Implementation of statewide election data to examine fairness of South Carolina district maps: A comparative analysis of approaches for approximating results in uncontested races

Alfie-Louise Brownless[*]   Project Advisor: Dr. Anne Catllá[†]

July 2021

Abstract

After each census, researchers analyze election data to provide information relevant to the redistricting process. South Carolina is among a collection of states which face certain issues regarding election analysis of fairness due to the presence of a large percentage of uncontested races. Although uncontested results are known to create analysis challenges, there is not a universal consensus on how to best handle these situations. Here we explore quantification of partisan fairness and the impact of using statewide election county-level data as a proxy for estimating uncontested results. We develop a district approximation method using statewide elections at the county scale and use known metrics to qualitatively and quantitatively evaluate resulting election characteristics in historical and simulated election contexts. The same metrics were then used to perform a thorough comparative analysis of other common approximation methods. We find county-level election data to be an effective tool in approximating uncontested elections by providing evidence to support the notion that county-level data is effective under multiple election conditions. Furthermore, analysis of different approximation methods show how measures of partisan fairness for a particular election can change based upon a particular approximation method, potentially affecting future interpretations of uncontested election results.

1 Introduction

As a representative democracy, the United States divides larger areas of the country into districts, where the residential population has the ability to vote for those who will represent them in federal or state governments. The design of these districts must follow both state and federal guidelines. Federal guidelines include maintaining nearly equal populations in districts and complying with the Voting Rights Act [6]. Additionally, at the state level, guidance may include maintenance of compactness, preserving communities of interest, and protecting incumbents, which often further complicate the procedure for drawing new and updated districts [6].

Extreme partisan unfairness resulting from a particular district plan via gerrymandering can occur even under seemingly ideal circumstances, and so it is valuable to have quantitative tools to analyze the fairness of districting plans [8]. Gerrymandering is the process of drawing district lines to artificially favor a particular group or party, and is a strong indicating factor used in determining electoral fairness. This often accomplished by "packing" and "cracking", where one

---

[*]Wofford College, 429 N Church St. Spartanburg SC, 29303 USA, `brownlessar@email.wofford.edu`

[†]Wofford College, Department of Mathematics, 429 N Church St. Spartanburg SC, 29303 USA, `catllaaj@wofford.edu`

party is "packed" in great numbers into a small number of districts while the remainder of party members are "cracked" throughout the remaining districts, such that the majority of districts within the state are given in artificial favor of the other party [21].

This report aims to quantitatively explore partisan fairness, specifically in regards to South Carolina, and to investigate relationships among election data. The resulting information should be useful as redistricting tools are developed following the 2020 census. More specifically, the goal of this research is to expand on methods for analyzing gerrymandered elections in relation to uncontested races. County-level statewide election data (senate and gubernatorial elections from multiple years) are proposed as potential sources for data which could be used to approximate uncontested results. The validity of approaches are analyzed via comparisons to real data and simulated election data. Results indicate that county-level data can be used as an effective proxy for approximating uncontested results.

In this paper, Section 2 is dedicated to outlining fundamental ideas central to this study, including specific South Carolina information discussed in Section 2.1 and particular analysis metrics outlined in Sections 2.2 and 2.3. Recent contested federal House data is also analyzed in this section to provide an illustration of the use of the various metrics. Section 3 is dedicated to understanding the potential role of county-level statewide election data in approximating districted data. We analyze particular characteristics of county-level statewide election data and compare with districted results in Section 3.1; we evaluate effectiveness of county-level statewide election data as approximations to unknown data in Section 3.2. Section 4 then provides an application for these approximations in uncontested districted results and analysis of subsequent results.

2 Background and examples

In this section we outline geographical and demographic information pertinent to understanding election dynamics in South Carolina and outline methods used to analyze partisan fairness.

2.1 South Carolina

South Carolina is divided into seven federal districts corresponding to the seven allotted federal House of Representative seats (Figure 1) [21]. Each district is shaped according to both federal and state guidelines relating to compactness, equality in population size, maintenance of communities of interest, and incumbency protection [21]. South Carolina's seven-district map was developed in 2011 following the previous six-district plan [19]. South Carolina also has 46 counties, 124 state House districts, and 46 state Senate districts. A large percentage of state House and Senate districts frequently remain uncontested [5], potentially affecting interpretations of electoral fairness.

The Voting Rights Act requires for the purposeful drawing of a majority-minority district to provide appropriate representation for minority groups [1]. In South Carolina, District 6 has been created as a majority-minority African American district [1].

As a frequently Republican-leaning state [21], the majority of federal and state representative seats have been allocated to Republican representatives since 1994 [7]. However, as the demographics and population centers of a state naturally change, its population distribution can become appreciably different from previously recorded [19]. Ensuring that the population in each district is nearly equal and that each population can elect an official to accurately represent them warrants ongoing evaluation of previously-drawn districts, which typically occurs on a 10-year cycle following the census.

Figure 1: Image of the 2021 South Carolina federal House of Representative 7 district layout.

2.2 Methods for identifying symmetry and bias

A variety of metrics have been used to evaluate partisan bias within a districting plan. Methods include analyzing partisan symmetry, computing a mean-median (MM) difference, determining an efficiency gap (EG), and performing other statistical tests such as outlier analysis via ensemble tests [6]. These methods have been used in previous studies in order to evaluate the partisanship of maps in both Wisconsin and North Carolina, among other states [9] [8]. In this section, we introduce and describe evaluation methods including seats-votes curves, mean-median gaps, bias parameters, and declination metrics. This study functions under the fundamental assumption of a two-party system, with the Republican party featured as the reference party.

Partisan symmetry occurs when, in a two-party system, both parties have the same level of inherent advantage or disadvantage [21]. In a perfectly symmetric election, a certain percent of the vote corresponds to a certain percent of the seats, regardless of which party receives that percent of votes [21]. An asymmetric election, on the other hand, occurs when the number of obtained seats for one party is not equivalent to that won by the other party under the same vote percentage conditions [21]. In other words, under symmetric conditions, the equation

$$S(V) = 1 - S(1 - V), \tag{1}$$

where V represents the vote percentage and $S(V)$ represents the corresponding seat allocation percentage, holds [11]. Creating a seats-votes curve, which can be qualitatively assessed for symmetry, allows for further calculations to quantify the level of symmetry (or asymmetry) of a particular election [21].

