Redistricting with Endogenous Candidates*

Paola Moscariello†

September 1, 2025

LINK TO MOST RECENT VERSION

Abstract

I study partisan gerrymandering when district composition affects candidates' policy positions and, consequently, voters' behavior. In the U.S., primary elections determine which candidates compete in general elections, with a district's ideological composition shaping who emerges as the nominee. Thus, redistricting affects not only which party wins but also the ideology of competing candidates. I find that classical gerrymandering strategies can backfire when candidates emerge endogenously, particularly in districts where extreme voters may select non-viable candidates. However, when properly designed to account for both voter affiliation and preference intensity, gerrymandering can be a more powerful instrument than traditional approaches that consider only party affiliation. I show how methods from optimal transport theory can be used to characterize the optimal redistricting plan, which creates districts that maximize ideological distance between competing candidates. Using these findings, I analyze two implications for the U.S. House of Representatives: how gerrymandering contributes to political polarization and its consequences for minority representation.

Keywords: Gerrymandering, optimal transport, information design, polarization.

---

*A special recognition goes to my advisor Leeat Yariv. I am grateful to Roland Bénabou, Roberto Corrao, Sylvain Chassang, Germán Gieczewski, Faruk Gul, Matias Iaryczower, Alessandro Lizzeri, Sofia Moroni, Xiaosheng Mu, Pietro Ortoleva, Jacopo Perego, Wolfgang Pesendorfer, Alejandro Sabal, Fëdor Sandorminskiy, Carol Shou and to participants in seminars and conferences for their constructive comments. I gratefully acknowledge the funding provided by the William S. Dietrich II Economic Theory Center.

†Department of Economics, Princeton University. Email: pm27@princeton.edu.

1 Introduction

Partisan gerrymandering is the deliberate manipulation of electoral district boundaries to create an unfair advantage for a particular political party. A notable example is the 2012 Republican Party's Redistricting Majority Project (REDMAP). Despite Republican candidates receiving 1.4 million fewer votes than their Democratic counterparts in U.S. House elections, REDMAP's strategic redistricting efforts resulted in a 33-seat majority for the Republicans (Daley, 2020). Traditionally, redistricters use data from presidential elections to predict whether a certain geographic area is predominantly Republican or Democratic. Perhaps as a consequence, voters' behavior in each district is assumed to be impervious to redistricting efforts themselves.

This paper studies optimal redistricting when candidates' policy positions and, consequently, voters' behavior, respond to the districts in place. I embed a two-stage model of elections at the district level into an optimal redistricting problem and show how traditional gerrymandering can backfire when extreme voters select non-viable primary candidates. Using methods from optimal transport theory (Monge, 1781; Kantorovich, 1942), I show how savvy redistricters can create more powerful plans than standard redistricting allows for. Their strategy? Crafting district boundaries that drive a wedge between moderates and extremists in the opposing party, effectively turning their rivals' diversity into a liability.

Illustrative examples. To illustrate the forces in my model, it is useful to recall the standard redistricting setting. In the example in Figure 1, there's a U.S. state with a finite population of voters. Two-thirds are Democrats (in blue) and one-third are Republicans (in red). A Republican redistricter has to partition voters into, say, three equipopulous districts, and wants to maximize the number of districts with a Republican majority. She applies the standard "pack-and-crack" technique. That is, she creates two "cracked" districts, District 1 and 2, consisting of just enough Republicans to have a majority, and one "packed" district, made up of only Democrats. Thanks to gerrymandering, Republicans win two-thirds of districts with just one-third of overall Republican supporters. This is a classic result and one core point of the effects of gerrymandering.

The approach above, however, assumes that party positions are fixed and exogenous. Instead, suppose that party positions in each district are endogenous. Consider Figure 2.

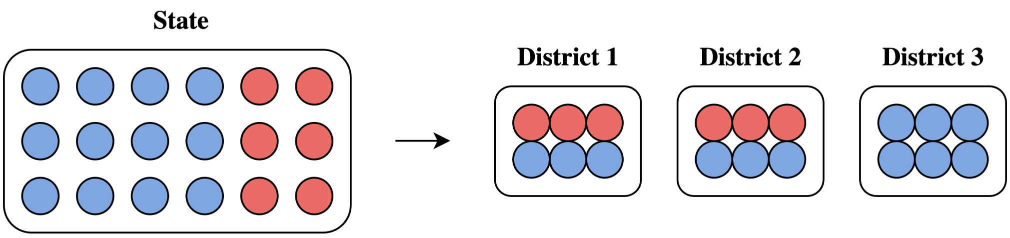


Figure 1: Standard pack-and-crack redistricting.

