A Computational Approach to Measuring Vote Elasticity and Competitiveness

Daryl DeFord*
CSAIL, Massachusetts Institute of Technology
Moon Duchin
Department of Mathematics, Tufts University
Justin Solomon
CSAIL, Massachusetts Institute of Technology

October 14, 2019

Abstract

The recent wave of attention to partisan gerrymandering has come with a push to refine or replace the laws that govern political redistricting around the country. A common element in several states' reform efforts has been the inclusion of *competitiveness metrics*, or scores that evaluate a districting plan based on the extent to which district-level outcomes are in play or are likely to be closely contested.

In this paper, we examine several classes of competitiveness metrics motivated by recent reform proposals and then evaluate their potential outcomes across large ensembles of districting plans at the Congressional and state Senate levels. This is part of a growing literature using MCMC techniques from applied statistics to situate plans and criteria in the context of valid redistricting alternatives. Our empirical analysis focuses on five states—Utah, Georgia, Wisconsin, Virginia, and Massachusetts—chosen to represent a range of partisan attributes. We study the feasibility and the partisan consequences of adopting and tightening various kinds of competitiveness requirements. We show that optimizing for competitiveness sometimes pulls partisan metrics to one side of their neutral baseline. These results highlight the importance of investigating possible new rules with careful mathematical modeling on real geospatial data to look for unintended consequences and for tensions with other redistricting goals.

Keywords: Redistricting, Gerrymandering, Markov Chains, Competitiveness

Contents

1 Introduction

In 2018 alone, five states saw redistricting reform measures passed by voters at the ballot box, and 2019 has already seen movement on laws or constitutional amendments in several more. A common feature in this surge of reform efforts is redoubled scrutiny of the rules, priorities, and criteria for new districting plans. In addition to reconsidering the treatment of traditional districting principles like population balance, compactness, and preservation of municipal boundaries, most of these state reforms explicitly address the use of partisan data and metrics in the redistricting process.

One reform approach in the recent trend has been to bar mapmakers from partisan considerations. Several states broadly prohibit favoring political parties or incumbents, while others go further and restrict mapmakers from directly using any partisan data while drawing the lines. At the opposite extreme, other reform approaches mandate that certain types of partisan symmetry or fairness metrics should be important decision criteria, effectively requiring that partisan data be taken into account during map construction. Competitiveness rules fall into the second category.[1]

This new attention to redistricting criteria is taking place in the context of an explosion of data and computational power in redistricting. One consequence of this new era of computational redistricting is that methods for locking in partisan advantage—and data proxies that make this possible without technically inputting election outcomes—create major opportunities for abuse. At the same time, computationally sophisticated ensemble techniques are improving dramatically at detecting partisan outliers. In this paper we devote our attention to a second, complementary application of algorithmically-generated ensembles of plans: we use them to study the effects of enacting different sets of rules and priorities.

Our main focus here is on rules favoring electoral competitiveness. Competitiveness is

---

[1]Some reform measures explicitly provide for a firewall between partisan-blind and partisan-aware parts of the process. For instance, Utah's successful ballot initiative in 2018 requires the maps to be drawn without considering partisan data but requires the expected partisan behavior to be measured before a plan can be approved [11]. Arizona requires similar behavior from its redistricting commission; the relevant portion of the redistricting law says that "Party registration and voting history data shall be excluded from the initial phase of the mapping process but may be used to test maps for compliance..." [31]. In these cases, the process is designed to prevent mapmakers from directly optimizing for the metrics and instead sets up a failsafe against intended or unintended partisan skew.

frequently regarded as a healthy attribute of elections that promotes voter engagement, responsiveness, and accountability of representatives, among other benefits.[2] During an election, competitiveness of a race might be measured by close polling or by candidate spending; after the fact, it is often assessed by a close outcome. It is much harder to capture the competitiveness level of districting plans at earlier stages, since they are created before there are outcomes, polling, or even a choice of candidates for the coming cycle. Accordingly, there is no agreed-upon means of measuring competitiveness and states have operationalized it in different ways when they have defined it at all. It is far from clear that a high competitiveness score can be achieved by a plan without other partisan consequences or unintended costs to other districting criteria, or that it will bring about outcomes that increase legitimacy and public trust. With these questions in mind, in this paper we develop statistical techniques to probe the consequences of various ways to quantify competitiveness of districting plans.

The primary reason that it is difficult to evaluate the potential consequences of altering or reordering redistricting criteria is that each set of rules and constraints constructs a different universe of valid plans, and those spaces are vast and complex. Since the set of valid plans typically is too large to enumerate, Markov chain Monte Carlo (MCMC) methods are standard techniques in applied mathematics and statistics for sampling in these situations, used to gather representative samples from the space of interest. Markov chain sampling is currently used by several groups of researchers to evaluate the partisan nature of proposed or enacted districting plans in court challenges [9, 10, 16, 20, 21]. These methods work by constructing large ensembles of plans that satisfy the rules set forth in each state and comparing the statistics of enacted plans to aggregate statistics of the distribution. All such techniques, however, require the choice of mathematical specifications of the legal constraints, which are rarely written with enough precision to determine an unambiguous formulation. Understanding the interactions between various districting criteria is fundamental for drawing meaningful quantitative conclusions from ensemble analysis, and we

---

[2]For instance, consider this language in Colorado's recent redistricting reform legislation: "Competitive elections for members of the United States House of Representatives provide voters with a meaningful choice among candidates, promote a healthy democracy, help ensure that constituents receive fair and effective representation, and contribute to the political well-being of key communities of interest and political subdivisions" [32].

hope that this study provides an invitation for more research in this direction.

1.1 Contributions

We analyze several variants of competitiveness metrics and compare the effects of prioritizing them in different ways. We use both *neutral ensembles* of alternative districting plans (i.e., generated without considering partisan data) and *heuristically optimized* plans that seek to extremize the metrics for Congressional and state Senate plans in five states.[3] We aim this discussion at two primary audiences. The first is legislators and reform groups who are actively working to formulate new redistricting rules, particularly those considering imposing competitiveness criteria. The second audience is researchers and quantitative analysts who are interested in assessing how well proposed plans uphold the stated criteria from which they were constructed. A message that emerges for both groups is the major impact of seemingly innocuous modeling choices against the idiosyncratic political geography of individual states. We argue that o universal metric of competitiveness should be adopted across the country, but that there is a fresh choice to be made in each circumstance. On the other hand, we do find that certain metrics are unreasonable and should *not* be adopted.

For instance, packing and cracking are the tools of the trade for gerrymandering, so it is notable that in certain circumstances, a competitiveness metric can approve or even incentivize packed and cracked plans. For instance, in a state with 68-32 voting for Party A, a plan that arranges A voters into some districts at 48% share and compensates that with other districts at 88% share could be textbook vote dilution while looking excellent to a metric that merely counts nearly-balanced districts. If an arrangement like that were the only or the predominant way to satisfy a competitiveness priority, then this would be important to know in advance of enactment. A principal contribution of our ensemble analysis is that it escapes the tendency to treat votes as though they are geographically unbound and fluidly rearrangeable, instead studying the actual elasticity of the vote by

---

[3]The only reason not to consider state House plans is that House districts tend to be too small to be made out of whole precincts while maintaining tolerable population deviation. Precincts are the smallest level at which we have authoritatively accurate election outcomes, which are a crucial element of this style of analysis. Vote results can be prorated to smaller units, but this is done at the cost of a potentially significant loss of accuracy.

comparing to alternative geographic partitions.

In our treatment, a general discussion of the reasons for preferring competitive districts in terms of broad alignment with societal and democratic goals will remain in the background. These normative questions are extremely important, but they are beyond the scope of the present paper and outside the reach of ensemble techniques, except to the extent that these techniques clarify the desirable and undesirable consequences of the norms.

1.2 Related work

Approaches to competitiveness in the political science literature vary widely. Although most recent reform efforts have focused on making more competitive districts, arguments have been made for instead packing districts with members of a single party to the extent practicable [3]. The opposite situation was considered in [17], where regression analysis was used to detect correlations between traditional districting criteria and competitive districts in the 1990 and 2000 cycles. They found that rules prioritizing population constraints as well as preservation of communities of interest and municipal boundaries were associated with more competitive districts. Additionally, their analysis helped expose the complexity of interactions between multiple redistricting criteria and the tradeoffs that are inherent in the process. Our work extends theirs by using the ensemble method to consider the potential impacts of directly incentivizing competitiveness, evaluating the properties of many alternative districting plans, rather than the currently enacted maps. Our analysis does not model the impacts of whether the lines are drawn by legislatures, courts, or commissions. Several groups of political scientists have studied the relationship between the identity of the line-drawers and the degree of competitiveness in subsequent elections [6, 13, 18, 19, 28], with mixed conclusions.

A recent article by Cottrell [12] presents an ensemble analysis to evaluate the impact of gerrymandering on the level of political competition around the country, concluding that gerrymandering has only a small cost in competitiveness because neutral processes also produce safe seats at rates similar to what can now be observed. We refer the reader to his references for a broad discussion of the arguments around competitive standards in the political science literature. This differs from our approach both in goals and methods,

but we view our problem domain as complementary to his. In terms of goals, Cottrell is looking for evidence of gerrymandering, while we are studying the impact of rules that favor competitive outcomes. That means he is comparing neutral ensembles to enacted plans, and we are focused on comparing neutral ensembles to methods for favoring competitive outcomes. Our analysis also uses a newly-developed Markov chain technique [14] with strong evidence of effective sampling.

1.3 Review of competitiveness language in legal criteria

Our research is motivated by several recent examples of redistricting rules introduced at the state level to promote competitive districts.

1.3.1 Washington, New York, Arizona

Some states had competitiveness language on the books before the recent reform wave, but not with quantitative details. Washington requires its commission to "provide fair and effective representation and to encourage electoral competition" and New York's rules require that "districts not be drawn to discourage competition" [34, 40].

Arizona's instructions to its commission include the following language: "To the extent practicable, competitive districts should be favored where to do so would create no significant detriment to the other goals" [31]. In practice, out of 9 districts, that seems to translate to a conspicuous effort to craft 3 competitive districts (53-47 outcomes or closer in 2012), leaving 4 safe Republican districts and 2 safe Democratic districts (all 60-40 or farther).