In order to visually produce a metric to determine partisan symmetry for this study, a seats-votes curve was produced for multiple recent SC elections, obtained from CNN [4]. To accomplish this, the average vote, described as

$$V_{ave} = \frac{1}{L} \sum_{d=1}^{L} v_d \tag{2}$$

[21], was calculated for each election. Here, v_d refers to vote percentages towards the Republican party in district d, and L refers to the total number of districts. The corresponding seat proportion, $S(V)$ was then defined as

$$S(V) = \frac{1}{L} \sum_{d=1}^{L} \mathbb{1}(v_d > 0.5), \tag{3}$$

where $\mathbb{1}$ is the indicator function [21]. In this case, we compute an average over districts with a majority vote over the total number of districts within the election in order to determine the corresponding number of average seats to the average vote proportion. Then, the average vote for each individual district is modified by a "uniform swing" ranging from -1.00 to 1.00, in discrete steps of 0.01. Defining the true election result as having a swing of 0.00, one step of 0.01 results in each district vote percentage increasing by 0.01, and then resulting seats are recalculated. For each level of swing, V and $S(V)$ are calculated and plotted to obtain a seats-votes curve. The resulting curves provide a qualitative representation of symmetry for each election and the corresponding districting map. In the following seats-votes curves presented, we use V rather than V_{ave}, following convention. Additionally, the steepness of the inner curve, the number and extent of plateaus, and the vertical and horizontal asymmetrical discrepancy at the 50% mark can describe particular election characteristics with greater asymmetry generally indicating decreased fairness.

The seats-votes curve for the 2020 House results determined using a uniform swing assumption is shown in Figure 2. This curve highlights the existence of asymmetry since the reference party requires only 47% of the vote to obtain 50% of the seats. Additionally, the existence of small district numbers results in relatively long plateau lengths, creating a step-like appearance. Since a discrete number of seats means there will likely exist a discrepancy between the average vote percentage and the number of seats obtained, understanding the extent of this discrepancy is key to understanding the inherent fairness of an election.

Certain attributes of election properties can be quantified and used to directly compare the results of one election to other similar elections. A bias parameter was calculated for each election to provide a quantitative estimate for symmetry. Under the fundamental assumption that partisan symmetry occurs when the system $S(V) = 1 - S(1 - V)$ is satisfied for all V, a bias parameter, $\beta(V)$ can be defined as

$$\beta(V) = \frac{S(V) - [1 - S(1 - V)]}{2} \tag{4}$$

[21]. This metric quantitatively shows a deviation from perfect partisan symmetry and therefore provides a numerical metric for determining the extent of partisan asymmetry at a certain V_{ave}. Notably, implementing a uniform swing assumption can produce calculated V values corresponding to each finite change of swing, potentially resulting in two adjacent calculated V which are more than 0.01 vote percentage apart. In this case, $S(V)$ for a specific V which could be directly calculated at a finer degree of swing can be approximated under a continuous assumption using averaging.
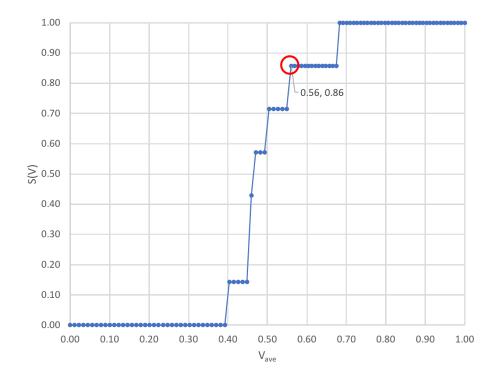
Figure 2: Seats-votes curve of the 2020 South Carolina House of Representatives election assuming uniform swing. The curve can be used to visualize the relationship between votes obtained and number of seats obtained. Notably, the small quantity and large lengths of plateaus can be attributed to the low number of districts. Similarly, the jumps between plateaus occur because of a winner-take-all property implemented within each district such that only whole-numbered seats can be obtained by a particular party. The red circle indicates the true election result.

Mean-median difference is also frequently used to quantify fairness as a measure of asymmetrical skew [2]. Mean-median gap is computed as

$$MM = V_{ave} - M, \tag{5}$$

where V_{ave} indicates the average district vote and M the median across $\{v_1, v_2, ...v_L\}$ [11]. Though a mean-median gap of zero suggests a symmetric election, this measurement can contain a notable level of inaccuracy [2] and thus should be used in tandem with other tests, rather than as a singular tool [11].

In addition to assessing the potential extent of partisan gerrymandering as a deviation from symmetry, a declination metric, as described by Warrington [22], can be calculated to determine the magnitude of discrepancy for a particular election at the 50% mark. To compute the declination, district votes are plotted in increasing order, a horizontal line is drawn at $V = 0.5$, and then the angles from the point indicating a switch in partisanship to the centers of mass of Republican majority and minority parties are compared. The declination metric, δ, then compares the two angles with the underlying assumption that a perfectly fair election would warrant both angles being equal such that $\delta = 0$ [22]. Development of the resulting metric stems from the fundamental idea that in a fair election, substantial change in bias should not occur consistently around the 50/50 mark [22]. Thus, it is reasonable to focus analysis of a potentially gerrymandered map at this point.

To compute declination, all district percentage votes for an election are plotted in numerical order, as in Figure 3. Then, a horizontal line is drawn at $V = 0.5$. Two reference points on either side of the $V = 0.5$ line are noted by stars in Figure 3. The leftmost star is placed horizontally in the center of mass of points which the reference party did not win and vertically at the average of vote percentage for those districts. The rightmost point is plotted in a similar manner for the districts that the reference party did win. One final point is placed on $(k/N, 0.5)$, where k represents the number of districts lost and N the total number of districts. Declination is then described as the deviation of this point from the line connecting the two starred points, whereby deviation of this point below the line indicates an advantage for the Republican party [22]. These angles can be precisely calculated as,

$$\theta_A = \arctan \frac{2\bar{z} - 1}{k'/N} \tag{6a}$$

$$\theta_B = \arctan \frac{1 - 2\bar{y}}{k/N}, \tag{6b}$$

where θ_A corresponds to the angle of deviation in favor of party A, θ_B to the angle of deviation in favor of party B, k to the number of districts won by party A out of L districts, and k' to the number of districts won by party B out of L districts. The average district vote over the districts where party A won and the average district vote over the districts where party B won are indicated by $\bar{z}$ and $\bar{y}$, respectively. The difference between these two values produces a number that describes the extent of partisan bias towards a reference party, which is divided by $\frac{\pi}{2}$, producing the standardized metric, δ, ranging from -1 to 1. In context of this study, a value close to 1 indicates a strong Republican bias, and a value close to -1 indicates a strong Democrat bias.