In the population, a small fraction of Democrats and an even smaller fraction of Republicans are moderates, represented by less intense shades of blue and red, respectively; all other voters are partisans (or extremists). Suppose primary elections precede the statewide elections. In each district and for each party, primaries determine the candidate, who can be either a moderate or an extremist, depending on which voter type constitutes the majority within their party. In general elections, while partisans consistently vote for their party's candidate, moderate voters prefer a moderate candidate from *either* party over an extreme candidate from their own party. The standard "pack-and-crack" gerrymandering strategy can backfire. For example, consider District 1 and District 2 in Figure 2. In both districts, there is a majority of partisans among Republicans and a majority of moderates among Democrats. Hence, the Republican candidate caters to partisans, while the Democratic candidate caters to moderates. The only moderate Republican voter prefers voting for the moderate Democratic candidate rather than for a partisan Republican. Redistricting effectively converts the moderate Republican voter to a Democratic voter! Ultimately, with endogenous candidates, classical gerrymandering fails to achieve its goal: in this example, *all* districts end up selecting a Democratic candidate.

A real-life equivalent of such "dummymandering", as a bad gerrymandering is usually referred to in policy circles, can be found in the case of Oregon's fifth district: after Democrats redrew boundaries in anticipation of the 2022 mid-term elections, progressive Jamie McLeod-Skinner unexpectedly defeated seven-term centrist incumbent Kurt Schrader in the Democratic primary. Crucially, McLeod-Skinner's victory hinged on a forty-point advantage in Deschutes County, which was only added to the district through a recent redistricting effort. This shift in Democratic candidate allowed Republican Lori Chavez-De Remer to appeal to moderate voters and flip the district for the first time since 1994 (Flaccus, 2022; Glueck, 2022; Scott and Weigel, 2022).

Do such examples and real-life instances mean that gerrymandering is less powerful

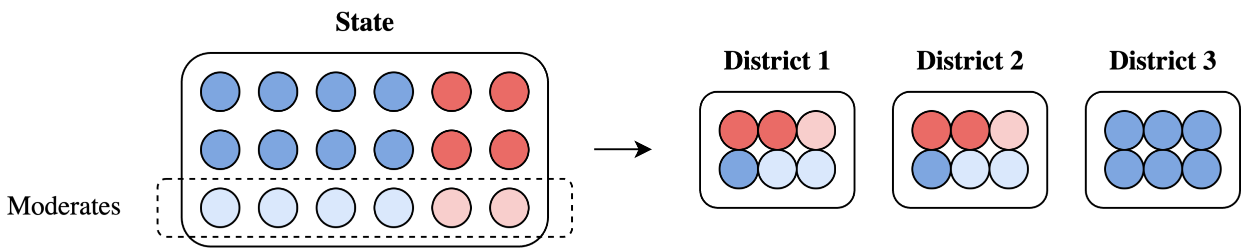


Figure 2: Pack-and-crack backfires.

than previously thought? Quite the opposite. Consider a scenario where a Republican redistricter is aware of and exploits the endogeneity of voter behavior. Figure 3 illustrates an optimal gerrymandering plan. In District 1, despite having a Democratic majority, the composition is heterogeneous. There are sufficient Democratic partisans so that the Democratic candidate caters to their preferred policy, while the two moderate Democrats align more closely with the Republican candidate, who caters to the single moderate Republican voter in the district. Consequently, moderate Democrats prefer to vote Republican, even though they would have preferred a moderate Democratic candidate to a moderate Republican. As a result, Republicans win this district. District 3 consists of a majority of partisan Republicans, ensuring a straightforward Republican win. By strategically considering the endogeneity of candidates' selection and voters' behavior, the redistricter manages to secure *all districts* for the Republican party, despite Republicans constituting only one-third of the overall population. That is, gerrymandering can be more powerful than previously thought if its impact on voting behavior is taken into account.