1.3.2 Colorado

In 2018, Colorado voters approved a ballot initiative that included a requirement that plans "maximize the number of politically competitive districts," where competitive is defined as "having a reasonable potential for the party affiliation of the district's representative to change at least once between federal decennial censuses" [32]. A literal reading of this language suggests that any district in which both parties have at least a 13% projected chance of winning might match the Colorado law, since $(1 - 0.13)^5 \approx 0.5$. While this interpre-

tation makes several potentially unrealistic assumptions, including that a point prediction would carry through over the full cycle and that there is no impact of incumbency on the elections, the text of the legislation does not provide a significant amount of guidance.[4]

1.3.3 Missouri

Missouri's 2018 constitutional amendment defines competitiveness in terms of a partisan metric called the *efficiency gap*, or *EG* [33]. It is one of the more detailed descriptions of a metric in the recent reform measures, requiring the state demographer to create a blended voting index by computing the turnout-weighted averages of the Presidential, Senatorial, and Gubernatorial elections that occurred during the previous census cycle. This pattern of votes is then modified by applying a uniform partisan swing of $\{-5, -4, \ldots, +4, +5\}$ and recomputing the efficiency gap against each modified vote dataset. Proposed districting plans are to be constructed so that each of the corresponding *EG* values is "as close to zero as practicable"—we call this *EG/swing zeroing*. We will show in §2.4 that this definition of competitiveness promotes drastically different qualitative properties of districting plans than the other metrics, and if taken literally it actually puts an *upper limit* on the number of districts that are allowed to fall in the 45-55% range for its voting index.

Missouri's reform is also noteworthy in that it places a higher priority on compliance with the priority on competitiveness than on contiguity or compactness of the plan.

1.3.4 New Jersey

Finally, in 2019, New Jersey's legislature debated a constitutional amendment that defined a competitive district to be one where the expected voting outcome is within 5% of the statewide average over the previous census cycle [22]. A plan was defined to be sufficiently competitive if 25% of its districts were competitive. The bill was widely pilloried (and ultimately defeated) because in a state whose recent voting trends $\approx 57\%$ Democratic, a district must have a Democratic majority to count as competitive under this language.

---

[4]This rule compares interestingly to the measure of competitiveness used by the 538 Atlas of Redistricting [5]. The Atlas project defined districts in the range from D+5 to R+5 as competitive, using the CPVI metric described in §2.3. Under their vote modeling, this corresponds to projected probabilities of approximately $18\% - 82\%$ of electing a candidate from a given party.

These examples and others demonstrate a range of efforts intended to quantify and promote "competitiveness" in districting plan design. Our analysis below defines mathematical formalizations in keeping with the legal language reviewed above to study potential impacts of implementation.

2 Competitiveness metrics

We consider several types of competitiveness metrics in this paper, motivated by the recently proposed legislation in several states discussed above.

2.1 Plans and votes

Characterizing a districting plan or individual district as competitive necessarily requires some assumptions about the underlying voter behavior. That is, if we denote a districting plan by $\mathcal{D}$ and a vote pattern by Δ, then all scores C of competitiveness are really functions of both: $C = C(\mathcal{D}, \Delta)$. The choice of Δ that allows the most robust assessment of a Congressional or Legislative districting plan is a matter of some debate. Some political scientists insist that it is important to use only *endogenous* data, or vote patterns from the same type of election, but the partisan preference information in these patterns is confounded by other variables such as incumbency, uncontested races, and particular candidate characteristics. Other scholars prefer the use of statewide *exogenous* vote patterns, but then must make a case about the choice of which election or elections to use to create an informative voting index. See for instance [2, 26, 1] for a paper and series of responses on this issue. In all cases, care is required in terms of predictive claims about future voting. We discuss the choice of Δ in more detail in §3.2.

Several states that require partisan analyses of proposed plans have addressed the choice of election data in their legislation. As noted above, Missouri specifies the exact election data that must be used to evaluate partisan performance [33], while Colorado simply requires [32] that "[c]ompetitiveness may be measured by factors such as a proposed district's past election results, a proposed district's political party registration data, and evidence-based analyses of proposed districts." New Jersey's proposed amendment [22], had it passed,

also would have specified exactly which data was to be considered. As with many of the other variables in redistricting legislation, there does not appear to be a consensus around the optimal modeling choice.

The results and visualizations reported below are based on a fixed vote pattern: we use Δ =Pres16, i.e., the precinct-level election returns from the 2016 Presidential election. No aspect of our method hinges on this choice, and it can be repeated for any other voting pattern in place of Δ.

2.2 Evenness and typicality

Perhaps the most intuitive measure of competitiveness for a district with respect to a vote pattern is the difference between the two-way vote share and 50%. This metric is implicit in Arizona's and Colorado's legal language. Up to scale, this measures the number of votes needed to change the outcome in the district. We might consider this a measurement of *evenness*, and it matches up well with standard ideas about competitiveness.

The Cook Partisan Voting Index (CPVI) is a widely used metric that is constructed by comparing the major-party vote share in each geographical unit—precinct, district, etc.—to the nationwide average across the two most recent Presidential elections [39]. Frequently, a district is described as competitive if its CPVI value is between D+5 and R+5, meaning that the district's vote is within 5% of the nationwide average. Upon inspection, most observers would agree that in the case that the average is skewed away from 50-50, this is not measuring competitiveness but rather *typicality*. Nonetheless, CPVI is reported by the Cook Political Report after each election as a measure of competitiveness across the nation; it also was used by 538's Atlas of Redistricting to count competitive districts [5]. New Jersey's proposed measure described above was based on an idea similar to CPVI, except that districts are compared to the statewide instead of nationwide average. Below, we will denote the statewide share of Democratic voters with D_0.

To measure the effects of quantifying and promoting competitiveness, we compute and compare metrics of evenness and typicality below with a generalized formulation that we call vote-band metrics.

2.3 Vote-band metrics

We start by defining a district to be in the (y, z) **band** with respect to voting pattern Δ, or simply to be a (y, z) district, if its Democratic vote share is within y of a target z. By default we will let y, z be denominated in percentage points, so that for instance a district is in the $(5, 50)$ band with respect to a voting pattern Δ if its vote share is between 45 and 55%.

Aggregating this behavior over the districts in a plan, we define $(\mathcal{D}, \Delta)$ to be (x, y, z) **compressed** if at least x share of the districts are in the (y, z) band. By varying the parameters x, y, and z, we can understand the elasticity of a vote pattern Δ in terms of its divisibility into districts at a certain scale.

A legal standard that employs this type of metric might require that adopted plans be (x, y, z) compressed, where x, y, z, and the choice of Δ are set by statute. This encompasses both many competitiveness rules (where the target is $z = 50$) and typicality rules (where the target is $z = D_0$ for state typicality). We will consider how plans that satisfy compression properties relate to unconstrained plans, along with studying how varying the parameters affects expected outcomes as a function of political geography.

Figure 1 shows how the currently enacted Congressional and state Senate plans in Utah, Georgia, Wisconsin, Virginia, and Massachusetts relate to competitiveness and state-typicality bands at $y = 5$, using 2016 Presidential vote data. Table 2 records the number of districts in those plans that fall into each band. Although none of these states had requirements governing competitiveness when the plans were drawn, they provide a starting example for considering the consequences of vote-band rules.

State	# Cong Districts	# Sen Districts	D_0 in Pres16	2012-16 CVPI
Massachusetts	9	40	64.7	D+12
Virginia	11	40	52.8	D+1
Wisconsin	8	33	49.6	Even
Georgia	14	56	47.3	R+5
Utah	4	29	37.6	R+20

Table 1: The five states considered in our study, chosen for a wide range of partisan tilts. The Democratic share D_0 is the two-party share in Pres16, so that for instance the sizeable third party vote in Utah is not reflected.
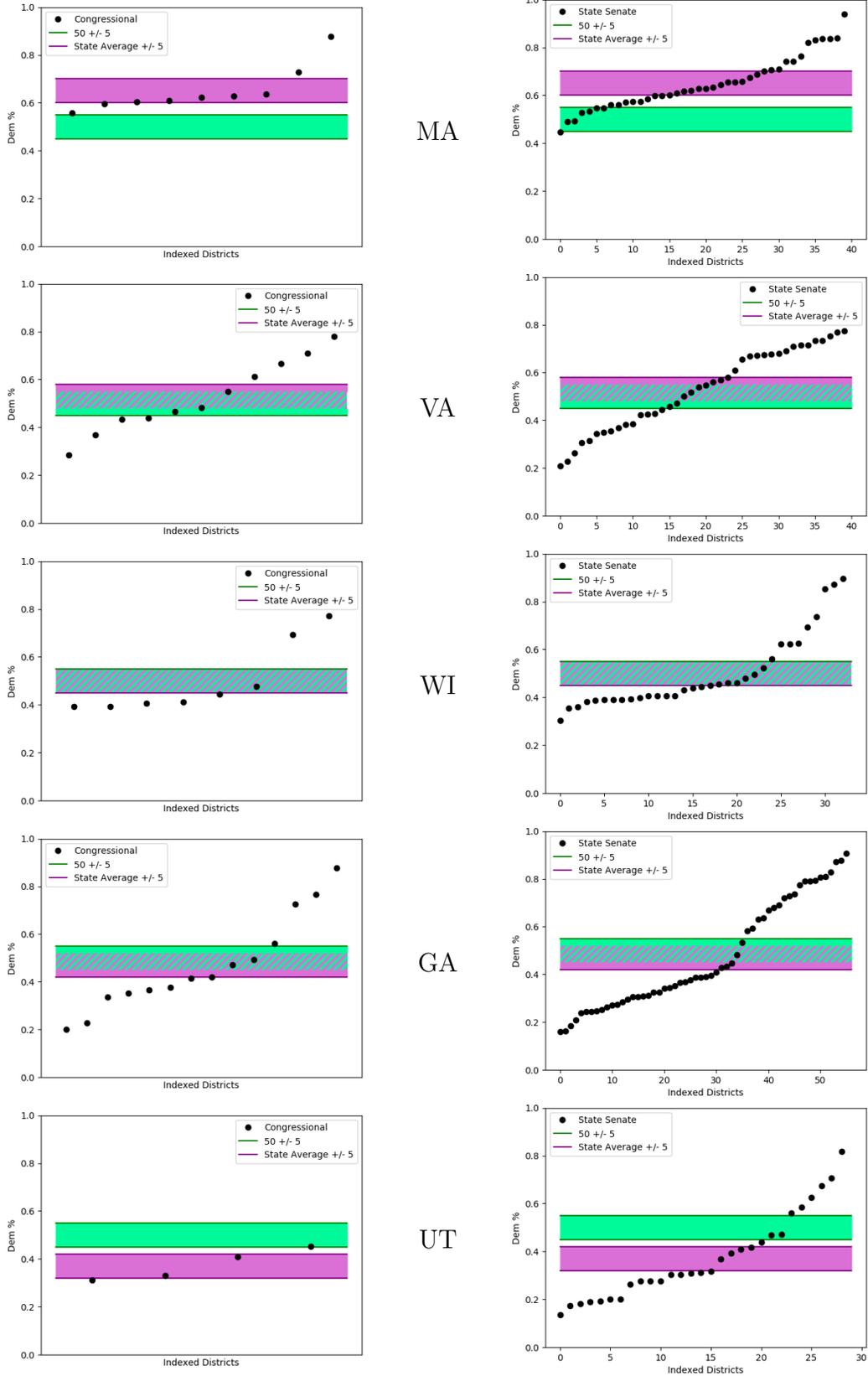
Figure 1: Comparison of the enacted Congressional and state Senate plans to $(5, 50)$ bands (green, competitive) and $(5, D_0)$ bands (purple, state-typical).

