Chapter 2
The Gerrymandering Model: A Monte Carlo Simulation

To simulate the effects of redistricting institutions on electoral outcomes for use in this paper and subsequent chapters, I have built a model of electoral competition that allows for adjusting various parameters to reflect different conditions with respect to voter ideology, voter welfare, partisan trends, candidate polarization, and districting. The model is built on the foundation of a model from Keith Krehbiel, Adam Meirowitz, and Tom Romer, significantly altered and expanded in order to simulate the effects of gerrymandering institutions, and to measure the effect of various electoral conditions on differing dimensions of democratic representation. The chapter will briefly introduce the source for the foundation of the model in Part I, detail the new “Gerrymandering Model” used throughout the dissertation in Part II, and discuss how this model is used and adapted throughout subsequent Chapters in Part III.

I. Foundation: The “KMR” Model”

The inspiration for the model used throughout the dissertation was originally developed in the 2005 paper “Parties in Elections, Parties in Government, and Partisan Bias” by Krehbiel, Meirowitz, and Romer (hereinafter the KMR model), modified and expanded to analyze gerrymandering and voter welfare. This model tackles the “analytically unwieldy” issue of partisan influence on both election outcomes and legislative policy outcomes by alternately including and excluding several conditions relating to party and voter behavior, each of which amends common formal theory assumptions about voter rationality. Through a Monte Carlo simulation of resulting legislatures under various conditions, the authors find that policy
outcomes only deviate from the median preference when several of these conditions are included in the model. The electoral conditions are as follows:

- **Separation of Candidates:** When turned “off”, each party’s candidate has preferences the same as the district median. When “on”, the parties run candidates equidistant from, but on opposite sides of, the median. The condition reflects the reality that candidates from the parties are pressured (or prefer to) to take distinct stances, but that they still somewhat adjust these stances to the ideology of their district.

- **Ideology-Based Partisanship:** When turned “on”, some voters identify strongly with a particular party, causing them to be more likely to vote for that party regardless of the ideology of that party’s candidate. The result is that each party has a different chance of winning in each district, depending on the preference of the district median. This condition might represent enduring party ID-based voting, or expectations about party performance in the legislature regardless of individual candidate platforms.

- **Electoral Tides:** When turned “on”, this condition creates a shift across districts increasing the chance that one party will win. The increase in the party’s chance is constant across districts. This represents the possibility that one party may be placed at a national advantage or disadvantage as a result of things like past performance or scandal.

Along with the legislative party condition, each of these three can be turned “on” or “off” to create 16 different combinations for which the authors each simulate an outcome. Each of these conditions is operationalized as a single parameter in the model. For each “on” condition, a parameter value is chosen (somewhat arbitrarily, although the authors provide comparative statics for these parameters), and legislatures are simulated for each “political environment” combination.
The authors find that multiple conditions must be simultaneously turned “on” in order to generate partisan bias in the legislature (that is, a policy outcome different from the median voter preferences). Their model is unusual in allowing what are essentially assumptions about rationality and subconscious incentives in voting to be toggled, and particularly useful in its flexibility. The paper also incorporates a normative element, by measuring “district median dissatisfaction” for each environment.

My model, hereinafter the Gerrymandering model, involves several significant changes to the KMR model. Most importantly, my model includes a new gerrymander module, and significantly expands the scope of voter welfare measures to allow for a wide-ranging normative analysis. Additionally, it modifies the basic electoral function to allow more flexibility, and changes the way in which candidates polarize around a district’s median voter to simulate this aspect more realistically (particularly so that candidates are never located wildly exterior to the entire electorate). The entire model is detailed in the next section.

II. The Gerrymandering Model

Given a set of voters with ideologies along a single dimension and several other parameter values, the Gerrymandering model works as follows:

1) Gerrymander: Given a gerrymander parameter $\gamma$ and a number of districts $d$, each voter is assigned a district and the median voter in each district is determined.

2) Candidate positions: Two candidates, one from the D party and one from the R party, run in the election in each district. Candidates position themselves symmetrically around the median voter in each district according to a polarization parameter $\delta$. The particular value of the $\delta$ parameter is not relevant to the simulations in this chapter.
3) Election: An election is held in each district, where the probability of a given candidate winning is determined as a function of the median voter’s ideology, along with three other parameters as described below.

