.

3.3 Input Data

The five state districts in this model are drawn from within Tulsa County (with one district combining with part of adjacent Osage County) in Oklahoma (OK). Not all districts from the county were used because Oklahoma conducts elections for state representatives every four years and staggers half of the districts in biennial elections (i.e., half of state representative elections in 2020 and half in 2022). These particular districts are attractive because Oklahoma voter registration and elections data are easily available and the districts themselves collectively represent a partisan balance [34], [35].

In 2020, two of these districts were strongly Republican, two were strongly Democratic, and one was evenly divided. The population of these districts is large enough to allow for variation but small enough to be manageable within NetLogo. Additionally, Oklahoma is a state that requires voters to declare a party (or no party) in their voter registration and shows party affiliation in voter turnout data.

There were 114,135 registered voters in these five districts in 2020 [34]. This is the total voter population initialized in the model. The state voter turnout rate of 69% was applied across each of the five districts, resulting in a total of 78,753 active voters across all districts in each run of the model [35]. This is the second stochastic element of the model in that the blend of voters by party varies with each model run.

3.4 Initialization

Voter Party Assignment. Voters were assigned a party according to one of three categories: Republican, Democrat, and Independent. The state also allows for voters to register as Libertarians. These voters were grouped with Independents. These party affiliations for each voter agent are assigned according to the real percentages of party affiliation according to voter registration data for each district [35]. These party affiliations are used as the starting point for identifying the ideology preference scores used to motivate voting behavior in the model.

Candidate Ideology. Candidate NOMINATE scores are assumed to align with similar candidates' scores. In other words, this paper assumes that Republican and Democratic candidates within a given region will generally have the same ideological profile as a reflection of the underlying political culture of their region. To be more specific, this assumption allows for the NOMINATE scores of Republican state-level candidates to be inferred from the ideological profile of the region's federal Republican incumbents. This assumption is useful because the NOMINATE scores already exist for federal incumbents. Since the legislative districts used in this model reside within OK Congressional District 1 (CD 1), the NOMINATE scores for the CD 1 incumbent and the neighboring Congressional District 2 (CD 2) are used from a possible range of scores for the Republican candidates. The scores of these two Republican incumbents are used to establish the range of potential NOMINATE scores for the Republican candidates in the model.

For Democratic candidates, the model draws on historical NOMINATE scores, which show that Oklahoma Democratic legislators, like other legislators in the American South, are more conservative than other Democrats nationally [37]. Consequently, this model considers NOMINATE scores of the Democratic candidates using those of the two most economically conservative Democrats in the 2020 U.S. House of Representatives.

The scores listed in Table 1 are a naïve blend of the available range of candidate scores. There is no specific knowledge of the relationship between these particular scores and the voter preferences of that particular district.

Table 1. Candidate Ideal Points by Party Identification and NOMINATE Dimension.

District	Candidate	Candidate NOMINATE Scores (dim1 / dim2)
66	Cand 1 (D)	-0.14 / 0.29
	Cand 2 (R)	0.541 /0.227
68	Cand 3 (D)	-0.085 / 0.41
	Cand 4 (R)	0.69 / -0.048
71	Cand 5 (D)	-0.085 / 0.41
	Cand 6 (R)	0.69 / -0.048
78	Cand 7 (D)	-0.14 / 0.29
	Cand 8 (R)	0.541 /0.227
79	Cand 9 (D)	-0.085 / 0.41
	Cand 10 (R)	0.541 /0.227

Voter Ideology. NOMINATE scores were assigned to voters according to one of several criteria, including their party affiliation. The problem of determining voters' utility function, and whether or not voters attempt to maximize their votes rationally, has been a vexing and unresolved problem for political scientists [37]. This model honors only the simple rule that agents will vote for the candidate closest to their own ideological ideal points. To this end, the model assumes low-information voters are primarily driven by partisan identification.

Accordingly, the model applies the concepts of affective polarization and party identification [36, cf. 37] to voting behavior. The implication of these theories for the model, Voter NOMINATE scores are clustered around, but not perfectly aligned with, the score of the candidate with whom they share party affiliation.