In most states, several districts lie outside both bands. Given fixed (y, z), there is no guarantee that it is even possible to construct a plan with a large number of (y, z) districts—even if z is near the statewide average—while adhering to reasonable compactness and boundary preservation norms. In Virginia, the enacted plan almost fails to have any districts that are competitive under both definitions, even though there is a large overlap between the bands.

One straightforward observation is that there must be some districts at least as far from 50% as the state average. Additionally, as shown in [15], discretization of geographical units can play a large role in determining the range of possible outcomes, since the voting population is not infinitely divisible.

Given a fixed vote share target z, it is natural to ask how the number of (y, z) districts varies as a function of the band width parameter y. Figure 2 shows the number of districts in the enacted plans that lie in competitive and state-typical bands as the width is increased. For most of the current plans, y values of approximately 15-20% would be needed to make half of the districts competitive; that is, about half the districts have outcomes even more skewed than 70-30. Also, as one would expect, it is harder to make even districts in the states that are themselves more partisan-tilted. Comparing Utah and Massachusetts to the other states shows the extent of this effect.

In our experiments, we will explore the interaction of vote-band metrics with political geography. Although this formulation is simple, we will see that the impacts of enforcing a choice of parameters x, y, and z do not fall uniformly over our test states.

2.4 EG/swing zeroing

As discussed above, Missouri has adopted a definition of competitiveness for redistricting that requires the State Demographer to compute a metric called the *efficiency gap* (described below) for any proposed districting plan, with respect to a particular synthetic vote pattern created by a weighted combination of recent election outcomes. They are also required to consider ten alterations of that vote pattern, obtained by uniformly swinging the district vote shares by $\{\pm 1, \pm 2, \pm 3, \pm 4, \pm 5\}$ and computing the corresponding efficiency gap values. A literal reading of the constitutional text requires that each of those
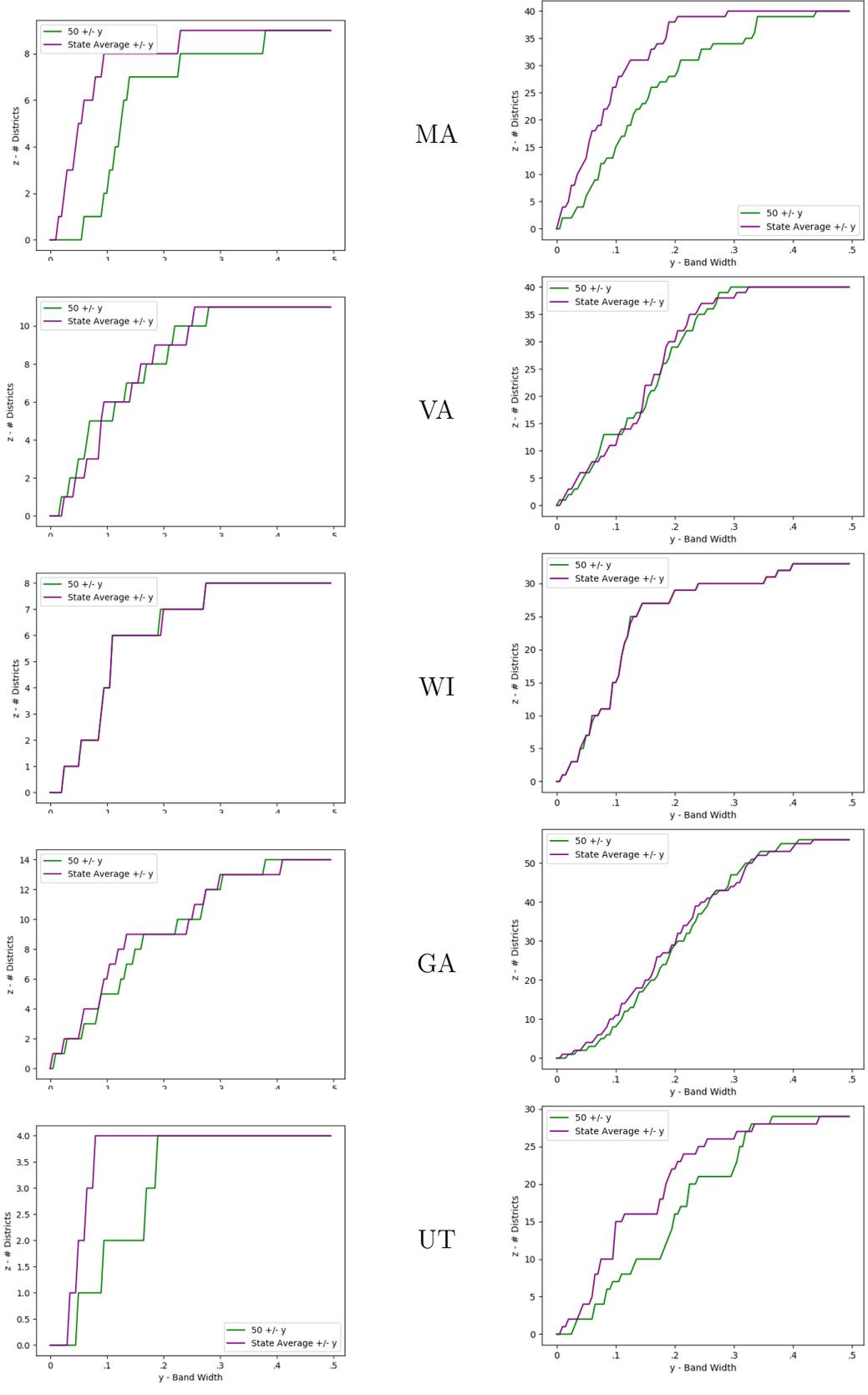
Figure 2: Number of districts in the enacted Congressional and state Senate plans that are $(y, 50)$ (green, competitive) and (y, D_0) (purple, state-typical), as y varies. For instance, all four of Utah's congressional districts are within 10% of the state average.

14

eleven efficiency gap values should be made as small as practicable. Due to the close relationship between efficiency gap and seat outcomes, many authors have noted that $EG \approx 0$ prescribes a "winner's bonus" with slope two in the votes to seats conversion [35, 36]. Here, we show that this constraint is directly at odds with vote-band competitiveness.

Let D_0 represent the statewide percentage of Democratic voters and S represent the number of seats won by the Democratic party, out of k total seats. Then, as derived in [36], the efficiency gap is

$$EG(\mathcal{D}, \Delta) = 2\left(\frac{D_0}{100}\right) - \frac{S}{k} - \frac{1}{2} + N(\mathcal{D}, \Delta),$$

where $N(\mathcal{D}, \Delta)$ is a noise term depending S, k, and the average turnout in districts won by each party, which is zero in the case of equal turnout. Noting that D_0 is fixed and neglecting the noise term, we must minimize this expression to satisfy the constitutional requirement, which amounts to securing a number of Democratic seats S as close as possible to $k(2\frac{D_0}{100} - \frac{1}{2})$. Repeating this procedure for each of $\{D_0 - 5, D_0 - 4, \ldots, D_0 + 4, D_0 + 5\}$ gives us a sequence of eleven prescribed seat values, $S_{-5} \leq S_{-4} \leq \cdots S_4 \leq S_5$. But we further observe that prescribing how the seat values change as you shift the votes 5 points up and down is precisely equivalent to prescribing the number and position of seats in the $(5, 50)$ band.

To show the potential impacts of this EG/swing zeroing rule, consider Missouri and Wisconsin. With respect to Pres16, Missouri had roughly $D_0 = 40$. Continuing to neglect turnout noise, we observe that $EG(\mathcal{D}, \Delta + i) = 2(\frac{D_0 + i}{100}) - \frac{S}{8} - \frac{1}{2}$, which prescribes $S = 2$ Democratic seats for shifts between $R + 5$ and $R + 0$ and by $S = 3$ seats for $D + 1$ through $D + 5$. This can *only* be satisfied by having one district with a Democratic share of between 49 and 50 percent and no others between 45 and 55 percent. If the Democratic shares by district were roughly $(28, 28, 28, 28, 49.5, 80, 80)$, the plan would earn perfect marks for EG-swing competitiveness. Having more districts close to even would rate strictly worse.

In Wisconsin, which had roughly $D_0 = 49.6$, the EG-swing standard requires two districts between 45 and 55 percent and six districts outside that range. Again, this optimization promotes the creation of many safe districts and is indifferent to the possibility that those are actually made into landslide districts.

15

Matters are even worse in Utah ($D_0 = 37$, $k = 4$), where adherence to this rule would require that Democrats are ahead in precisely one Congressional district out of four, for all of the swing values. That is, a compliant Utah plan would not be allowed to have *any* districts within 45 and 55 percent.

This construction directly opposes the vote-band (and common-sense) definition of competitiveness, which prefers having more districts close to even. This seems so clearly problematic that we will not continue to model more refined consequences of adopting this kind of metric below, but will focus on vote-band definitions.