Recent South Carolina House of Representative elections were analyzed using the above methods, with results summarized in Table 1. Focusing on the most recent 2020 election, a fairly large declination value (0.3429) suggests the existence of substantial Republican bias, further supported by the notable positive value of the calculated MM gap (0.04). However, the average $\beta(V)$ is small
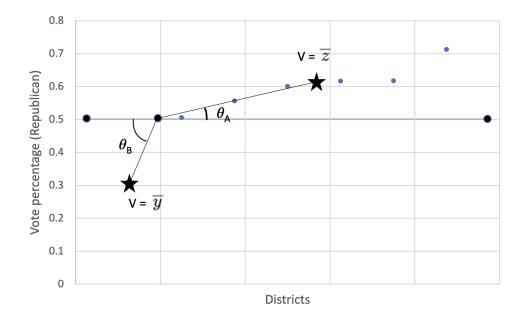
Figure 3: Visualization of the declination metric calculation for the recent 2020 House election results. District vote percentages are plotted in order of Republican vote percentages. Points are then notated at the center of mass of Republican majority and Republican minority districts. Then, angles θ_A and θ_B are calculated as indicated by Equation 6, and declination δ is calculated from the difference.

(-0.01) compared to the $\beta(V)$ at 50% (0.21). This suggests bias exists condensed around the 50% mark, rather than being equally distributed throughout the whole curve.

It is worth noting that all forms of testing described above can show a level of partisan bias present which may not be a result of gerrymandering, but rather a simple consequence of geography and the distribution of voters across the state [22]. In this case, it may be useful to analyze competitiveness or responsiveness of the election along with looking at population demographic distribution to describe a more holistic picture [21]. Though the existence of bias or asymmetry does not necessarily imply that the election results were unfair or a result of ill-intentioned gerrymandering, continuing to analyze the election to obtain more information is, regardless, a valuable pursuit.

Election	Average (V)	Variance	$\rho(V)$	MM Gap	Declination (δ)	Average $\beta(V)$	$\beta(V)$ at 50%
2020	0.56	0.01402	2.15	0.04	0.3429	0.013	0.21
2018	0.54	0.01358	3.6	0.03	0.1986	0.057	0.21
2016	0.58	0.01781	2.15	0.02	0.3782	0.045	0.36

Table 1: Quantitative results for recent South Carolina House of Representative elections. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments. Positive MM gap, declination, and $\beta(V)$ values indicate Republican advantage. Larger variance magnitudes indicate greater competitiveness, and greater $\rho(V)$ values indicate greater responsiveness.

2.3 Other measurements for analyzing election data

In addition to quantifying partisan bias, other metrics can provide additional information that is useful in analyzing election data. For instance, the variance parameter,

$$\text{VAR} = \frac{1}{L} \sum_{d=1}^{L} (v_d)^2 - (V_{ave})^2, \tag{7}$$

can quantify the level of competitiveness exhibited within a particular election [21]. This competitiveness parameter is helpful when determining if election results reflect high competition in many districts favoring one party, indicating cracking, and low competition in few districts favoring the other party, indicating packing.

A responsiveness parameter, ρ, has previously been defined by Katz et al. as

$$\rho(V) = \frac{\partial S(V)}{\partial V}, \tag{8}$$

[11]. An election with a higher responsiveness value will be more reactive to a change in election result V, reflected in the resulting value $S(V)$. However, since this continuous definition cannot be directly applied to discrete seat-vote curves, we have approximated ρ using a the difference quotient approximation

$$\rho(V) = \frac{S(V^L) - S(V^R)}{V^L - V^R}, \tag{9}$$

where V^L and V^R indicate average votes to the left and right of the true election result, respectively [11]. For the intents of this study, we chose V^L and V^R to be equal to the true $V_{ave} \pm 10\%$ in order to capture a range of reasonable election results, as discussed by Katz et. al [11].

We have computed variance and responsiveness for the 2020 House results. Responsiveness was calculated where $V^L = 0.46$ and $V^R = 0.66$ since $V_{ave} = 0.56$ (Table 1). When modeling idealized seats-votes curves, ρ is frequently set to a value of 3, corresponding with the cube law [21]. In comparison, the responsiveness of this election ($\rho = 2.15$) appears low. However, it is worth noting that low responsiveness is likely partially due to the fact that, since the House results of 2020 led to six Republican seats, an increase in Republican votes likely would not easily cause the last seat to flip. This concept is further supported by the low variance (VAR = 0.01) calculated for this election, reflecting a low level of competitiveness. The combination of these two parameters suggest that the election was not particularly competitive or responsive, meaning a small change in V likely would not correlate to an equal change in $S(V)$ in either direction of swing. The collection of data for multiple elections was relevant in determining if the information outlined for the 2020 House election was particularly asymmetric, biased, or noncompetitive compared to other similar elections.

All comparison metrics were computed for the 2018 and 2016 elections in addition to the 2020 election in order to illustrate meaningful analysis of multiple election results (Table 1). Data for 2018 and 2016 elections were obtained from [23] and [16]. The seats-votes curves for both elections are shown in Figure 4. Vote percentages indicated by V in Figure 4 were calculated as V_{ave} from Equation 2, but relabeled to follow convention. Additionally, the 2018 election resulted in two Democrat seats won versus only one in 2020 and 2016, almost certainly contributing to the differences among certain metrics. For instance, the differences in location of true election result on the seats-votes curves between the 2018 federal House and other recent House elections is likely a key factor affecting responsiveness discrepancies between the elections (Figure 4). The average $\beta(V)$ values that we computed were similar across elections. However, we lack sufficient
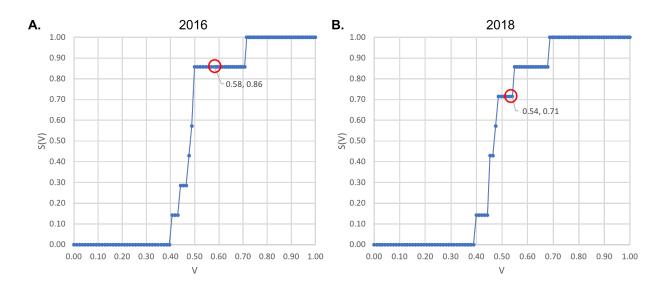
Figure 4: Seats-votes curves of the 2016 (A.) and 2018 (B.) South Carolina federal House of Representative elections. The curves can be used to visualize the relationship between votes obtained and number of seats obtained. Red circles indicate true election results. The 2018 election (A.) is notable due to a Republican win of five seats, rather than the typical six-seat win.

data to determine if small differences in these values are significant, an issue that we will address via simulations in Section 3.2.1.

3 County-level statewide election data as an approximation tool for incomplete district election data

Election data analysis can be hindered by the existence of incomplete or unknown data. Approximation strategies are often recommended to mitigate this problem [2]. The following section is dedicated to the assessment of using known county-based statewide election results as a partisanship proxy and thus a potential approximation tool for unknown districted data. We find that dynamics of county-level data are fundamentally different from districted data. However, despite these characteristic differences, implementation of county-level data can be used as a potential method in approximating districted results in uncontested circumstances.