Lastly, Independent voters, who represent a significant percentage of the electorate in this sample, are initialized using several concepts in the existing voting behavior literature. Independents are known to generally have consistently liberal or conservative tendencies in their voting despite not having a party affiliation [38] However, Independents are the most likely of any voting bloc to switch parties between elections [39, 40]. Consequently, the model treats Independents as the most ideologically flexible group. These observations are implemented by centering the NOMINATE scores around 0 (the midpoint of the scale) with a wide range for Independents.

3.5 Process Overview

3.5.1 First Round Calibration Method

Candidate NOMINATE scores are held constant for each model run using the values in Table 1. One advantage of agent-based simulations is that all possible combinations of variable values can be tested. The range of values for voters in each party is tested at a consistent increment, noted in Table 2, for each district. These ranges are the full ranges of values assigned to candidates in Table 1. Each district is calibrated separately since this model does not include agent interactions across districts.

Table 2. Voter NOMINATE Bounds by Party Identification and NOMINATE Dimension.

Party	***Dim1* range, increment**	***Dim2* range, increment**
Democrat	(0.14) - (0.085), (0.3)	0.29 – 0.41, (0.3)
Independent	(0.2) - 0.2, (0.1)	(0.1) – 0.41 (0.1)
Republican	0.541 - 0.69, (0.3)	(.048) – 0.227 (0.3)

This first round of calibration is done by a parameter sweep, allowing potential agent NOMINATE values to be tested in combination with all others. The parameter sweep varies the value of each dimension score by the increments in the table above. Using the parameter values and increments above, each district has 714 possible combinations of values. These combinations are fixed within each time step since each time step represents an election. The values are varied between time steps.

3.5.2 Second Round Calibration

The parameter sweep treats each party's voting bloc as homogeneous. In other words, the parameter sweep will (with every combination of scores) assign the same scores to every agent within a given party. This method is referred to subsequently in this paper as the "first round" of calibration. This homogeneity may closely represent real election results in districts in which there is high party loyalty. However, voters in other districts may demonstrate more heterogeneity in their NOMINATE scores even within one party.

A second calibration method to account for greater agent heterogeneity uses a given ideal point within the NOMINATE ranges above (in Table 2) and assigns that as the mean score in a random normal distribution of voters for that party and a standard deviation that allows the entire distribution of values to be spread across the whole range of values in Table 2. The specific ideal points used in this second step are drawn from the NOMINATE values that appear most frequently for each party in the first of round calibration. Within each range, voter ideology scores are tested using this new ideal point within the range as the mean for a random normal distribution.

3.5.3 Election Method

The election mechanism is a procedure wherein the voter will vote for the candidate who is ideologically closest to them on both of the NOMINATE dimensions. Thus, the model tests voting behavior and election outcomes with the two NOMINATE dimensions (*nominate_dim1* and *nominate_dim2*). The ideological distance between a given voter and the two possible candidates of their choosing is calculated (within each district) as the shortest distance of either sqrt [(voter$_i$ *nominate_dim1* – C1 *nominate_dim1*)^2 + (voter$_i$ *nominate_dim2* – C1 *nominate_dim2*)^2] or sqrt [(voter$_i$ *nominate_dim1* – C2 nominate_dim1)^2 + (voter$_i$ *nominate_dim2* – C2 *nominate_dim2*)^2]. The winner of the race is determined by a simple count of the number of voter choices made via the formulae above.

4 RESULTS

The parameter sweep using the uniform ranges above yielded election results that ranged within an absolute value of 0.91% and 5.19% of the real election results for each district. While plausible, these results demonstrated that it is be important to leverage additional agent characteristics in order to develop more realistic outcomes. This was especially true in District 68, District 71, and District 78 (see Table 4). The second calibration method was then implemented using a range of ideal points identified in the best model runs in the first round (the values displayed in Table 3).

Table 3. Most Frequent Ideal Points in First Calibration Run.

Party	***Dim1*** **common points**	***Dim2*** **common points**
Democrat	-0.14, -0.085	0.29, 0.30
Independent	-0.20, 0.10	-0.10, 0.10, 0.41
Republican	0.69, 0.541, 0.30	-0.048, 0.227, 0.30

A standard deviation is shown for second-round calibration because the ideal point is now used as a mean in the normal distribution. The results of this second round are assigned using the election results averaged over 50 model runs. Table 4 quantifies the best results from the two calibration methods. The first calibration method uses scores from the model run (out of the 714 combinations) that produced election results closest to the real-world vote total. The second calibration method uses the scores that generated an election result using the method described above.