4) Legislature: This final module calculates the utility that each voter gets from the legislature under four different welfare standards.

For a given set of parameter values, the model is iterated in MATLAB over several thousand Monte Carlo simulations of resulting legislatures to yield average partisan compositions (and average policy and voter welfare outcomes in other applications) under various electoral conditions.

**Voter Ideologies:** The model takes a sorted vector of voter ideology scores. The ideology of each voter is always identified along a single dimension, and we shall assume that positive ideologies represent conservatives that generally prefer Republican candidates, while negative ideologies are liberals that prefer Democratic candidates (the probabilistic translation of ideologies into votes is described in the “Election” section below). For all simulations excluding the modified version of the model in the VRA chapters, voter ideologies are assumed to be uniformly distributed \( U\left[-\frac{n-1}{2}, \frac{n-1}{2}\right]\) with a mean of 0 and a range of \(n-1\) (where \(n\) is the number of voters). For convenience, I set \(n = 435\) in all simulations, created a uniform and symmetric distribution of voters with ideologies ranging \([-217, 217]\).

**Gerrymander:** This module, entirely new to the Gerrymandering model, assigns all voters in the population to a district based on a gerrymandering parameter \(\gamma\). There are no “geographic” or other constraints as to which district a voter can be assigned to. The \(\gamma\) parameter is an integer
between 0 and \( d \) (the number of districts). \( \gamma \) represents the number of districts that will be ideologically packed, so if \( \gamma = 1 \), the \( n/d \) voters with the most liberal ideology will be assigned to a single district, while all other voters will be spread in a balanced way among all districts. I.e. the gerrymander will create one “safe Democratic” district and the remaining districts will lean slightly toward the R party. The result is that a value of \( \gamma = 0 \) represents a nonpartisan gerrymander where all districts are microcosms of the state as a whole to the greatest degree possible. A value of \( \gamma = d \) represents a bipartisan gerrymander where all voters are packed into districts with other voters of similar ideologies, and legislators of both parties are protected. For the sake of clarity, I assume that partisan gerrymanders are controlled by the Republican party, and that negative ideology values are preferred by the Democratic party and positive ones preferred by the Republican party.\(^1\) Thus, \( 0 < \gamma < (d/2) \) represents a partisan gerrymander favorable to the Republican party in which voters with Democratic-leaning ideologies are packed into \( \gamma \) districts, while moderate and Republican-leaning ideology voters are spread around the remaining \( (d - \gamma) \) districts, giving a majority of districts a median voter favorable to Republicans. Values \( (d/2) < \gamma < d \) are also possible, but do not represent gerrymanders that would realistically be in the interest of any institution; I would expect the results from these gerrymanders to be similar to those under \( \gamma = d \).

All simulations throughout the dissertation assign \( d = 15 \), representing a medium-sized state with 15 districts (analogous to states like New Jersey, North Carolina, or Michigan). Each district is assigned 29 voters through the gerrymandering algorithm. Specifically, the algorithm first assigns voters to the first \( \gamma \) districts, sequentially taking the most liberal voter still

---

\(^1\) Average utility results for Democratic partisan gerrymanders are generated by reversing the tide parameter for a given Republican gerrymander, as long as the population is symmetric. When the asymmetric population extension is used in section IV.E, this asymmetry must also be reversed to simulate Democratic gerrymanders.
unassigned and placing them in the first district not yet filled. So when $\gamma = 2$, the algorithm takes the 29 most liberal voters and places them in the first district, and then takes the 29 next-most liberal voters and places them in the second district. When $\gamma = d = 15$, all voters will be assigned this way, so that the fifteenth district will be assigned the 29 most conservative voters in the state.

After this part of the process is completed, the algorithm fills in the remaining $d - \gamma$ districts with the remaining voters. This is done by taking the $d - \gamma$ most liberal remaining voters, and assigning one sequentially to each of the remaining districts. It then takes the next batch of $d - \gamma$ most liberal remaining voters and assigns one to each of the remaining districts in reverse order. The order is then reversed again for each remaining batch of voters until each district is filled, and each voter assigned so that all the non-ideologically packed districts are as evenly balanced as possible. The model also identifies the ideology of the median voter in each district, as this will be crucial to the election function.