3 Data and methods for ensemble analysis

We analyze the potential effects of enforcing a vote-band rule as described in §2.3 in two sets of experiments, using data from five states. First, we generate a large neutral ensemble of districting plans for each state and filter the plans according to increasingly strict vote-band constraints, comparing the properties of the compressed plans to those of the full ensemble. Next, we use two heuristic optimization techniques to generate plans that score well on the given competitiveness measure and compare the statistics of the optimized plans to that of the full ensemble.

3.1 State data

In our study, we focus on Georgia, Massachusetts, Utah, Virginia, and Wisconsin, as exemplar states that display a wide range of partisan leans (see Table 1). The CVPI baseline over this time period is approximately 51.5%, as the Democratic party candidate narrowly won the popular vote in both the 2012 and 2016 elections.

Our methodology uses a discrete formulation of redistricting, viewing a plan as a graph partition for each given state. This is defined with respect to the *precinct dual graph*, which has a node for each precinct and an edge between two nodes if the corresponding precincts are adjacent. Given a fixed number of districts k, a districting plan is a partition of this graph into k connected subgraphs. Each node is decorated with the population of the corresponding precinct and its vote counts with respect to the election Δ, so the

population and partisan performance of each district can be computed by summing over the nodes. This discrete approach has enabled a growing literature on sampling techniques in court cases and in scientific literature [8, 10, 14, 20].

To construct the dual graph for each state, we use precinct shapefiles with population and vote data joined.[5]

3.2   Partisan metrics

The choice of Pres16 as the vote pattern Δ for these runs makes it easier to compare results across states and is consonant with the CVPI approach and with several other political science analyses. However we note that the resemblance between Presidential and Congressional or Legislative voting patterns varies greatly from state to state (and year to year), and we do not feel that the lessons learned from this analysis hinge on there being a close match.

Nevertheless, we can observe that the number of districts with a Democratic majority in Pres16 matches the actual Congressional 2016 outcome in MA, GA, and UT (with 9/9, 4/14, and 0/4 Democratic-majority districts, respectively), while assigning one D seat too many in Virginia and one too few in Wisconsin (where the enacted plan laid over Pres16 voting gives 5/11 and 2/8 Democratic-majority districts, respectively). We have also carried out the trials reported below with respect to recent U.S. Senate and Gubernatorial elections where data was available; results are available on request, and the findings are broadly similar, with the same big-picture messages about the relative elasticity and the between-state variability of voting patterns.

To evaluate the partisan side effects of competitiveness rules on districting plans, we

---

[5]The underlying geographic and partisan data for each state was originally obtained from the following sources. **Georgia:** Geography from the Georgia General Assembly Legislative and Congressional Reapportionment Office [24]; partisan data from MIT Elections Data and Science Lab (MEDSL) [23]. **Utah:** Geography from the Utah Automated Geographic Reference Center [7]; partisan data from MEDSL. **Wisconsin:** All geographic and partisan data from the Wisconsin Legislative Technology Services Bureau [4]. **Virginia:** All geographic and partisan data from the Princeton Gerrymandering Project OpenPrecincts project [30]. **Massachusetts:** Geography from the Massachusetts Secretary of the Commonwealth and partisan data from the Massachusetts Secretary of the Commonwealth Elections Division [29]. Staff of the Metric Geometry and Gerrymandering Group joined population data from the 2010 US Census to these datasets by aggregating up from census blocks, then used geospatial tools to clean the geographical data. The processed shapefiles and metadata are publicly available for download [27].

use three partisan metrics, again without taking a stance on which of these is the best or most useful. Rather, we are looking for the presence of interrelations. The first partisan metric is simply the number of districts in which the number of Democratic votes exceeds the number of Republican votes (discussed above in this section). A second metric, defined above in §2.4, is the efficiency gap. Finally, a third popular partisan metric is the *mean-median (MM) score*, a measure of asymmetry that is defined by subtracting the mean Democratic vote share across the districts of a plan from the median of the same list of values. Popularized by Gary King and others, this metric is sometimes described as measuring how much one party could fall short of 50% of the statewide votes while still receiving 50% of the seats [25, 38]. Symmetry metrics are explicitly included in Utah's reform language from 2018 [11].

3.3   Ensemble generation

We use open-source software `GerryChain` [37] both for generating the neutral ensembles and for optimizing for competitive districts. Replication code and data are included in the Supplementary Information. `GerryChain` uses a Markov chain to generate districting plans, beginning with an initial state and making iterative changes to the assignments of the nodes. Our ensembles are constructed using the `ReCom` Markov chain procedure overviewed in [14], which uses a spanning tree method to bipartition pairs of adjacent districts at each step.

For the neutral ensembles, we required the districts to be connected and population-balanced, imposing no further restrictions.[6] We constrain the population to within 2% of ideal for Congressional districts and within 5% for state Senate districts. For each state, we took 100,000 `ReCom` steps for the Congressional plans and 1,000,000 steps for state Senate plans, beginning at a random seed. See [14] for a workflow describing standard heuristic convergence tests such as independence of seed and stability under increases in run-length.

For the heuristic optimization ensembles, we used two types of hill-climbing proposals

---

[6]As described in [14], `ReCom` produces plans with compactness scores in range of human-made plans without any need to impose additional compactness constraints. Compactness is measured by a discrete metric called *cut edges*, which counts how many pairs of units were adjacent in the state but are assigned to different districts—these are the edges that would have to be cut to separate the whole graph into its district subgraphs.

to find elevated numbers of $(5, 50)$ districts. We call these methods Opt1 and Opt2 and construct 100 plans using each. To generate each plan, we start with a number of unconstrained ReCom steps (200 for Congress, 1000 for Senate) and then propose 1,000,000 Flip steps, each attempting to change the assignment of a single node on the boundary. In the Opt1 run, proposed plans are rejected if they decrease the number of $(5, 50)$ districts. In the Opt2 run, a proposed plan is accepted if it decreases the sum of the distances between the district Dem shares and the 45-55 percent band. Because Flip steps tend towards poor compactness scores, these steps are constrained to have no more than twice as many cut edges as the partition after the ReCom steps.

4 Ensemble results

In this section, we report the outcomes of the chain runs described above. The key takeaway is that partisan metrics are sensitive to both the formulation of the competitiveness metric and to the underlying political geography, which controls the elasticity of the vote. For this reason, great care should be taken to evaluate potential impacts of these rules, using data representing the actual distribution of voters across the state, before adoption.

4.1 Partisan baselines

The best-known application of ensemble methods is to construct a baseline for any districting metric of interest, across a wide variety of plans that incorporate the state's political and physical geography. Figure 3 shows the distributions of the vote share across districts for each of the full ensembles, and Figure 4 shows the distribution of an auxiliary partisan score of symmetry introduced above, called the mean-median score.

In Table 2 we show the average number of $(5, 50)$ and $(5, D_0)$ districts across the ensemble, compared to the enacted plans. The distribution of values observed across the ensemble is similar to those of the enacted plans, with the exception of Georgia, where the ensemble finds many more districts in both bands.

The baselines reveal the extent to which the various states' political geographies shape the ensemble statistics. With the exception of Virginia, whose MM distribution is centered
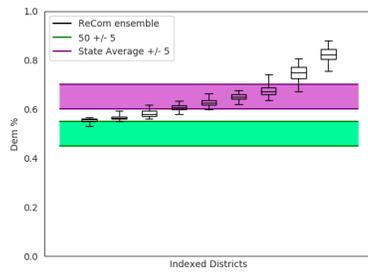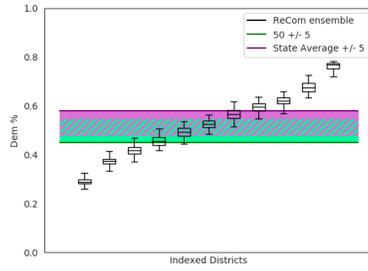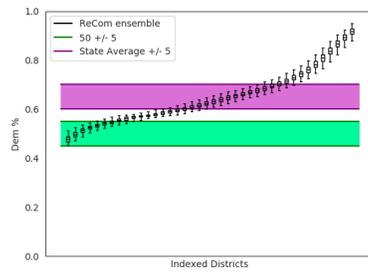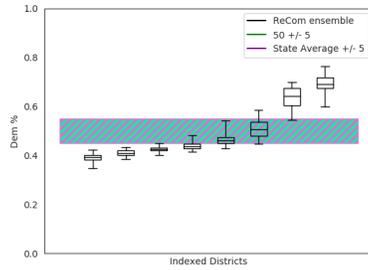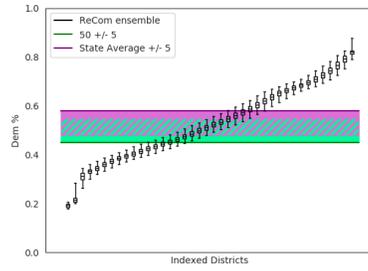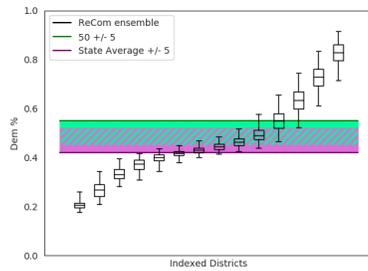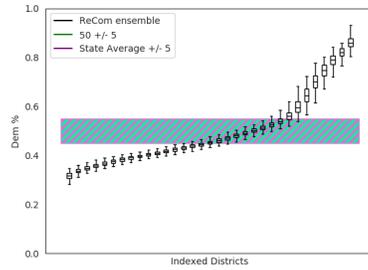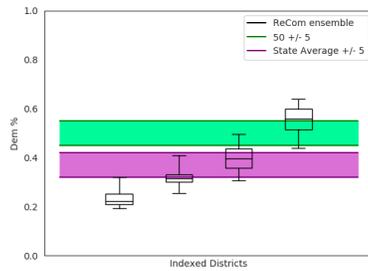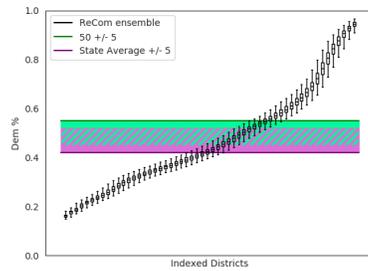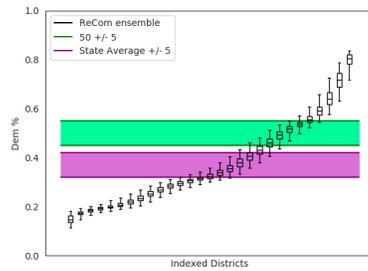
Figure 3: Partisan share boxplots by district for the neutral ensembles in Congress (left) and state Senate (right).