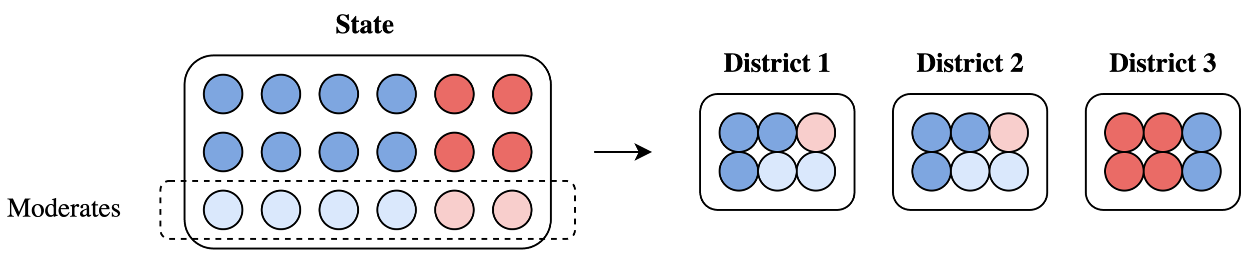


Figure 3: Gerrymandering is even more powerful than previously thought.

Model. There is a continuum of voters with single-peaked preferences over a unidimensional policy space. Voters are identified by their bliss point, which I sometimes refer to as their "type." Republicans are voters with bliss points above an exogenous cutoff, while Democrats are voters with bliss points below the cutoff.

A Republican designer partitions voters into equipopulous districts to maximize the number of districts won by Republicans.

The model incorporates a two-stage election process at the district level to determine voting behavior. First, in the primary elections, Democratic and Republican voters separately select their party candidates. Then, in the general election, all district voters choose between the two primary winners. Drawing on the work of Owen and Grofman (2006), I model electoral incentives so that primary candidates position themselves to appeal to their respective party medians.[1] To close the model, in districts lacking Republican or Democratic voters, I assume the corresponding party's candidate defaults to the most moderate position within their ideological spectrum. The general election outcome is decided by the district's median voter. The candidate whose position most closely aligns with the district median wins the election.

Given that redistricting in the U.S. takes place roughly every ten years, it is conceivable that redistricters cannot perfectly anticipate future populations' preferences at the time of redistricting. Therefore, I allow the designer to face uncertainty about voters' preferences, parametrized by a one-dimensional aggregate shock. In other words, after the designer commits to their redistricting plan, but before elections take place, a random shock occurs that uniformly shifts all voters' preferences by the same amount.

Results. In Section 2, I show that the problem can be simplified by restricting attention to a specific subset of feasible plans. Proposition 1 establishes that each optimal district must contain exactly two voter types: one type from below the population median and one type from above it, in equal proportions.

The intuition for this result emerges from the examples above. First, Figure 2 demonstrates how pack-and-crack backfires when District 1 mixes moderate and extreme Republican types. The designer can perfectly hedge against this possibility by creating cracked districts with no more than one Republican type, ensuring the optimal plan performs at least as well as standard pack-and-crack. Second, Figure 3 reveals a novel opportunity: In Districts 1 and 2, the ideological divide between moderate and extreme Democrats al-

[1] In the Appendix, I provide microfoundations. The strict dependence on party medians is not necessary for my results: the driving force behind the model is responsiveness of candidates to party preferences. In Section 5, I provide an extension where candidates' positions are determined by a general quantile of the preference distribution within each party.

lows Republicans to win despite being in the minority. This same mechanism extends to districts packed exclusively with Democrats, where a moderate Republican candidate can peel off enough moderate Democratic voters to secure victory. To maximize this advantage, each such district should contain exactly two voter types: a higher type serving as moderate Democrat and a lower type as extreme Democrat. This hybrid strategy—packing Democrats while cracking them between moderates and extremists—allows the optimal plan to strictly outperform traditional pack-and-crack.

The effectiveness of gerrymandering depends critically on voter type differences within a district. The larger the wedge between the two voter types, the higher the probability of winning the district, making districts with significant gaps "safer" than more homogeneous ones. While the designer would ideally maximize this gap for all districts, she is constrained by the overall distribution of voter types. She therefore has to decide which voter types to allocate to safer, more heterogeneous, districts, and which to less safe, more homogeneous, districts.