3.1 Analysis of county-level election data in South Carolina

Since certain statewide elections do not rely on districts, we plotted a counties-votes curve to visually compare with the seats-votes curves derived for other elections. Gubernatorial data from the 2018 elections obtained from The New York Times [18] were used to characterize and differentiate county-based data from districted House data. Visually, the 2018 gubernatorial counties-votes curve appears much smoother and much more symmetric than the House curves, warranting further examination of these differences (Figures 2, 4, 5). The relative smoothness of the resulting curve can be attributed to the larger number of counties when compared to federal districts. Overlaying the counties-votes curve on a federal seats-votes curve allows for an intuitive examination of how federal election results diverge from the general partisanship of the state. The overlay of the 2020 federal House of Representative election and the 2020 Senate election is shown in Figure 6. Although

	Election	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
Governor	2018	0.54	-0.0051	3.15	0	0.0938	0.0110	0.09
	2014	0.56	0.0263	2.65	-0.03	0.0881	-0.0070	0
Senate	2020	0.55	0.0020	3.25	0.01	0.1633	-0.022	0.02
	2016	0.61	-0.0412	2.5	-0.03	0.2665	-0.085	-0.02
	2014	0.55	-0.0252	2.95	-0.02	0.0227	-0.067	-0.07

Table 2: Quantitative results for recent South Carolina county-based elections. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments. Positive MM gap, declination, and $\beta(V)$ values indicate Republican advantage. Larger variance magnitudes indicate greater competitiveness, and greater $\rho(V)$ values indicate greater responsiveness. Both elections reflect relatively high responsiveness and comparatively low $\beta(V)$ and declination values when compared to federal House elections.

there exist locations where the county-based curve lies above the 2020 House curve, there are more dramatic locations where the 2020 House curve lies above the Senate curve. This supports the notion of the Republican bias noted previously, which is further supported by the comparatively large declination magnitude (Table 1).

Quantitative metrics were calculated using results from 2018 [18] and 2014 [17] gubernatorial elections and 2020 [3], 2016 [14], and 2014 [17] federal Senate races and summarized in Table 2. The 2018 average vote percentage was similar to the 2018 federal House election, suggesting the existence of a relationship between partisanship exhibited during House and gubernatorial races (Tables 1 and 2). The calculated variance was of a similar magnitude to that of the federal House elections analyzed previously. Despite these similarities, however, the 2018 gubernatorial election showed a much greater responsiveness and much lower MM gap, declination, and bias parameter results when compared to the federal House elections. Since these parameters describe a district-seats relationship, they cannot be used for isolated county-based election analysis; rather, the calculated parameters can be used to compare certain properties of county-based election data compared to districted data with the understanding that the values of the parameters are absent of inherent meaning for county-based data. These results suggest that gubernatorial and Senate data likely mirror House data despite certain differences. Additionally, the step-like progression of the House curve when compared to the smoothed county-based curve reflects the difference in responsiveness between the two sets of data, as the Senate data was shown to have a responsiveness parameter closer to $\rho = 3$ (Table 2).

Consistency among these results and bias metrics discussed in Sections 2.2 and 2.3 encourage further use for county data in understanding inherent South Carolina partisanship and in election results.

3.2 Evaluation of approximations

Two methods were used to evaluate the effectiveness of county data as a tool for district vote percentage approximation. Simulated results were created to test the impact of a county-level approximation under consistent and replicable conditions. True historical data was then used to examine the effectiveness of approximation on non-randomized data. County-level approximations produced results similar to true data in both simulated and historical contexts, providing validation for the use of this and other similar approximation methods.
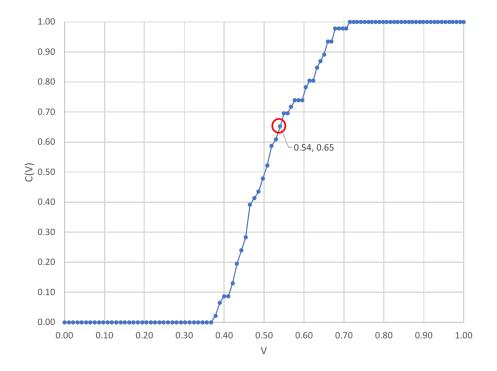
Figure 5: Counties-votes curve of the 2018 South Carolina gubernatorial election assuming uniform swing. The red circle indicates the true election result. The symmetry exhibited here suggests that South Carolina county data is relatively unbiased. Smoothness of the resulting curve is due to the comparatively large number of counties (46) compared to the number of federal districts (7).
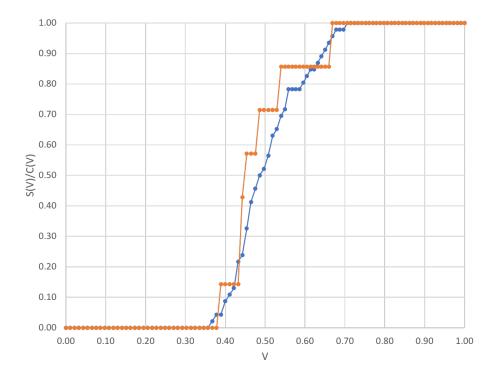
Figure 6: Overlay of the 2020 federal House election and 2020 Senate election data. Using the underlying assumption that the county-based data generally reflects the inherent partisanship of South Carolina, diverges from this curve can reflect a degree of bias. The blue curve represents 2020 Senate data, and the orange curve reflects 2020 federal House data.

Election	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
Original	0.56	0.01688	2.56	0.02	0.3232	0.02010	0.19
Counties (Sen)	0.57	0.01773	1.94*	0.03	0.3928	0.01701	0.15
Counties (Gov)	0.55	0.01701	2.76	0.02	0.2913	0.01613	0.12

Table 3: Average simulated results for multiple election metrics. Thirty simulations were run and each original election was compared to one district replaced with an approximated vote percentage. Resulting metrics were compared to the original data to determine if county data can be an effective proxy to approximate districted results. Results notated with * are significantly different from original data. Responsiveness parameters were computed for $\pm10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments.

3.2.1 Simulated results

To evaluate the effectiveness of using county results as a proxy for otherwise unknown districted results, we ran simulations to create a sample of elections, and computed metrics from Sections 2.2 and 2.3 to compare with actual election results. We simulated election results on the seven-district House of Representatives map 30 times where each district vote percentage was randomly generated from a reasonable range determined by the South Carolina House of Representative results from 2012 to 2020. We created these ranges by noting the lowest and highest vote percentage for each district and rounding to the nearest percent. For each set of simulated results, metrics were calculated for the original (true) data and then for data including an approximated vote percentage using Senate and gubernatorial county data.

Results of simulations are summarized in Table 3. A one-way ANOVA followed by a Tukey HSD test were performed to determine statistically significant differences between groups. Both gubernatorial and Senate data provided approximations which resulted in statistically similar calculated metrics with an exception being responsiveness ($p > 0.05$). In this case, Senate county-level results reflected a lower responsiveness than original data ($p = 0.0084$).