State, Type (#)	Enacted $(x, 5, 50)$	Ensemble $(x, 5, 50)$	D_0	Enacted $(x, 5, D_0)$	Ensemble $(x, 5, D_0)$
MA Cong (9)	0	0.24	64.7	5	3.85
VA Cong (11)	3	2.79	52.8	2	2.82
WI Cong (8)	1	1.76	49.6	1	1.81
GA Cong (14)	2	2.63	47.3	2	4.43
UT Cong (4)	1	0.62	37.6	2	1.05
MA Sen (40)	6	6.85	64.7	13	12.82
VA Sen (40)	6	8.60	52.8	6	8.00
WI Sen (33)	7	8.80	49.6	7	9.10
GA Sen (56)	2	8.57	47.3	4	8.78
UT Sen (29)	2	4.09	37.6	4	4.57

Table 2: Comparison of the number of $(5, 50)$ and $(5, D_0)$ districts in the enacted plans to the averages across the neutral ensembles. Most cases, the enacted plan has slightly fewer $(5, 50)$ districts than the ensemble average—the only examples where the enacted plan has more than the mean are Virginia's and Utah's Congressional plans. Georgia stands out from the group in the other direction, showing signs that the lines have been arranged to significantly reduce the number of seats that are competitive or near the statewide average.

at zero, all of the states have significantly Republican-favoring average mean–median scores of between two and five percent.[7] For several ensembles, more than 90% of the plans have a Republican favoring score. This adds to a rapidly growing body of evidence that shows that applying uniform standards across states, without accounting for how voters are distributed throughout each state, can lead to misleading results and unattainable ideals. For example, a plan whose MM value is typical for the Wisconsin ensemble would be a significant outlier if evaluated against the Virginia data, even though the difference in statewide averages is small.

The seats outcomes let us measure the extent of partisan advantage that is entailed by political geography. For instance, Wisconsin's vote pattern which was very close to 50/50 in Pres16, but the most Democratic-favoring plans found by the state Senate ensemble have $14/33 \approx 42\%$ Democratic seats.

---

[7]The standard narrative around mean–median scores holds that $MM = .05$ means that the Republican party could expect half of the seats with only 45% of the votes.

4.2 Satisfiable vote-band rules

Too loose of a vote-band rule will impose no constraint at all on redistricting; too tight of a rule will be impossible to meet. In this section, we examine the level of constraint imposed by $(y, 50)$ and (y, D_0) rules. This effect is explored in Figure 6. The left column shows the existence and proportion of plans in the ensemble that are $(x, y, 50)$ compressed, and the right column shows similar plots for (x, y, D_0). The transition regions for each fixed x value are relatively small as a function of y; but it is precisely these transition regions in which a vote-band rule would have any impact. Outside these zones, a vote-band requirement is either vacuous or is so hard to satisfy that a large ensemble finds no examples. At a minimum, states must perform modeling of this kind using recent electoral patterns before adopting a numerically specific vote-band rule for competitiveness or typicality.

For all of the states, the neutral ensemble found Congressional plans with no $(5, 50)$ districts—that is, it is always possible to draw a plan with no districts at all between 45 and 55% partisan share. Even within a fixed state, different vote elasticity may be observed at different scales. Comparing the behavior of Georgia's Congressional ensemble to the state Senate ensemble shows that requiring a fixed proportion of seats to be competitive would impose a much more constraining condition on the Congressional districts than the state Senate.

4.3 Partisan impacts of vote-band winnowing

One way that vote-band rules might play out is that among contending plans, those that best satisfy the rules would be selected. To model this scenario, we now extract subsets of plans that are (x, y, z) compressed and compare the aggregate properties of these plans to the full ensemble distributions. This allows us to observe whether the competitive plans differ from the neutral ensemble in a systematic fashion under the partisan metrics.

We begin by computing the average number of Democratic seats in districting plans that are $(x, 5, 50)$ and $(x, 5, D_0)$ compressed, as a function of x. Figure 7 shows these results for each state we analyzed. For states with a strong statewide party skew, increasing the number of districts close to the statewide average unsurprisingly tends to shift the seats outcome in favor of the party with the larger statewide percentage. However, in tilted

states, vying for competitiveness can actually hurt the minority party. The Massachusetts state Senate results show an example of a potentially surprising consequence of enforcing a competitiveness constraint. For this ensemble, increasing the number of $(5, 50)$ districts increases the expected number of Democratic seats. This is because it is easier to create more seats with a Democratic share of roughly 55% by taking scarce Republican votes from the few Republican-leaning districts.

The case of Wisconsin ($D_0 \approx 50$) is particularly interesting, since a vote-band competitive rule for Congress imposes a Democratic advantage, while a similar rule for state Senate has a Republican lean.

Next, we compare the distributions of mean-median values on the winnowed sub-ensembles, shown in Figure 8. Again, the histograms show a range of different behaviors as the constraints are tightened; several cases are worth highlighting. For the case $z = 50$, both Wisconsin and Georgia, whose full ensemble averages are Republican-favoring, are forced towards the center as the constraint tightens, while Virginia, which was balanced in the full ensemble, becomes more Republican-favoring in both the Congressional and state Senate ensembles under progressively tighter winnowing. Contrast this behavior with Utah, where the winnowed Congressional ensemble leans approximately four points towards Democratic-favoring scores, while the state Senate ensemble looks even more extremely Republican-favoring.

4.4 Partisan impacts of heuristic optimization

While the previous trials investigated the properties of competitive plans that were observed in a neutral ensemble, laws mandating competitiveness will also incentivize mapmakers to *optimize* for competitive plans. Plans constructed with this aim in mind likely will be significantly different than those drawn without access to partisan information. To expose some of these differences, we use the partisan-aware optimization methods Opt1 and Opt2 (described in §3.3) to construct plans that have significantly more $(5, 50)$ competitive districts than were found by the neutral ensemble.

To give an extremely crude sense of the success of these short runs with simple optimization heuristics, we compare our the optimization runs to those expressly constructed

for the 538 Atlas of Redistricting to minimize CPVI values (Table 3). The comparison is not perfect, since the underlying data, models, and evaluation mechanism differ.[8] The similarity between our results and theirs, however, suggests that our simple optimization heuristics have some power. An interesting direction for future work would be to use the precinct-sorting methods from [15] to get global upper bounds on the possible number of competitive districts.

State, Type, (Total #)	538 Atlas of Redistricting	Neutral Max	Opt1 Max	Opt1 Mean	Opt2 Max	Opt2 Mean
MA Cong (9)	**4**	2	3	2.11	1	0.05
VA Cong (11)	**9**	7	8	7.33	**9**	6.61
WI Cong (8)	6	5	6	3.99	**8**	4.28
GA Cong (14)	7	8	12	9.41	**14**	10.68
UT Cong (4)	1	**2**	**2**	1.99	**2**	.51
MA Sen (40)	–	12	**17**	15.35	9	5.12
VA Sen (40)	–	17	26	23.63	**28**	20.32
WI Sen (33)	–	17	20	17.33	**25**	15.83
GA Sen (56)	–	19	**37**	32.99	31	20.85
UT Sen (29)	–	9	**14**	11.42	13	6.93

Table 3: Comparison of the number of $(5, 50)$ districts found in the neutral and two styles of hill-climbing ensembles to the 538 Atlas of Redistricting's highly competitive plans. Maximum values highlighted in bold. In each state except Massachusetts, hill-climbing found plans where at least half the districts are $(5, 50)$ competitive, which differs significantly from the results of the neutral ensemble.

Comparing to the box plots of Figure 3 gives a measure of how far these optimized plans are from those in the neutral ensemble. For Georgia Congressional plans, the 1%-99% whiskers of the box plots only touch the $(5, 50)$ band for five out of 14 districts, but the optimization method found plans where all 14 districts are in the $(5, 50)$ band. Once again, we find Georgia to be particularly elastic in its vote distribution and to have enacted plans that are conspicuously uncompetitive. (On the other hand, we have not made an effort to include the full suite of Georgia districting criteria in our model.)

Figure 9 shows the number of Democratic seats against the number of $(5, 50)$ districts found by the two optimization methods. As with our previous examples, the effects of promoting plan competitiveness are unpredictable at best, both within states or between

---

[8]For example, the CVPI values use a slightly Democratic-favoring baseline as a result of the popular vote in the 2016 presidential election, which makes it slightly easier to make competitive districts in Democratic-leaning states and slightly harder in Republican-leaning ones. Additionally, the 538 maps had constraints on the number of majority-minority districts.

Congressional and state Senate plans in the same state. The plot further highlights that different optimization methods targeting competitiveness can drastically change the partisan character of the plans that are found.

For Georgia's Congressional plan, higher numbers of competitive districts correlates with Republican seat advantage. Wisconsin's Senate plans show the opposite correlation (at least for these optimization techniques), tending to find more Democratic seats as the number of competitive districts increases. The optimization ensembles also found seats outcomes that did not appear at all in the significantly larger neutral ensembles, which is a reminder that outcomes that are extremely unlikely under partisan-blind circumstances may nonetheless be relatively easy to construct intentionally.

As with the winnowing results, these optimization runs show that the potentially extreme impacts of enforcing a competitiveness measure depend heavily and unpredictably on political geography and scale.

5   Conclusion

In the national conversation around redistricting reform, competitiveness is only one among numerous priorities that are being debated and sometimes hastily written into law. Mapping out the interactions between various values, criteria, and constraints is becoming increasingly important. Data-intensive modeling is indispensable to make informed decisions about these interactions and to identify unintended consequences of quantitatively specific rules.

Ensemble analysis—and Markov chain sampling in particular—is a useful tool for exploring these complex spaces and intertwined constraints. The present study sets up a model methodology for surveying the state-specific partisan landscape in advance of setting quantitative thresholds and priorities for state reform.