As Theorem 1 summarizes, this task can be formalized as a Monge-Kantorovich optimal transport problem, where the designer determines how to optimally "transport" voter types from above the population median to match with types below it. The solution to this problem depends on the distribution of uncertainty about voter preferences.

In Section 3, I characterize such solution under the assumption that the aggregate shock is S-shaped around zero.[2] Proposition 3 demonstrates that there is a unique optimal redistricting plan. Voters far from the population median are matched in a "positive assortative" or "comonotonic" manner: high types above the median are paired with high types below it. However, for voters close to the median, the matching reverses to "negative assortative" or "anti-comonotonic": high types above the median are paired with low types below it.

The intuition is as follows. The districts most likely to be won are those with large ideological gaps between their two voter types, which requires matching voters from the right tail of the distribution with voters from the left tail. When working with these extreme voters, there's no benefit in creating an excessively large gap in one district at the expense of another. Instead, the designer uses positive assortative matching to maintain

[2]The cumulative distribution function (CDF) of the shock is convex below zero and concave above it. Prominent cases of symmetric noise considered in the literature—e.g., normally-distributed noise—satisfy this assumption.

consistent, sufficient gaps across all such districts. However, closer to the median, ideological gaps become harder to create and districts become less likely to be won. Here, the designer switches to negative assortative matching: creating some very safe districts with large gaps at the expense of other districts that are given up as lost.

Under the additional assumption that the shock is symmetric, I am able to determine further which voters are matched positively versus negatively assortatively across the distribution. Proposition 4 serves as the analogue of a first-order condition in my framework: the designer must maximize the number of districts with positive assortative matching, where extreme types above the median pair with moderate types below it.

Intuitively, because the designer's objective is linear in the number of districts won, the symmetry assumption on the shock effectively makes the problem independent of the shock variance. Then, the designer's optimal strategy is to create as many winning districts as possible by maintaining just enough ideological gap to secure victory in each. Positive assortative matching, pairing extreme types above the median with moderate types below it, achieves exactly this: it creates consistent, sufficient gaps without making any district excessively safe.

This characterization yields clear comparative statics. When Republicans are sufficiently numerous, the designer can afford to match voters positively assortatively throughout the entire distribution. As the Democratic share of the population grows, the designer must increasingly resort to negative assortative matching, pairing extreme voters from both sides of the median together.

Implications. In Section 4, I explore implications of my gerrymandering model that are pertinent to current political and legal debates.

First, I examine how optimal gerrymandering strategies can exacerbate polarization in the U.S. House of Representatives. Proposition 5 predicts a notable ideological gap among elected representatives, who are either moderate Republicans or extreme Democrats. Notably, extreme Democrats emerge in districts with a significant share of moderate voters, challenging the conventional view that political extremism stems from segregating extreme voters into homogeneous districts.

Second, I discuss the implications for minority representation in congressional districts. In "minority opportunity" districts, where minorities make up 40-50% of voters, minority candidates' success depends critically on white voter support. My model suggests this dependence can trigger “white backlash,” where white voters unite against minority-preferred candidates, benefiting Republican candidates as predicted by my framework.

1.1 Related Literature

Optimal partisan gerrymandering. This paper primarily relates to the economic theory literature on optimal partisan gerrymandering, starting with Owen and Grofman (1988). While their work focuses on a binary voter type model, subsequent research has emphasized the importance of considering voters’ preference intensities in redistricting strategies. In existing models, preference intensities play a role mostly due to uncertainty about voter preferences, typically assuming moderate voters are more susceptible to preference swings. Friedman and Holden (2008) show that with significant aggregate preference uncertainty, the optimal redistricting plan assigns extreme supporting voters to the same districts as extreme opposing voters, effectively neutralizing the latter’s influence. The authors term this approach “matching slices,” which is reminiscent of what I term negative assortative matching. Kolotilin and Wolitzky (2024) develop a more comprehensive model of gerrymandering that allows for both aggregate preference uncertainty and substantial shocks that affect voters independently, namely idiosyncratic uncertainty.[3] Their findings suggest that when idiosyncratic uncertainty dominates, the optimal redistricting involves segregating the most extreme opposing voters while matching the remaining voters in a negative assortative manner. Importantly, their empirical analysis indicates that scenarios where idiosyncratic uncertainty outweighs aggregate uncertainty are most relevant in practice. Gul and Pesendorfer (2010) examine competition between two parties, each controlling redistricting in distinct areas. In their optimal plan, opposing voters are segregated and more favorable voters are all pooled together.