Similarities between approximated results and original simulated data suggest that using county data as a source of approximation is a valid approach and can be used to address situations where there is a lack of present districted data. To mitigate computational strain, differences in county populations were not taken into account when implementing this method.

3.2.2 Real election data

Randomized simulations described above and only fully contested federal election results (2016, 2018, 2020) were used to further evaluate the effect of county-based approximation methods on general accuracy. For all three elections, we took original contested election data and then changed one district's vote percentage to be unknown or incomplete. We then used county-based approximation methods and determined the resulting accuracies of calculated metrics. Results are summarized in Table 4. For all three years, county data appeared to produce very similar average vote percentages, variances, responsiveness, and mean-median gap values when compared to true data. Additionally, declination values appeared similar between original and county-based data but not to the same degree as other metrics. Bias parameters, however, appeared fairly consistent between county and true data for 2020 and 2016, but were very different for the 2018 election. Since there were only three contested elections, it is uncertain whether the 2018 bias parameter results were notable outliers. However, the results do appear to follow the trends established by the

Election	Method	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
2020	Original	0.56	0.01402	2.15	0.04	0.3429	0.013	0.21
	Counties (Sen)	0.55	0.01161	2.15	0.05	0.3082	0.043	0.21
	Counties (Gov)	0.55	0.01159	2.15	0.05	0.3082	0.043	0.21
2018	Original	0.54	0.01358	3.60	0.03	0.1986	0.057	0.21
	Counties (Sen)	0.53	0.01353	3.60	0.04	0.1348	0.029	0.07
	Counties (Gov)	0.53	0.01355	3.60	0.04	0.0998	0.029	0.07
2016	Original	0.58	0.01781	2.15	0.02	0.3782	0.045	0.36
	Counties (Sen)	0.58	0.01776	2.15	0.02	0.3735	0.029	0.21
	Counties (Gov)	0.57	0.01821	2.15	0.02	0.3516	0.014	0.21

Table 4: Summary results for multiple real-data approximation comparisons. At each election year, metrics were calculated and then county-level approximation methods were used to provide an estimated vote percentage for a particular district. Metrics were then calculated for each approximated dataset, and comparisons were made to the original data. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments.

previous simulation results, further suggesting that county data can be used as an effective proxy to approximate districted results.

4 Addressing uncontested districts using county data and state-wide elections

Using metrics to compare the properties of particular elections can be helpful in evaluating the fairness of a current districting map. However, these metrics are only valid to compare elections where each district is contested, since the relationship between two parties cannot be determined when there is not a candidate running from each reference party. Although the issue of uncontested results is well-known and extremely prevalent throughout states such as South Carolina [2], there is not a standardized method for altering uncontested results to allow for analysis. Multiple methods have been proposed, such as removal of the districts completely or simple replacement with a 0/100 or 25/75 split [2]. Stiff split methods such as these are conducted by replacing the vote percentage of the winning party by a constant value such as 100% or 75%, regardless of previous voting history of the district. More complex methods have also been suggested whereby the result of the uncontested district is approximated using the results of neighboring districts or adjacent years [13].

Here we present the seats-votes curves and bias metrics obtained using multiple approximation approaches. Uncontested districts could be replaced with a 0/100 split and removed completely. These methods are commonly believed to be generally inaccurate yet mechanically simple, and so were analyzed to determine the validity of this belief [2]. To maintain the same degree of simplicity but provide a more accurate representation, implementing a 25/75 split has been suggested [2], and so this method was also analyzed. Additionally, it is widely believed that recent elections can be used as a proxy to determine an approximated value closer to what the true contested result would be [2], and so a proportional method was conducted using the same district result from an adjacent year. The proportional vote percentage to the previous year was calculated in regards to the total average vote, such that

	Method	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
2014	0/100	0.66	0.04714	0.75	-0.04	0.4742	-0.020	-0.07
	Removed	0.61	-0.02432	0.80	0.01	0.3790	-0.0180	0.00
	25/75	0.63	0.03066	0.75	0.00	0.4382	0.0150	0.07
	Proportional	0.61	0.02859	1.10	0.01	0.4171	0.0010	0.07
	Counties	0.60	0.02834	1.45	0.00	0.3965	0.0010	0.07
2012	0/100	0.58	0.07553	1.45	0.05	0.2031	0.0430	0.07
	Removed	0.61	0.0023	2.00	0.01	N/A	0.0000	0.10
	25/75	0.58	0.02196	2.15	0.05	0.3501	0.0150	0.07
	Proportional	0.55	0.01908	2.15	0.01	0.2470	0.0010	0.07
	Counties	0.56	0.0080	2.85	-0.01	0.4175	0.0130	-0.07

Table 5: Comparison of the five methods of approximating uncontested results on analysis parameters. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments. Positive MM gap, declination, and $\beta(V)$ values indicate Republican advantage. Larger variance magnitudes indicate greater competitiveness, and greater $\rho(V)$ values indicate greater responsiveness.

$$v_d^y = \frac{v_d^{y+1} V^y}{V^{y+1}}, \tag{10}$$

where v_d is the average vote for a particular district and V is the average vote for the entire election. In this equation, y reflects the specific election year and the reference proportion was taken from the adjacent year. We previously confirmed in Section 3.2 the validity of using state-wide county data as an approximation tool, and so the results of the above methods were compared to county-based results. Following the previous confirmation of the validity of county-based approximation strategies, we propose a method of approximating uncontested districted results by computing an average of statewide county-level election results contained within the uncontested district. We find this method to be effective in both state and federal election contexts.

4.1 Federal House implementation

Due to the relatively small number of federal House seats, one of the simplest cases of this uncontested problem surfaces when attempting to analyze the federal House of Representative elections prior to 2016. Both the 2014 and 2012 results show at least one uncontested district which must be accounted for before comparing the results from uncontested elections to contested results [17][15]. The five methods outlined above were used to handle the presence of uncontested districts for both the 2014 and 2012 House elections, and results were compared to county-based approximation strategies. Both Senate and gubernatorial county data were confirmed to produce similar results, and so only 2014 gubernatorial county data were used for 2014 House election analysis.

We summarize comparisons of metrics between these different approximation methods for 2014 and 2012 federal House results in Table 5. To determine method effectiveness, we compare the resulting metrics to county-based methods, which we have already confirmed are valid. Simulations outlined in Section 3.2.1 were used to provide statistical support for the following results.

We created and visually evaluated seats-votes curves to see the effect of different methods of approximating uncontested results had on election asymmetry, responsiveness close to the 50% mark, and number and characteristics of any plateaus compared to contested results. Seat-votes

curves for the different approximation methods were compared to the county-based method (Figure 7). The proportional method (Figure 7C) appeared to be most similar to the county-based method, shown in Figure 7E, due to the number and location of plateaus, the consistent $(0.50, 0.57)$ result, and the relative location of the true election result. In contrast, none of these characteristics are similar when comparing the 0/100 split (Figure 7A) to the county-based curve. Although the characteristics for the removal curve (Figure 7B) are similar to those of the county-level curve, the removal of one district produced a curve that contains fewer plateaus as a result. Lastly, although the 25/75 method (Figure 7D) mimicked the characteristics of the county-based curve better than the 0/100 split, the proportional method was still preferred due to the responsiveness close to the 50% mark.