Table 4. Final Calibrated NOMINATE Scores by Original Districts.

District	**Voter NOMINATE Scores (sd)**	**Election round (First or Second)**	**1st Method Error**	**2nd Method Error**
66	R: dim1 0.69 / dim2 -0.048 D: dim1 -0.14 / dim2 0.29 I: dim1 0.2 / dim2 -0.1	First	**1.44%**	9.54%
68	R: dim1 0.69 (0.1) / dim2 -0.048 (0.1) D: dim1 -0.085 (0.2) / dim2 0.29 (0.2) I: dim1 0.3 (0.2) / dim2 0.0 (0.2)	Second	4.15%	**0.31%**
71	R: dim1 0.69 (0.1) / dim2 -0.048 (0.1) D: dim1 -0.14 (0.2) / dim2 0.29 (0.2) I: dim1 0.1 (0.2) / dim2 0.0 (0.2)	Second	5.19%	**1.61%**
78	R: dim1 0.69 (0.1) / dim2 -0.048 (0.1) D: dim1 -0.14 (0.2) / dim2 0.29 (0.2) I: dim1 0.1 (0.2) / dim2 0.0 (0.2)	Second	2.68%	**2.23%**
79	R: dim1 0.69 / dim2 -0.048 D: dim1 0.3 / dim2 0.3 I: dim1 -0.2 / dim2 0.41	First	**0.91%**	4.73%

The error columns show the absolute value of the difference between the first round's best simulation and the real results and then the difference between the average of the second round's simulations and the real results. Figure 2 visualizes the combination of the best results of the two calibration methods shown in Table 4. This figure helps visualize how the error figures are calculated; each candidate's simulated vote percentage is placed next to the real vote percentage received in 2020. The error represents the difference in real and simulated vote percentages. Absolute values are used because that value is the same for each candidate. In other words, in District 66, 1.44% represents that one candidate won by 1.44% more votes and one candidate lost by 1.44% more votes in the simulation than in the real election.

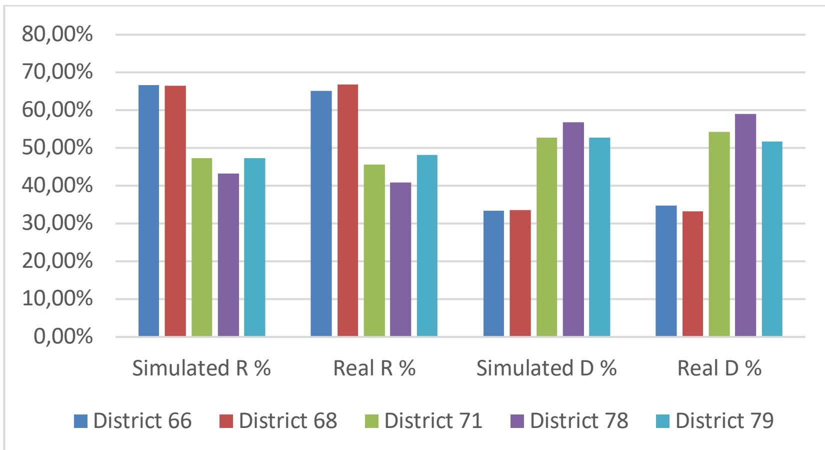

Figure 2. Simulated 2020 Election Results versus Real Results.

5 DISCUSSION AND FUTURE WORK

This model has produced exploratory data in a simulated environment with behaviorally limited agents. We argue that the voter values here meet the test of face validity because they are very similar to what we expected based on previous voting behavior research. Several interesting findings and questions result from data created by implementing existing methods to quantify the ideological distance between voters and candidates. While the similarity of the simulated and real election results shows that simulations can be fit to real data with relatively few rules, the value of those results is the creation of a synthetic electorate to model a variety of candidates and election institutions. The election results are the outcome measure, but successfully identifying voter preferences is important. In order to test the validity of voter preference values, the method presented here would need several additional steps of validation. An important test is to demonstrate the usefulness of the synthetic electorate in an out-of-sample experiment. One method for this would be to simulate the results of other votes in the same election. Since 2020 was also a presidential election year, it would be possible to simulate the results of congressional and presidential elections as an additional test.