**Candidate Positions:** Two candidates, one Republican ("R") and one Democrat ("D") run in the election, and each adopts an ideological position. This position is a function of the ideology of the median voter in the district (which we will call $x$), the absolute value of the most extreme voter in the vote universe (assigned $K$ in the original KMR model, in this case $\frac{n-1}{2} = 217$), and an exogenous party polarization parameter ($\delta$, ranging from 0 to 100).

The position of D is:

$$D = \frac{\delta}{100} (-K) + (1 - \frac{\delta}{100})x$$
And the position of R is:

\[ R = \frac{\delta}{100}(K) + (1 - \frac{\delta}{100})x \]

So when \( \delta = 0 \) (no polarization), both candidates locate at the ideology of the median voter in all districts. As \( \delta \) increases, candidates move progressively away from the district median, toward the respective poles of the voters’ ideological range. When \( \delta = 100 \), all Democratic candidates locate at the left-most extreme of the population, while all Republican candidates locate at the right-most extreme, regardless of the ideologies of the voters in the district.

As a further example, consider \( \delta = 40 \) and \( K = 100 \), so ideologies range from -100 to 100. In a district with \( x = 0 \) (median voter is an exact moderate), the D and R would locate symmetrically at -30 and 30 respectively. In a district with \( x = 40 \) (a somewhat conservative district), R would be very conservative, locating at 64 while D would be a liberal-leaning moderate, locating at -16. If the district median were the most-conservative \( x = 100 \), R would locate at 100 and D would locate at 20 (a legitimately right-of-center conservative Democrat).

**Election:** This function takes an input of district medians, and values for three parameters \( \alpha \), \( \rho \), and \( \tau \), and randomly generates a vector of legislator ideologies with \( d \) members. The probability that R wins a given district in the model is:

\[ p_{R\text{win}}(x) = 0.5 + (\alpha - 0.5) \left( \frac{x}{K} + \frac{\tau}{100} \right)^{1/\rho} \]

As in the polarization module, \( K \) is the absolute value of the ideological range of the voter universe, and acts to normalize this range to \([-1,1]\), while \( x \) is the ideology of the median voter. \( \tau \)
is the global partisan tide parameter which is the same across all districts; when \( \tau \) is negative, an ideologically moderate \((x = 0)\) voter is more likely to vote for the Democratic candidate, while when \( \tau \) is positive, the same voter is more likely to vote for the Republican candidate.\(^2\) The \( \alpha \) parameter, a holdover from the KMR model is a measure of the degree of ideological partisanship of the voting population (i.e. the weight that voters put on partisan labels, as opposed to individual candidate ideology, in deciding their vote). For all simulations across the dissertation, I assign \( \alpha = 1 \).

Finally, the \( \rho \) parameter, new to the Gerrymandering model, represents the extent to which the seat share in the electoral system is sensitive to the voting results. The original KMR model does not use this parameter, but might be thought of as a specific case where \( \rho = 1 \).

Following the research of Taagepera and Shugart, I use \( \rho = 3 \) for the simulations in this paper, which increases the marginal effect of both partisan tides and changes in the median voter when the median is closer to 0, with little effect when the median is close to \(-K\) or \(K\). The choice of \( \rho = 3 \) was made because under a nonpartisan gerrymander, it will generate the cube-root seats-votes curve commonly observed in single-member district elections over the range of \( \tau \) values.\(^3\)

Thus, when \( x \) and \( \tau \) are 0, the election is won by each party 50% of the time; as \( x \) and/or \( \tau \) increase, the probability of electing R increases according to a cube-root function. In cases where the function implies that the probability of R winning is greater than 100% (or less than 0%), R always wins (or D always wins).

---

\(^2\) \( \tau \) is designed to be analogous to the national partisan tide variable in the empirical chapters. Thus, \( \tau = 10 \) reflects a 10% partisan tide, or an environment where the 55\(^{th}\) percentile most liberal voter votes Republican 50% of the time, or a 55%-45% GOP win in the national popular vote.