I contribute to this literature by allowing individual voters’ behavior to depend on the districts in place. One implication of this approach is that it results in a different treatment of extreme opposing voters. While previous literature either predicts such extreme voters to be segregated or matched with extreme supporters, my model exploits them to turn moderate voters against their own party. To do so, the optimal plan prescribes “mis-

[3] In their companion paper, Kolotilin et al. (2023), they show how their problem can be connected to optimal transport.

matched slices," on either side of the preference distribution, resulting in novel political and legal implications, as I discuss in Section 4.

Information design. As Kolotilin and Wolitzky (2024) show, the gerrymandering problem can be mapped onto an information design problem. The distribution of voter preferences serves as the "prior," districts function as "posteriors," and a redistricting plan represents a distribution of districts that satisfies a constraint, which is mathematically equivalent to Bayes plausibility. Indeed, in the special case of exogenous policies, my model becomes a variant of Bayesian persuasion (Kamenica and Gentzkow, 2011) for medians, which can be solved with recent off-the-shelf tools (Yang and Zentefis, 2024), as I show in Section 2. My solution sheds light on information design problems where payoffs depend on more intricate aspects of posterior distributions, such as the relative positions of conditional medians, rather than single summary statistics like means or medians.

Optimal transport. To characterize the solution, I leverage a Monge-Kantorovich optimal transport representation of the redistricter's problem (Monge, 1781; Kantorovich, 1942). Drawing on results from Chiappori et al. (2010) and Santambrogio (2015), I establish the existence and uniqueness of the solution and characterize it for key benchmark cases. I use related techniques to characterize the solution to an optimal transport problem where the surplus function is symmetric and S-shaped.

Other topics in gerrymandering. The broader literature on gerrymandering tackles a variety of different issues. The effects of redistricting on policy choice is considered by Shotts (2002) and Besley and Preston (2007), while the impact of gerrymandering on polarization in the House of Representatives is addressed, for instance, by McCarty et al. (2009). Other important topics in redistricting that I do not explore in this paper relate to: including geographic constraints on gerrymandering (Puppe and Tasnádi, 2009), accounting for differential voter turnout (Bouton et al., 2023), and measures of electoral maldistricting (Gomberg et al., 2023).

2 The model

2.1 Setup and Statement of the Problem

Voters and Parties. Consider a continuum of voters with single-peaked preferences over uni-dimensional policy space $[\underline{v},\overline{v}]$, a closed interval on the real line $\mathbb{R}$. A voter's ideal point, sometimes referred to as her type, is denoted by $v \in [\underline{v},\overline{v}]$, with population distribution $\phi \in \Delta([\underline{v},\overline{v}])$[4], assumed to have a strictly increasing and continuously differentiable cumulative distribution function (CDF), F, on the interior of $[\underline{v},\overline{v}]$. I normalize the median of F to be $v^m = 0$. Moreover, I assume voters' preferences to be symmetric around their ideal points[5].

There are two parties, the Democratic and the Republican party. Party affiliation is determined by a threshold $k \in [\underline{v},\overline{v}]$. In particular, I call Republicans the voters such that $v \geqslant k$ and Democrats the voters such that $v < k$.

Gerrymandering. A Republican designer, or redistricter, is in charge of creating equipopulous districts so as to maximize his party's seat share. She allocates voters among a continuum of districts based on their type v, thus determining the distribution $\pi \in \Delta([\underline{v},\overline{v}])$, with CDF P, of voter types within a district. A redistricting plan $\mathcal{H} \in \Delta(\Delta([\underline{v},\overline{v}]))$[6] is, therefore, a distribution over distributions: it specifies the measure of districts with each distribution π of voter types. To satisfy the equipopulous requirement typical of gerrymandering, any plan must be such that the following budget constraint holds[7]:

$$\int \pi d\mathcal{H}(\pi) = \phi. \tag{BC}$$

For instance, uniform redistricting imposes all districts to be the same, thus $\mathcal{H}(\phi) = 1$, while perfect segregation imposes that each voter type $v \in [\underline{v},\overline{v}]$ constitutes a district on their own.