One additional issue to address in future work is the importance of accurately assigning candidate scores. The scores here were assigned on the basis of theory and a qualitative assessment of closest neighbors. However, in larger, subsequent studies, it is important to find a reliable method for quantifying candidate NOMINATE scores since these scores form the foundation of grouping voter ideal points in the method used here.

After further refining the model fit with these out-of-sample tests, the next step is to use the model to identify how election results would change under various counterfactual scenarios with voting rules that address the various boundary and choice rules discussed earlier. A more rigorous test would be to model the performance of the synthetic electorate in a subsequent election, such as the 2022 congressional election. This test would add the difficulty of simulating changes in the electorate through migration, death, and newly of-age voters, as well as different voter turnout rates. Future work will investigate the fairness of existing redistricting plans and past changes. This framework will provide a more objective view of redistricting and fair elections.

REFERENCES

[1] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge, UK: Cambridge Univ. Press, 1991.

[2] State of New Hampshire, “State Government Overview,” Accessed 12 January 2024. [Online]. Available: https://www.nh.gov/almanac/government.htm.

[3] California State Assembly, “Members.” Accessed 12 January 2024. [Online]. Available: https://www.assembly.ca.gov/assemblymembers.

[4] Baker v. Carr. No.6, 369 US 186, (U.S. Supreme Court). Mar. 26, 1962.

[5] Gray v. Sanders. No. 112, 372 U.S. 368, (U.S. Supreme Court). Mar. 18, 1963.

[6] Reynolds v. Sims. No. 23, 377 US 533, (U.S. Supreme Court). June 15, 1964.

[7] Wesberry v. Sanders. No. 22, 376 US 1, (U.S. Supreme Court). Feb 17, 1964.

[8] Avery v. Midland County. No. 39, 390 U.S. 474, (U.S. Supreme Court). Apr 1, 1968.

[9] J. Bernauer and A. Vatter, “Can’t Get No Satisfaction with the Westminster Model? Winners, Losers, and the Effects of Consensual and Direct Democratic Institutions on Satisfaction with Democracy,” *Eur. J. of Political Res.*, vol. 51, no. 4, pp.435-468, June 2012.

[10] D. Polsby, and R. Popper, “The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering,” *Yale Law & Policy Rev*., vol. 9, no. 20, pp. 301-353, 1991.

[11] Shelby County v Holder. 2013. 570 U.S. 529.

[12] J. De Benedictis-Kessner “Evidence in Voting Rights Act Litigation: Producing Accurate Estimates of Racial Voting Patterns,” *Election Law J.*, vol. 14, no. 4, pp. 361-381, Dec. 2015.

[13] K. Kueng, D. Mixon, S. Villar, “Fair Redistricting is Hard.” *Theoretical Comp. Sci*., vol. 791, pp.28-35, 29 Oct. 2019.

[14] Alexeev, B. and D. Mixon, “An Impossibility Theory for Gerrymandering,” *arXiv: 1710.04193v2*, 2017.

[15] J. Thoreson and J. Littschwager, ‘‘Computers in Behavioral Science: Legislative Districting by Computer Simulation,’’ *Behavioral Sci*., vol. 12, pp. 237–247, 1967.

[16] G. King, “Representation through Legislative Redistricting: A Stochastic Model,” *Am. J. of Political Sci*., vol. 33, no.4, pp.787-824, Nov. 1989.

[17] G. King and R. Browning, "Democratic Representation and Partisan Bias in Congressional Elections." *Am. Political Sci. Rev*., vol. 81, no. 4, pp. 1251-1273, Dec. 1987.

[18] J.F. Nagle, “Measures of Partisan Bias for Legislating Fair Elections,” *Election Law J*., vol. 14, no. 4, pp.346-360, Dec. 2015.

[19] J. Chen and J. Rodden, ‘‘Cutting through the Thicket: Redistricting Simulations and the Detection of Partisan Gerrymanders.’’ *Election Law J.*, vol. 14, no. 4, pp. 331–345, 2015.

[20] T. Cho and Y. Liu, “Toward A Talismanic Redistricting Tool: A Computational Method for Identifying Extreme Redistricting Plans,” *Election Law J.* vol. 15, no. 4, pp. 351-366, Dec. 2016.

[21] N.O. Stephanopoulos, N.O. and E.M. McGhee, ‘‘Partisan Gerrymandering and the Efficiency Gap.’’ *Univ. of Chicago Law Rev.*, vol. 82, no. 2, pp. 831-900, Spring 2015.