In addition to a qualitative analysis, a quantitative analysis was conducted, similar to that for the previous federal elections, producing results from real data summarized in Table 5 and results from simulated data summarized in Table 6. Each method for approximating the uncontested district in the 2014 House election shown in Table 5 was compared to the county-based approximation which we previously confirmed to be a valid strategy. Furthermore, our simulations in Table 6 were used to provide statistical evidence to support the results obtained from real data in Table 5. For all 2014 and 2012 parameters, the 0/100 method produced results substantially different from county-based approximations (Table 5), corresponding to results found significantly different from original data in simulation tests (Table 6, $p < 0.0001$). In addition, the proportional and county methods produced very similar results for each 2014 and 2012 House parameter (Table 5), supported by simulation results (Table 6, $p > 0.05$).

Another way of looking at this data is presented in Table 7. Here, we used methods outlined in Section 3.2.2 by taking fully contested races, choosing one district to be uncontested, and using different strategies to approximate the missing data. In contrast to simulation strategies, this approach benefits from the use of true historical data. Results show trends similar to simulation strategies, where the 0/100 split and removal methods produce metrics substantially different from original data.

From these results, we recommend using dynamic county-based or proportional methods when possible. If these methods cannot be used, the 25/75 split method should be used. The 0/100 split is not recommended since the 25/75 split is just as mechanically simple but was found to produce consistently better results. The removal method should not be used in order to approximate contested results due to the fact that it promotes the complete removal of pertinent information. Furthermore, if the removed districts are the only districts which were won by one of the two major parties within an election, the removal method would prevent certain computations, such as the declination, as shown when calculating the 2012 parameters (Table 5).

4.2   Uncontested elections at the state level

In addition to analyzing federal-level election data, state-level data were examined to determine if there existed notable differences in characteristics between federal and state elections. With much larger numbers of House districts and larger percentages of uncontested seats, it follows that state-level data behaves differently than federal election data. We confirm the existence of key differences in state data versus federal data and implement county-based data replacement strategies to mitigate the effects of extremely high uncontested percentages within these elections.
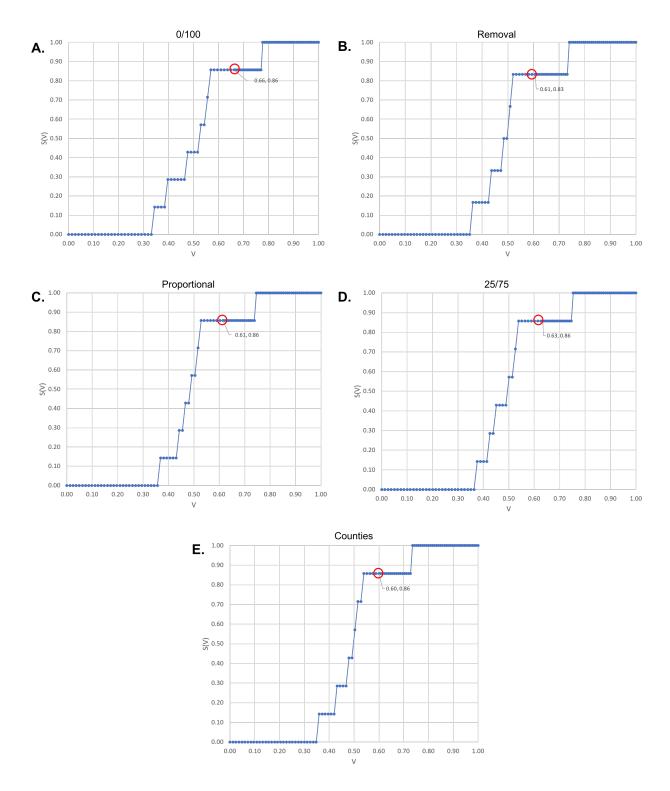
Figure 7: Comparison of all seats-votes curves used in addressing the uncontested district found in the 2014 South Carolina House election. Figure A shows the seats-votes curve resulting from implementing the 0/100 split method, Figure B from the removal method, Figure C from the proportional method, Figure D from the 25/75 split method, and Figure E from the county-based method. Red circles reflect the election result of each method.

4.2.1 State-level properties

The South Carolina state House of Representative was analyzed for 2020 using data obtained from USA Today [20]. In addition to the large number of districts (124), the 2020 state House election contained a much larger percentage of uncontested races ($> 50\%$) than the federal House elections analyzed. The presence of this larger percentage changes the characteristics of each approximation. For instance, a 0/100 or 25/75 split replacement method would result in a large percentage of districts showing the same results, which is extremely unlikely and can have substantial effects on election analysis results. Strengths of flexible methods, such as county-based and proportional methods, are especially evident in cases such as these as different approximations are able to be produced for multiple uncontested districts. However, large percentages of uncontested districts are found within most all recent elections, which we show prevents the effective implementation of certain key methods of proportional approximation.

4.2.2 Map analysis

Since over 56% of districts were uncontested in the 2020 South Carolina state House of Representative election, a map was created where the locations of these uncontested districts were provided in order to help in the ease of visualization (Figure 8). It was found that many uncontested districts were adjacent with few isolated uncontested districts, often presenting in quite large clusters (Figure 8). This clustering makes evident how certain techniques, such as using geographically close results to approximate uncontested results, would be unhelpful for approximating results for state elections. Additionally, the uncontested Democratic districts appear to be present where there exists larger populations of African Americans, and uncontested Republican districts appear to be present where there exists mostly White populations [10], revealing the connection between political geography and racial geography, which has implications for maintaining communities of interest when redistricting.

4.2.3 State House results

The previous analysis determined whether the recommended methods produced similar results in slightly different conditions, for one uncontested district. The same methods used to approximate uncontested results for the federal House elections were used for the state House elections. However, the proportional method was unable to be used due to the equally high level of uncontested districts in the previous 2018 election. County-based results were once again used as a reference to determine the recommendation for handling uncontested districts.

Seats-votes curves were compared between methods and are shown in Figure 9. Since a much larger percentage of districts were uncontested for state elections in comparison to federal elections, the results of implementing a stiff split approximation method include visually notable jumps on the ends (Figures 9A and 9C). In this case, a removal method can produce a smoother curve similar in appearance to the county-based generated curve, likely due to the similar number of both Democratic and Republican uncontested districts (Figure 9B and 9D).