[4]For any complete and separable metric space X, I let $\Delta(X)$ denote the set of Borel probability measures on X, endowed with the weak* topology.

[5]For instance, a voter type v's preference for policy $x \in [\underline{v},\overline{v}]$ could be represented by utility function $u(x;v) = -|x - v|$.

[6]Note that $\Delta([\underline{v},\overline{v}]])$ is metrizable as a complete and separable metric space with the weak* topology.

[7]The integral sign is used to denote the Lebesgue integral as defined in Aliprantis and Border (1994).

Vote Shares. In any district with distribution $\pi \in \text{supp}(\mathcal{H})$[8], two-stage elections are held. In primary elections, Republican voters choose a Republican candidate (R), while Democratic voters choose a Democratic candidate (D). During general elections, each voter chooses the candidate whose policy position is closer to their ideal point. The designer wins a district if and only if the Republican candidate receives at least a majority of the district vote.

I model electoral incentives so that D and R locate at the lowest median of $\pi(\cdot|v < k)$[9](the median of Democratic affiliates) and at the highest median of $\pi(\cdot|v \geqslant k)$ (the median of Republican affiliates)[10], respectively. There are various reasons why primary candidates might cater to their respective party medians rather than converge to the overall district median. In the Appendix, I provide a detailed model based on Owen and Grofman (2006), where both voters and candidates are uncertain about the position of the population median. However, it is important to note that the strict dependence on party medians is not necessary for my results to go through. For instance, the results are robust to any alternative rule that determines the position of candidate D (respectively, R) through a linear combination of the median of Democratic (respectively, Republican) affiliates and the population median.

To close the model, I assume that whenever $\text{supp}(\pi) \subseteq [k, \infty)$ (respectively, $\text{supp}(\pi) \subseteq (-\infty, k]$), D (respectively, R) takes position k. This assumption is equivalent to stating that even in a district with a very high Democratic (respectively, Republican) majority, there is always an arbitrarily small fraction of moderate Republicans (respectively, Democrats).

Formally, call $c_{\pi,D}$ and $c_{\pi,R}$ the position taken by D and R in district π. Then:

$$(c_{\pi,D}, c_{\pi,R}) = \begin{cases} (v^m_{\pi,D}, k) & if \ \text{supp}(\pi) \subseteq (-\infty, k] \\ (k, v^m_{\pi,R}) & if \ \text{supp}(\pi) \subseteq [k, \infty) \\ (v^m_{\pi,D}, v^m_{\pi,R}) & otherwise \end{cases},$$

[8]Throughout, supp$(\cdot)$ denotes the support of a probability measure, defined as the set of all points whose every open neighborhood has positive measure.

[9]Throughout, $M(Y|X)$ denotes the conditional probability distribution of Y given X according to measure M.

[10]Without loss of generality, the tie breaking rule is chosen so as to insure the existence of an optimum.

where:

$$v^m_{\pi,D} = \inf_a \quad \left\{a:\ \pi([a,\overline{v}]|v<k) \geqslant \frac{1}{2}\right\} \cap \left\{a:\ \pi([\underline{v},a]|v<k) \geqslant \frac{1}{2}\right\}$$
$$v^m_{\pi,R} = \sup_a \quad \left\{a:\ \pi([a,\overline{v}]|v \geqslant k) \geqslant \frac{1}{2}\right\} \cap \left\{a:\ \pi([\underline{v},a]|v \geqslant k) \geqslant \frac{1}{2}\right\}.$$

Voters choose their district representative by majority rule. Hence, the candidate who wins the elections is the one closer to the district's median. Formally, the position of the district representative, c_π, is:

$$c_\pi = \begin{cases} c_{\pi,D} & if\ \ v^m_\pi < \frac{c_{\pi,R}+c_{\pi,D}}{2} \\ c_{\pi,R} & if\ \ v^m_\pi \geqslant \frac{c_{\pi,R}+c_{\pi,D}}{2} \end{cases}$$

where:

$$v^m_\pi = \sup_a \quad \left\{a:\ \pi([a,\overline{v}]) \geqslant \frac{1}{2}\right\} \cap \left\{a:\ \pi([\underline{v},a]) \geqslant \frac{1}{2}\right\}.$$