[22] D. DeFord, M. Duchin and J. Solomon, “A Computational Approach to Measuring Vote Elasticity and Competitiveness.” *Stat. and Public Policy*, vol. 7, no. 1, pp. 69-86, 2020.

[23] M. Laver and E. Sergenti, *Party Competition: An Agent-Based Model.* Princeton, NJ: Princeton Univ. Press, 2012.

[24] S. Baltz, K. McAlister, A. Pineda, F. Vasslai, W.R. Mebane, “Agent-Based Models of Strategic Electoral Behavior in Election Forensics,” presented at the 2018 Annual Meeting of Midwest Political Sci. Assoc., Chicago, IL., USA, April 5-8, 2018.

[25] W.R. Mebane, S. Baltz, F. Vasselai, "Using Agent-Based Models to Simulate Strategic Behavior in Elections," presented at the PolMeth XXXVI, Massachusetts Institute of Technology, Cambridge, MA, USA, July 18-20, 2019.

[26] J. Silver, "Electoral College: Emergent Battlegrounds. An Agent-Based Model of Campaign Behavior Change with District Allocation of Electors," in *Proc. 2019 Int. Conf. Computational Social Sci. Soc. of the Americas*, Cham, CH: Springer, pp. 171-189.

[27] A. Kononovicius, "Empirical Analysis and Agent-Based Modeling of the Lithuanian Parliamentary Elections," *Complexity*, pp., 1-15, 2017.

[28] S. Baltz, "Computer Simulations of Elections, with Applications to Understanding Electoral System Reform. Dissertation," Ph.D. dissertation, Dept. of Political Sci. and Sci. Comp, Univ. Michigan, 2022.

[29] A. Downs, "An Economic Theory of Political Action in a Democracy," *J. of Political Econ*., vol. 65, no. 2, pp.135-150, April 1957.

[30] R. Holcombe, "The Median Voter Theory in Public Choice Theory," *Public Choice*, vol. 61, no. 2, pp.115-125, May 1989.

[31] K.Poole, *Spatial Models of Parliamentary Voting*. New York, NY, USA: Cambridge Univ. Press.

[32] K. Poole and H. Rosenthal, *Ideology and Congress: A Political Economic History of Roll Call Voting*, New York, NY, USA: Taylor & Francis, 2017.

[33] Voteview, "House of Representatives." [Online]. Accessed 1 December 2022. Available: https://voteview.com/congress/house.

[34] A. Bertelli and C. Grose, ''Secretaries of pork? A new theory of distributive public policy,'' *J. of Politics*, vol. 71, no. 3, pp.926–945, July 2009.

[35] Oklahoma State Election Board, "Voter Registration Statistics." [Online]. Accessed 1 December 2022. Available: https://oklahoma.gov/elections/voter-registration/voter-registration-statistics.html.

[36] Oklahoma State Election Board, "2020 Election Results ." [Online]. Accessed 1 December 2022. Available: https://oklahoma.gov/elections/elections-results/election-results/2020-election-results.html.

[37] N. Dias and Y. Lelkes, "The Nature of Affective Polarization: Disentangling Policy Disagreement from Partisan Identity," *Am. J. of Political Sci*., vol. 66, no. 3, pp.775-790, July 2022.

[38] Voteview, "Democratic Party." [Online]. Accessed 2 January 2024. Available: https://voteview.com/parties/100/democratic-party.

[39] T. Reilly and D. Hunting, "The Fluid Voter: Exploring Independent Voting Patterns Over Time," Politics & Policy, vol. 51, pp.59-90, 2023.

[40] T. Reilly and E.C. Hedberg, "Social Networks of Independents and Partisans: Are Independents A Moderating Force?" Politics & Policy, vol. 50, pp.225-243, 2022.

AUTHOR BIOGRAPHIES

MICHAEL E. HAMMER is a doctoral student in the Computational Social Science program at George Mason University. His research interests are in modeling & simulation related to public policy and administration (including electoral systems, resource management regimes, and political communication) and organization science. His email address is mhammer4@gmu.edu.

HAMDI KAVAK is an Assistant Professor in the Computational and Data Sciences Department and co-director of the Center for Social Complexity at George Mason University. His research interests lie at the intersection of data science and modeling & simulation. His email address is hkavak@gmu.edu and website address is www.hamdikavak.com.