Quantitative results for the 2020 state House of Representative election are summarized in Table 8. As before, results were compared to county-based results as statewide data was previously confirmed to produce good approximation methods. With this in mind, the 0/100 method was likely not effective at capturing an effective approximated average. The variance, responsiveness, MM gap, and declination results should mimic those of county-based results, and so the 25/75 split method and 0/100 split method were likely ineffective. Since none of these other methods match
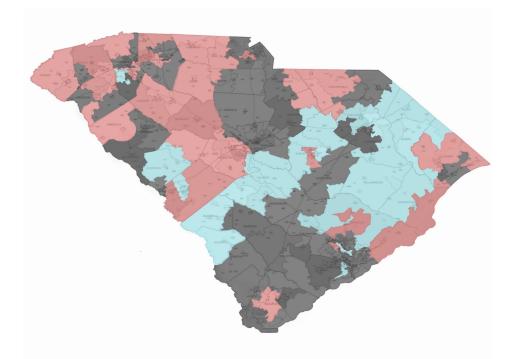
Figure 8: Visualization of the geographic location of uncontested districts. The South Carolina state house of representative uncontested locations are indicated with pink (Republican win) and blue (Democratic win). Darkened areas reflect contested districts. Base map obtained from [12].

the metrics resulting from county-level approximations, we would not expect them to match the true data.

5 Final discussion

In this paper, recent South Carolina election results were analyzed using seats-votes curves and calculated metrics to facilitate further understanding into particular election characteristics. Then, using data obtained for contested elections, quantitative analysis factors were used to assess the effectiveness of county-level statewide election data as a district vote approximation tool. After having validated county-based statewide election data as an approximation tool, other approximation methods were compared to understand how they are impacted by uncontested districting issues.

It was found that although there exist a variety of methods to approximate incomplete districted data, these methods can often simply be classified as dynamic or constant. In the context of this study, proportional and county-based methods were classified as dynamic and 0/100 split, 25/75 split, and removal methods were classified as constant methods. In this paper, we have shown that dynamic methods better match election data under various election conditions. This occurs due to the ability to include characteristics of election data, adjust approximations over time, and avoid issues associated with repeated identical results. However, dynamic methods are more computationally expensive than constant methods and are based on the assumptions that there exists continuity of voter behavior over time, across candidates, and between federal and state elections.
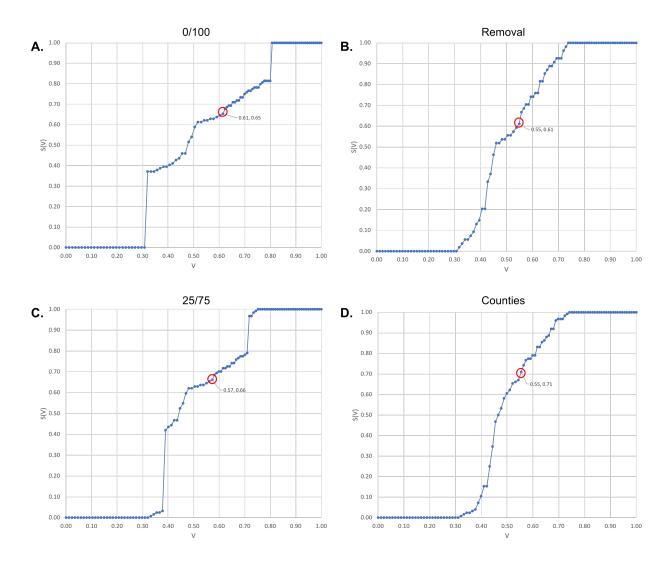
Figure 9: Comparison of seats-votes curves for all uncontested approximation methods. Due to the large percentage of uncontested districts, both the 0/100 split method (Figure A) and the 25/75 split method (Figure C) produced curves which are very asymmetric and contain very large jumps where the large number of approximated districts flip at the same time. Additionally, the removal method (Figure B) produces a smoothed curve but with much fewer datapoints due to the removal of over half of the districts. County-level replacement strategies (Figure D) result in a more visually ideal curve. Red circles indicate true election results.

Election	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
Original	0.56	0.01688	2.56	0.023	0.3232	0.0201	0.186
0/100	0.62**	0.04038**	1.42**	-0.024**	0.4500**	-0.0423**	-0.098**
Removed	0.56	-0.02825**	2.30	0.026	0.2846	0.0221	0.185
25/75	0.59**	0.02112*	1.68**	0.016	0.3850*	0.0132	0.089*
Proportional	0.57	0.01763	1.99*	0.025	0.3491	0.0315	0.179
Counties (Sen)	0.57	0.01773	1.94*	0.025	0.3928*	0.0170	0.146
Counties (Gov)	0.55	0.01701	2.76	0.022	0.2913	0.0161	0.121

Table 6: Average simulated results for multiple election metrics. Thirty simulations were run and each original election was compared to one district replaced with an approximated vote percentage. Resulting metrics were compared to the original data to determine if certain approximation methods were consistently superior to others. Notably, the 0/100 split method was consistently different from the original data, confirming the method to be unhelpful. Using county data and proportional methods resulted in values consistently similar to original data, supporting the validity of these approaches. Results notated with ** are statistically different from original data ($p << 0.0001$). Results notated with * are statistically different from original data ($p < 0.01$).

Election	Method	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
2020	Original	0.56	0.01402	2.15	0.04	0.3429	0.013	0.21
	0/100	0.60	0.03665	1.45	0.00	0.4489	-0.015	0.07
	Removed	0.53	-0.03070	4.15	0.04	-0.0449	0.033	0.17
	25/75	0.57	0.01580	1.45	0.04	0.3599	-0.015	0.07
	Proportional	0.57	0.01596	1.45	0.04	0.3626	-0.015	0.07
	Counties (Sen)	0.55	0.01161	2.15	0.05	0.3082	0.043	0.21
	Counties (Gov)	0.55	0.01159	2.15	0.05	0.3082	0.043	0.21
2018	Original	0.54	0.01358	3.60	0.03	0.1986	0.057	0.21
	0/100	0.60	0.03984	2.15	-0.01	0.3458	-0.043	-0.07
	Removed	0.54	-0.02758	3.30	0.05	0.1685	0.050	-0.17
	25/75	0.55	0.01909	2.15	0.03	0.2275	0.014	0.21
	Proportional	0.55	0.01396	3.60	0.05	0.1677	0.013	0.21
	Counties (Sen)	0.53	0.01353	3.60	0.04	0.1348	0.029	0.07
	Counties (Gov)	0.53	0.01355	3.60	0.04	0.0998	0.029	0.07
2016	Original	0.58	0.01781	2.15	0.02	0.3782	0.045	0.36
	0/100	0.64	0.03959	1.45	-0.03	0.4750	-0.013	-0.07
	Removed	0.58	-0.02991	1.85	0.02	0.3401	0.048	0.33
	25/75	0.60	0.02137	1.45	0.01	0.4261	-0.013	0.07
	Proportional	0.59	0.01871	1.45	0.02	0.3931	0.001	0.07
	Counties (Sen)	0.58	0.01776	2.15	0.02	0.3735	0.029	0.21
	Counties (Gov)	0.57	0.01821	2.15	0.02	0.3516	0.014	0.21

Table 7: Summary results for multiple real-data approximation comparisons. At each election year, metrics were calculated and then different approximation methods were used to approximate one particular district. Metrics were then calculated for each approximated dataset, and comparisons were made to the original data. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments.