\(^3\) Taagepera and Shugart (1989) provide a formal justification for both the use of \( \rho = 3 \) in this context, and for variations in \( \rho \) based on district size and electoral system. In response to Davis v. Bandemer (106 S.Ct 2797 (1986)), Browning and King (1987) advocate for the use of this “bilogit” model, with an unconstrained \( \rho \), as a standard against which to measure the partisan bias in a gerrymander.
**Voter Utility:** Once the election has been held in each district, the resulting utility of every voter is calculated under four different measures. This module is also new to the Gerrymandering model. All utility measures are calculated using the negative linear distance of the voter’s ideology from the legislator(s) of interest. With respect to interaction with tides, a voter’s ideology is the sum of their ideology from the original distribution and the partisan tides \( \tau \) parameter. Utilities that are less negative (closer to 0) represent greater voter satisfaction.

- **Policy Median:** A voter’s Policy Median Utility is their distance from the policy that would result if the median member of the legislator controls this policy outcome. So for a 15-member legislature, this is the distance between the voter’s ideal and the ideal of the 8th most liberal (or conservative) member of the legislature.

- **Personal:** A voter’s Personal Utility is the extent of their agreement with the legislator who personally represents them; i.e. the distance between their ideal and the ideal of the legislator elected from their district.

- **Compositional:** A voter’s Compositional Utility is their overall satisfaction with the entire composition of the legislator. It is calculated as the mean distance between the voter’s ideal and the ideal of every member of the legislature. (Note: this is not the voter’s distance from the average legislator; each of \( d \) distances is calculated independently and then averaged.)

---

4 Note that the original KMR model does not modify a voter’s ideology ideal for changes in partisan tides in calculating voter utility. The change was made conceiving of the tides parameter as representative of a sincere, if temporary, shift in public opinion (i.e. a strongly positive \( \tau \) value represents an election where voters do actually desire more conservative policies, as opposed to an exogenous shock whereby all voters vote more Republican for some reason despite no change in their underlying policy preferences). This modification is also necessary if we are to model the responsiveness of an electoral system to changes in public opinion. Candidate positions are still determined by the underlying ideological distribution of the district and the polarization parameter \( \delta \), without reference to \( \tau \).
- **Discursive:** Discursive Utility is conceptualized as the extent to which the voter feels that their point of view is perfectly represented by *some* legislator (i.e. their interest are present in the legislature), regardless of which district this legislator is in, or which party controls the body. It is calculated as the shortest distance between the voter’s ideal and any legislator.

For most results related to voter welfare, average utility is calculated across the entire range of voters for each of the four norms. The model never averages, sums, or compares values across different norms.

## III. How is the Model Used?

Each of the next three chapters of the dissertation address a different set of questions in redistricting research, and thus each manipulates a different set of variables and aspects of the model:

- Chapter 3 models the impact of partisan tides and wave elections on the party composition of the legislature under different districting regimes; thus, this chapter varies $\gamma$ (gerrymander) and $\tau$ (tide), while holding $\delta$ (polarization) constant, along with the ideological distribution of the population.

- Chapter 4 models the impact of the Voting Rights Act amendments on legislative composition within a state with a significant and ideologically discrete minority population. So this simulation portion of this chapter varies $\tau$ and the composition of the population, but also amends the gerrymandering module to allow for majority-minority districting requirements.
Chapter 5 models the impact of districting regimes and polarization on the measures of norms of representation, and thus varies $\gamma$, $\delta$, and (to a lesser extent) $\tau$, while incorporating the voter welfare module for the first time. This chapter concludes with a supplement to model majority-minority districting as in Chapter 4.

Note that $\alpha = 1$ and $\rho = 3$ are held constant throughout all chapters, as are $n = 435$ (the number of voters) and $d = 15$ (the number of districts). Each chapter details exactly which values for each parameter, along with the number of simulation iterations, were used to generate the results presented. The simulation results in each of these chapters are backed by empirical evidence from election data, survey results, and/or case studies.