For the specific case of competitiveness, our results show that vote-band rules can be set in a way that does impose nontrivial constraints on redistricters, but sometimes with ancillary effects on other partisan metrics. There is every reason to expect nontrivial effects on the ability to uphold other priorities, from racial statistics to political boundary integrity, which were not considered here.

Like for any other partisan quantity, securing a high score on a competitiveness metric does not follow effortlessly from the absence of nefarious intent. Democratic deliberation is called for both to clarify the type of competitiveness that is sought—more close outcomes? fewer landslides? etc.—and to weigh this against other priorities. As states continue to experiment with their rules and frameworks for redistricting, we hope that statistically robust ensemble methods become standard tools in the repertoire.

References

[1] Robin E. Best, Shawn J. Donahue, Jonathan Krasno, Daniel B. Magleby, and Michael D. McDonald. Authors' Response– Values and Validations: Proper Criteria for Comparing Standards for Packing Gerrymanders. *Election Law Journal: Rules, Politics, and Policy*, 17(1):82–84, November 2017.

[2] Robin E. Best, Shawn J. Donahue, Jonathan Krasno, Daniel B. Magleby, and Michael D. McDonald. Considering the Prospects for Establishing a Packing Gerrymandering Standard. *Election Law Journal: Rules, Politics, and Policy*, 17(1):1–20, September 2017.

[3] Thomas L. Brunell. Rethinking Redistricting: How Drawing Uncompetitive Districts Eliminates Gerrymanders, Enhances Representation, and Improves Attitudes toward Congress. *PS: Political Science & Politics*, 39(1):77–85, January 2006.

[4] Wisconsin Legislative Technology Services Bureau. Open data portal, 2019.

[5] Aaron Bycoffe, Ella Koeze, David Wasserman, and Julia Wolfe. The atlas of redistricting. *FiveThirtyEight*, 2018.

[6] Jamie L. Carson and Michael H. Crespin. The effect of state redistricting methods on electoral competition in united states house of representatives races. *State Politics & Policy Quarterly*, 4(4):455–469, 2004.

[7] Utah Automated Geographic Reference Center. Voting precincts, 2019.

[8] Jowei Chen and Jonathan Rodden. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8:239–269, 2013.

[9] Maria Chikina, Alan Frieze, Jonathan Mattingly, and Wesley Pegden. Practical tests for significance in Markov Chains. *arXiv:1904.04052 [math, stat]*, April 2019. arXiv: 1904.04052.

[10] Maria Chikina, Alan Frieze, and Wesley Pegden. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11):2860–2864, March 2017.

[11] Utah Code. Chapter 7 Title 20A Chapter 19 Part 1 Section 103.

[12] David Cottrell. Using Computer Simulations to Measure the Effect of Gerrymandering on Electoral Competition in the U.S. Congress. *Legislative Studies Quarterly*, 44(3):487–514, 2019.

[13] James B. Cottrill. The Effects of Non-Legislative Approaches to Redistricting on Competition in Congressional Elections. *Polity*, 44(1):32–50, January 2012.

[14] Daryl DeFord, Moon Duchin, and Justin Solomon. Recombination: A family of Markov chains for redistricting. *Preprint*, 2019.

[15] Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen. Locating the representational baseline: Republicans in Massachusetts. *arXiv:1810:09051*, 2018.

[16] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. A new automated redistricting simulator using Markov chain Monte Carlo. *Preprint*, 2019.

[17] Richard Forgette, Andrew Garner, and John Winkle. Do Redistricting Principles and Practices Affect U. S. State Legislative Electoral Competition? *State Politics & Policy Quarterly*, 9(2):151–175, 2009.

[18] Corbett A. Grainger. Redistricting and Polarization: Who Draws the Lines in California? *The Journal of Law & Economics*, 53(3):545–567, 2010.

[19] John A Henderson, Brian Hamel, and Aaron Goldzimer. Gerrymandering incumbency: Does non-partisan redistricting increase electoral competition? *Journal of Politics*, 80, 2018.

[20] Gregory Herschlag, Han Sung Kang, Justin Luo, Christy Vaughn Graves, Sachet Bangia, Robert Ravier, and Jonathan C. Mattingly. Quantifying Gerrymandering in North Carolina. *arXiv:1801.03783 [physics, stat]*, January 2018. arXiv: 1801.03783.

[21] Gregory Herschlag, Robert Ravier, and Jonathan C. Mattingly. Evaluating Partisan Gerrymandering in Wisconsin. *arXiv:1709.01596 [physics, stat]*, September 2017. arXiv: 1709.01596.

[22] New Jersey Legislature. Senate Concurrent Resolution 43.

[23] MIT Election Science Data Labs. Election data, 2019.

[24] Georgia Legislative and Congressional Reapportionment Office. Statewide voting precincts, 2019.

[25] Michael D. McDonald and Robin E. Best. Unfair Partisan Gerrymanders in Politics and Law: A Diagnostic Applied to Six Cases. *Election Law Journal: Rules, Politics, and Policy*, 14(4):312–330, November 2015.

[26] Eric McGhee. Rejoinder to "Considering the Prospects for Establishing a Packing Gerrymandering Standard". *Election Law Journal: Rules, Politics, and Policy*, 17(1):73–82, October 2017.

[27] MGGG. Mggg-States GitHub Repository, 2019.

[28] Peter Miller and Bernard Grofman. Redistricting Commissions in the Western United States. *UC Irvine Law Review*, 3(3):637, August 2013.

[29] Massachusetts Secretary of the Commonwealth Elections Division. Election results, 2019.

[30] Princeton Gerrymandering Project. Openprecincts.org, 2019.

[31] Arizona State Constitution. Article IV Part 2 Section 1.

[32] Colorado State Constitution. Article V section 44.

[33] Missouri State Constitution. Article III Section 3.

[34] Washington State Constitution. Article III Section 43.

[35] Nicholas Stephanopoulos and Eric McGhee. Partisan gerrymandering and the efficiency gap. *University of Chicago Law Review*, page 831–900, 2014.

[36] Ellen Veomett. Efficiency gap, voter turnout, and the efficiency principle. *Election Law Journal*, 17.

[37] VRDI. Gerrychain. *GitHub Repository: mggg/gerrychain*, 2018.

[38] Samuel S.-H. Wang. Three Practical Tests for Gerrymandering: Application to Maryland and Wisconsin. *Election Law Journal: Rules, Politics, and Policy*, 15(4):367–384, October 2016.

[39] David Wasserman and Ally Flinn. Introducing the 2017 Cook Political Report Partisan Voter Index. 2013.

[40] New York State Constitution. Article III Sections 4 and 5.
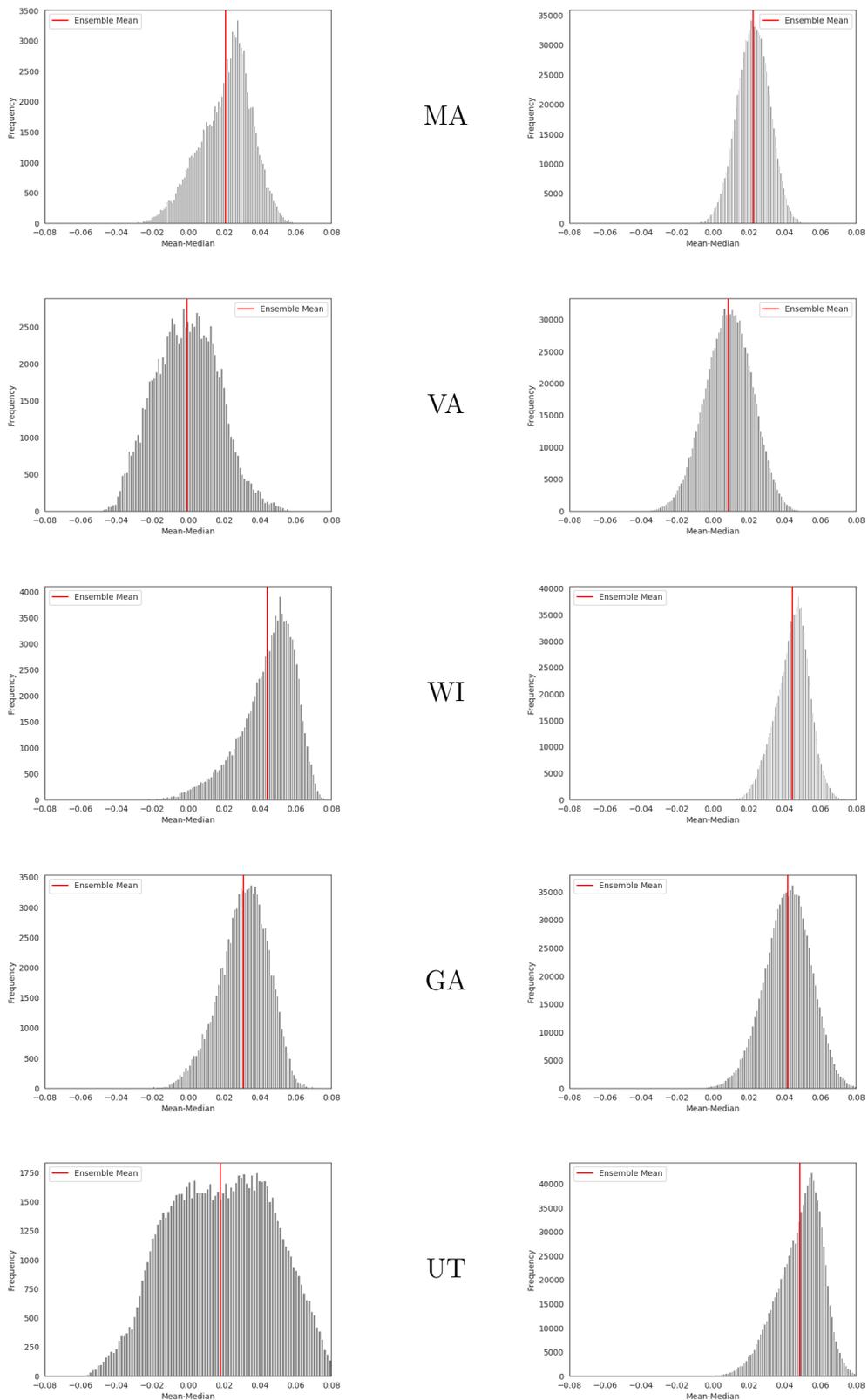
Figure 4: Mean-median baselines for the neutral ensembles with ensemble means marked in red. The x-axis runs from $-.08$ to $.08$ in each plot. All states except Virginia show pronounced average Republican advantages of approximately 2-5%.