Adjustment Method	Average (V)	Variance	$\rho(V)$	MM Gap	Declination	Average $\beta(V)$	$\beta(V)$ at 50%
0/100	0.61	0.1394	0.8	0.03	0.2	-0.0010	0.04
Removed	0.55	0.02833	1.95	0.06	0.2	-0.0060	0.06
25/75	0.57	0.04299	0.8	0.08	0.23	-0.0110	0.13
Counties	0.55	0.01952	2.05	0.04	0.24	0.0140	0.1

Table 8: Summary of the effects of different uncontested approximation on partisan analysis parameters for the 2020 South Carolina state House of Representative election. Responsiveness parameters were computed for $\pm 10\%$ of average vote and average $\beta(V)$ was computed for $V = 0.30$ to $V = 0.50$ at 0.05 increments. Positive MM gap, declination, and $\beta(V)$ values indicate Republican advantage. Larger variance magnitudes indicate greater competitiveness, and greater $\rho(V)$ values indicate greater responsiveness.

In cases where there exist few uncontested races, dynamic methods are generally superior to constant methods, but the potential consequences of using a constant method in this case are more modest. However, when there exists a high percentage of uncontested districts, the drawbacks of both types of methods are exacerbated, often preventing effective use of particular methods. For instance, constant methods can completely change the nature of election results due to a high level of replacement with inaccurate approximation methods. Additionally, the proportional method cannot be calculated when adjacent districts and election years contain a similarly high level of uncontested data.

Through this exploration, it was found that county-based methods were effective proxies for contested results, especially the statewide county-level method. In fact, this particular method reflected the advantages of other dynamic methods while still being implementable in cases of high uncontested percentages. Despite the fact that the methods implemented in this study did not adjust for differences in county population sizes, they still proved effective. Further investigation into adjusting for population sizes is worthwhile to develop a more rigorous uncontested approximation method. Further research into testing the use of different geographic regions such as precincts or census tracts would be valuable in further developing a series of methods to reduce the impact of uncontested results.

Furthermore, it is worth noting that there did not appear to exist an extreme difference in effectiveness between Senate county and gubernatorial county data overall, but gubernatorial data did appear more effective in certain contexts. This relationship could be analyzed further to create a more refined county-based method. Expanding on this particular method variation could produce a method which can be comfortably preferred over other approximation methods in most cases where limited districted data exists.

References

[1] BALLOTPEDIA, *Majority-minority districts.* https://ballotpedia.org/Majority-minority_districts.

[2] B. BURDEN AND C. SMIDT, *Evaluating legislative districts using measures of partisan bias and simulations*, SAGE Open, 10 (2020), p. 2158244020981054.

[3] CNN, *South Carolina 2020 Senate election results.* https://www.cnn.com/election/2020/results/state/south-carolina/senate, Mar 2021.

[4] CNN Politics, *South Carolina 2020 house election results.* `https://www.cnn.com/election/2020/results/state/south-carolina/house/district-1`, Mar 2021.

[5] A. Coyne, *Party primaries, uncontested races dominate general assembly.* `https://www.greenvilleonline.com/story/news/2016/03/30/party-primaries-uncontested-races-dominate-general-assembly/82444644/`, Mar 2016.

[6] M. Duchin, *Gerrymandering metrics: How to measure? What's the baseline?*, arXiv: Physics and Society, (2018).

[7] Electoral Ventures LLC. `https://www.270towin.com/states/South_Carolina`, 2022.

[8] G. Herschlag, H. S. Kang, J. Luo, C. V. Graves, S. Bangia, R. Ravier, and J. C. Mattingly, *Quantifying gerrymandering in North Carolina*, Statistics and Public Policy, 7 (2020), pp. 30–38.

[9] G. Herschlag, R. Ravier, and J. C. Mattingly, *Evaluating partisan gerrymandering in Wisconsin*, 2017.

[10] Index Mundi, *South Carolina black population percentage by county.* `https://www.indexmundi.com/facts/united-states/quick-facts/south-carolina/black-population-percentage#map`.

[11] J. N. Katz, G. King, and E. Rosenblatt, *Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies*, American Political Science Review, 114 (2020), pp. 164–178.

[12] Sciway, *South Carolina – government and election maps.* `https://www.sciway.net/maps/government.html`, 2021.

[13] N. O. Stephanopoulos and C. Warshaw, *The impact of partisan gerrymandering on political parties*, Legislative Studies Quarterly, 45 (2020), pp. 609–643.

[14] The New York Times, *South Carolina U.S. Senate results: Tim Scott wins.* `https://www.nytimes.com/elections/2016/results/south-carolina-senate-scott-dixon`, Aug 2017.

[15] The New York Times, *South Carolina.* `https://www.nytimes.com/elections/2012/results/states/south-carolina.html?mtrref=www.google.com&gwh=60EBE922DC8EEA6BCA153BAED311E983&gwt=regi&assetType=REGIWALL`, 2012.

[16] The New York Times, *South Carolina election results 2016.* `https://www.nytimes.com/elections/2016/results/south-carolina`, Aug 2017.

[17] The New York Times, *South Carolina election results.* `https://www.nytimes.com/elections/2014/south-carolina-elections`, Dec 2014.

[18] The New York Times, *South Carolina governor election results.* `https://www.nytimes.com/elections/results/south-carolina-governor`, 2019.

[19] US Census Bureau Public Information Office, *Census bureau delivers South Carolina's census 2000 population totals for legislative redistricting; first race and hispanic data - census 2000 - newsroom - U.S. Census Bureau.* `https://www.census.gov/newsroom/releases/archives/census_2000/cb01cn47.html`, May 2016.

[20] USA Today, *2020 South Carolina state house election results.* `https://www.usatoday.com/elections/results/2020-11-03/state/south-carolina/lower/`, Nov 2020.

[21] A. M. Vagnozzi, *Detecting partisan gerrymandering through mathematical analysis: A case study of South Carolina*, Master's thesis, Clemson University, May 2020.

[22] G. S. Warrington, *Quantifying gerrymandering using the vote distribution*, Election Law Journal: Rules, Politics, and Policy, 17 (2018), pp. 39–57.

[23] Washington Post Staff, *South Carolina election results 2018.* `https://www.washingtonpost.com/election-results/south-carolina/`, Feb 2019.