30
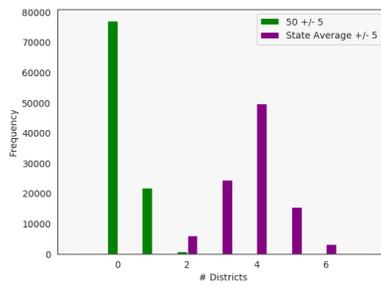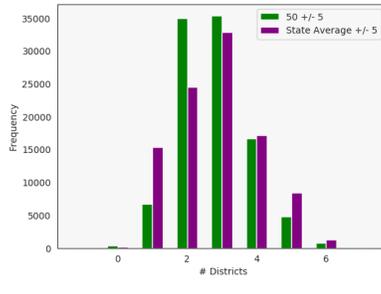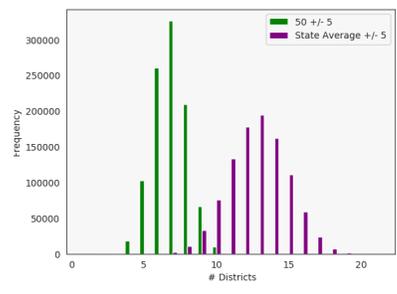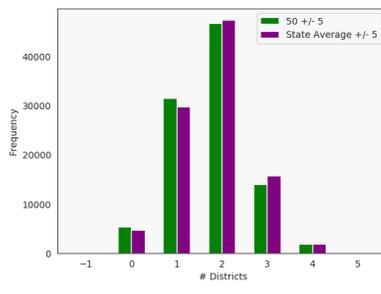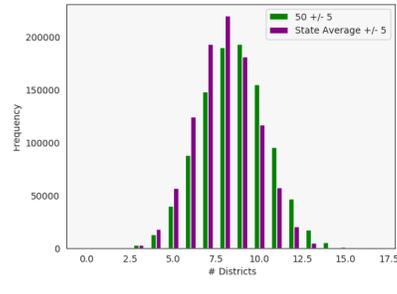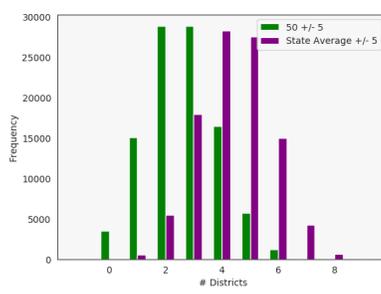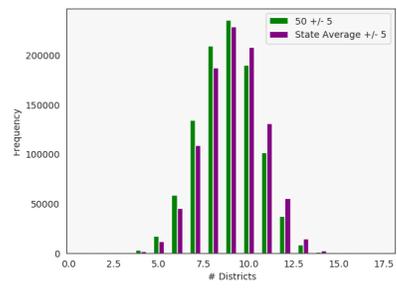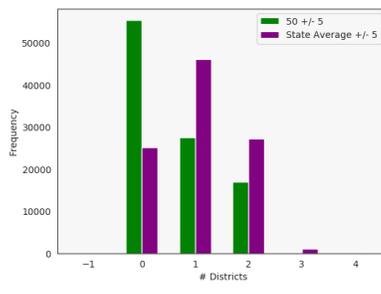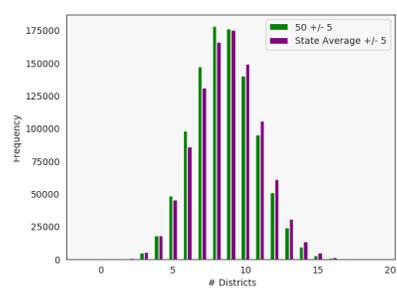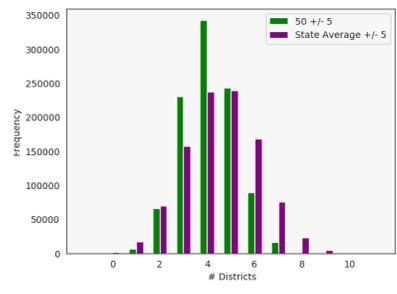
Figure 5: How often each number of $(5, 50)$ and $(5, D_0)$ districts was observed in the neutral ensemble. Scale effects are notable in Georgia especially, where the competitiveness and state-typicality statistics are the same for state Senate but noticeably different for Congress.
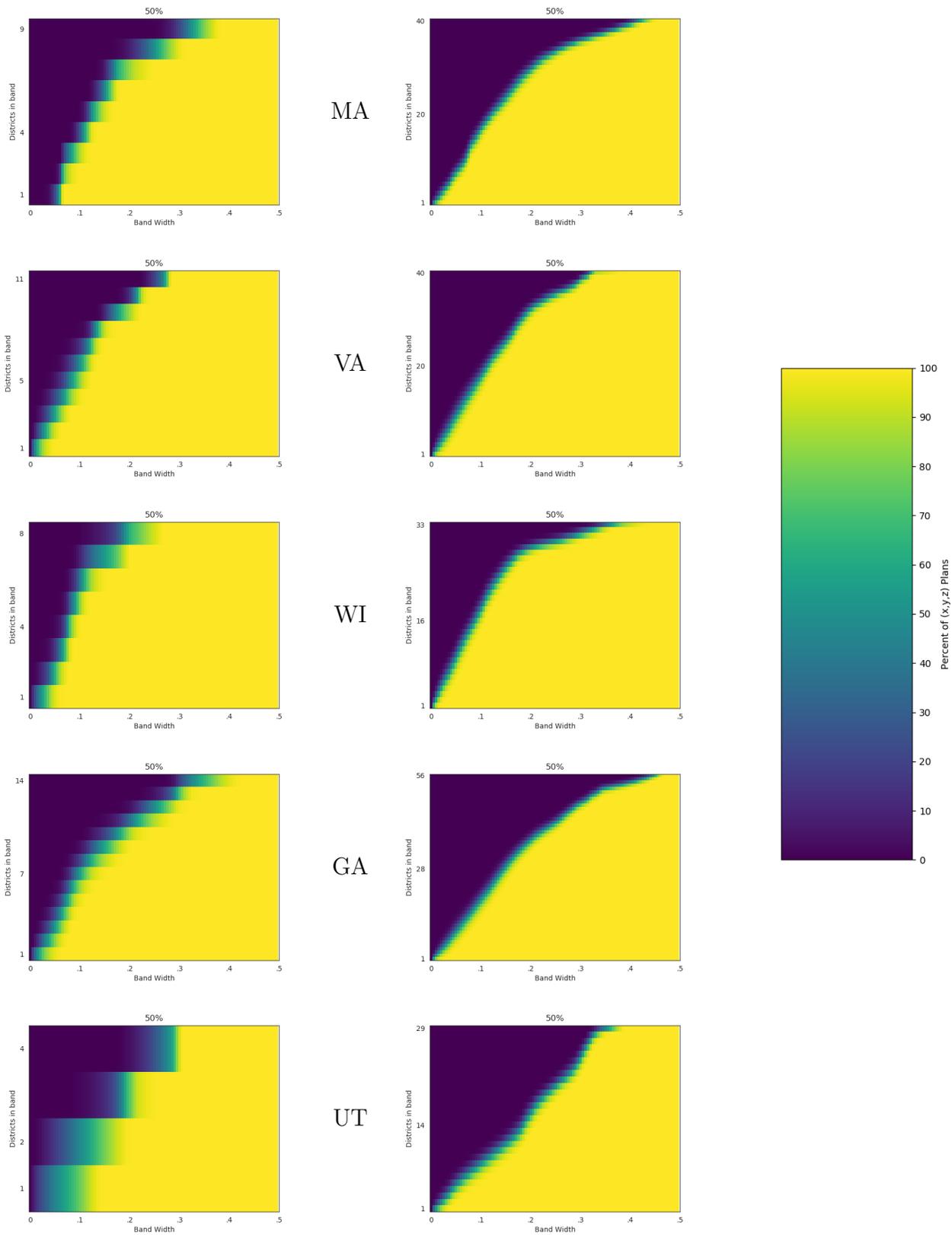
Figure 6: Visualizing the narrow range of vote-band rules that meaningfully constrain a neutral ensemble. Congress (left), state Senate (right).
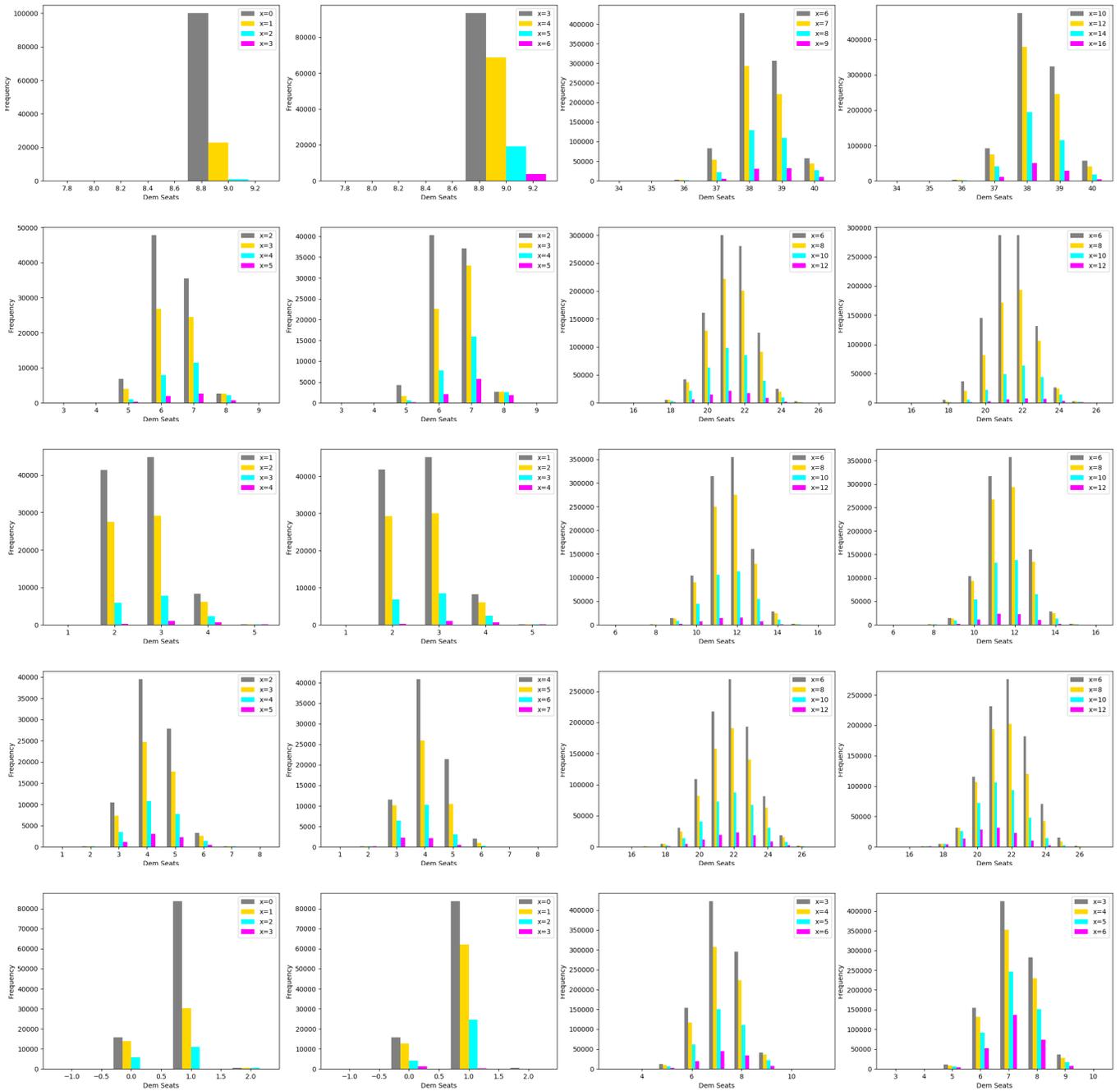
Figure 7: Histograms of Democratic seat outcomes in (x, y, z) compressed plans as a function of increasing x. Columns 1-2: Congress, $z = 50$ and $z = D_0$. Columns 3-4: state Senate, $z = 50$ and $z = D_0$. States, top to bottom: MA, VA, WI, GA, UT.
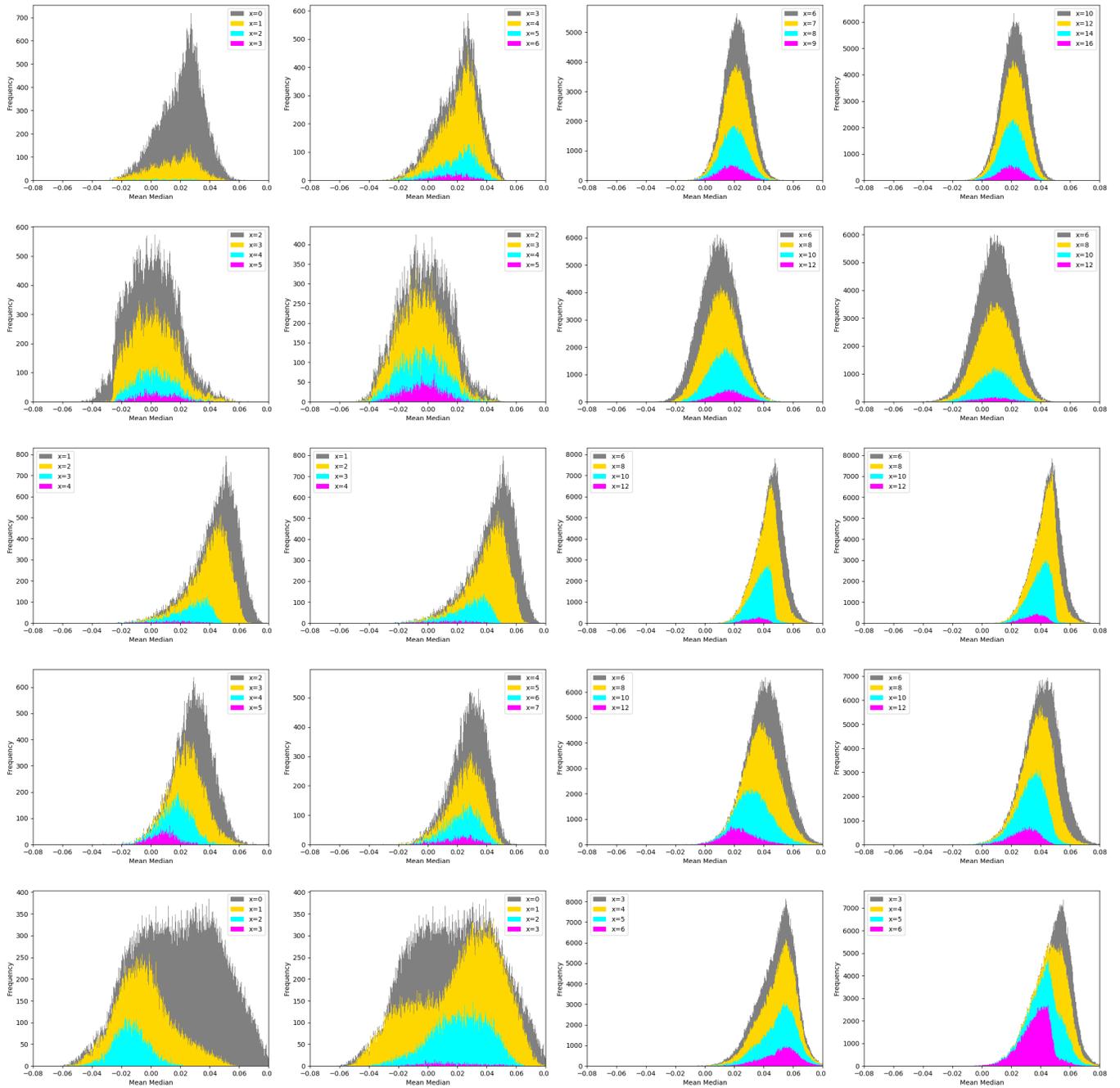
Figure 8: Successively winnowed mean-median distributions for $(x, 5, z)$ compressed plans as a function of increasing x. Columns 1-2: Congress, $z = 50$ and $z = D_0$. Columns 3-4: state Senate, $z = 50$ and $z = D_0$. States, top to bottom: MA, VA, WI, GA, UT.
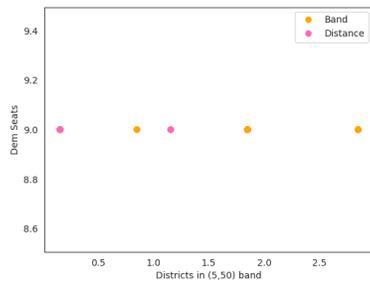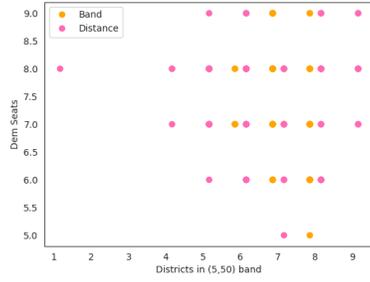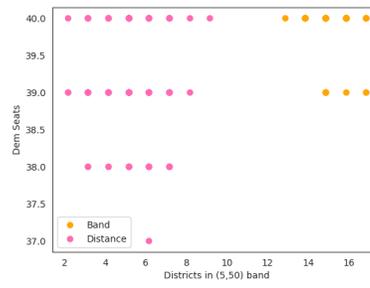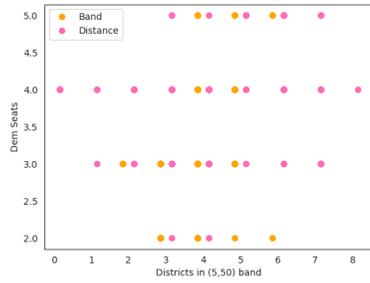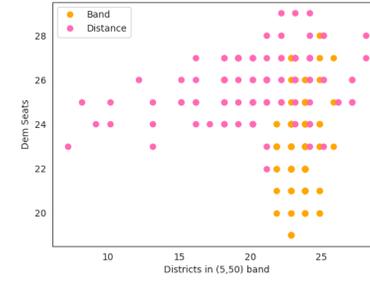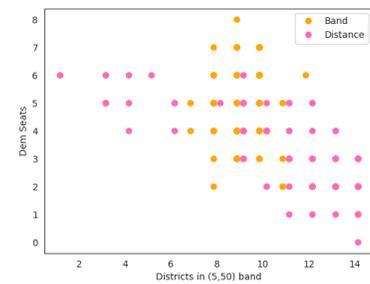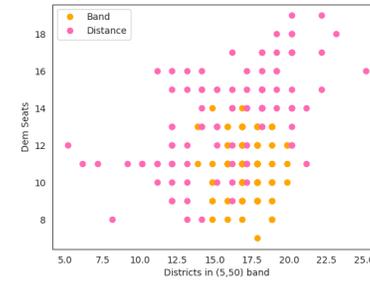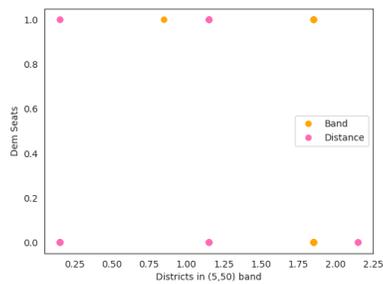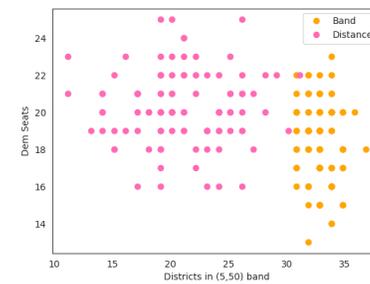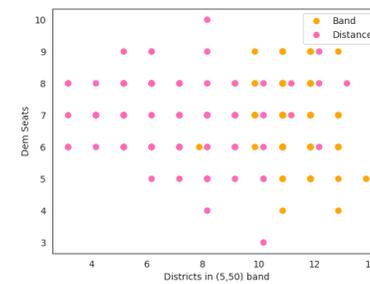
Figure 9: Number of Democratic seats in the plans found by the two optimization methods plotted against the number of $(5, 50)$ districts.