The designer wins district π if the winning candidate is the Republican candidate; that is, if $c_\pi \geqslant k$. Given redistricting plan $\mathcal{H}$, the designer's vote share is:

$$\int \mathbb{1}\left(c_\pi \geqslant k\right) d\mathcal{H}(\pi) = \int \mathbb{1}\left(v^m_\pi - \frac{c_{\pi,R}+c_{\pi,D}}{2} \geqslant 0\right) d\mathcal{H}(\pi).$$

Aggregate Uncertainty. Suppose that, after the designer commits to a plan, but before candidates choose their positions, an aggregate location shock affects all voters[11]. Formally, each voter experiences a common shock $\omega \in \mathbb{R}$, so that her ideal point becomes $v-\omega$. Assume that ω has distribution $\gamma \in \Delta(\mathbb{R})$, with CDF G, assumed to be Lipschitz continuous[12] and strictly increasing on $[2\underline{v}-2\overline{v}, 2\overline{v}-2\underline{v}]$. Any shock ω induces a new preference distribution in each district. I call π^ω such induced distribution, with CDF $P^\omega = P(v+\omega)$.

[11] A shock to preferences is particularly relevant in my setup, given that redistricting opportunities usually present themselves only every ten years. In any case, my solution encompasses the case in which the shock is arbitrarily small, so as to approximate the solution in the absence of a shock.

[12] For instance, if G is everywhere continuously differentiable, it satisfies the assumptions. I merely require Lipschitz continuity to allow for G to have a set of non-differentiability points of at most Lebesgue measure zero.

Redistricter's Problem. The redistricter, or designer, wishes to maximize the expected seat share won by Republicans. Her problem is:

$$\max_{\mathcal{H}\in\Delta(\Delta([\underline{v},\overline{v}]))} \int\int \mathbb{1}\left(v^m_{\pi^\omega} - \frac{c_{\pi^\omega,D}+c_{\pi^\omega,R}}{2} \geqslant 0\right) d\mathcal{H}(\pi)d\gamma(\omega)$$

$$\text{s.t.} \int \pi d\mathcal{H}(\pi) = \phi. \tag{RP}$$

A redistricting plan $\mathcal{H}$ is *feasible* if it satisfies constraint (BC). It is *optimal* if it is a solution to (RP).

It is instructive to compare (RP) to a redistricting problem with exogenous policies, whose solution can be found with off-the-shelf tools from recent literature. Suppose that candidates' positions are fixed: $c_{\pi^\omega,R} = \overline{v}$, $c_{\pi^\omega,D} = \underline{v}$. The redistricter's problem under exogenous policies is:

$$\max_{\mathcal{H}\in\Delta(\Delta([\underline{v},\overline{v}]))} \int\int \mathbb{1}\left(v^m_{\pi^\omega} \geqslant k^\star\right) d\mathcal{H}(\pi)d\gamma(\omega)$$

$$\text{s.t.} \int P d\mathcal{H}(\pi) = F. \tag{RPEx}$$

where $k^\star = \frac{\underline{v}+\overline{v}}{2}$. The object of interest to the designer is an element of $\Delta(\Delta([\underline{v},\overline{v}]))$, a distribution over distributions. However, in the case of exogenous policies, the objective function only depends on district medians. Hence, the problem can be mapped to a maximization problem over elements of $\Delta([\underline{v},\overline{v}])$, which are much simpler objects to work with. The set of feasible distributions of medians can be characterized using Theorem 2 in Yang and Zentefis (2024). The transformed problem is:

$$\max_{\chi\in\Delta([\underline{v},\overline{v}])} \int\int \mathbb{1}\left(v-\omega \geqslant k^\star\right) d\chi(v)d\gamma(\omega)$$

$$\text{s.t.} \quad \max\{2F(v)-1,0\} \leqslant X(v) \leqslant \min\{2F(v),1\}, \text{ for all } v \in [\underline{v},\overline{v}],$$

where X is the CDF associated with $\chi \in \Delta([\underline{v},\overline{v}])$. Switching the order of integration, it can be rewritten as:

$$\max_{\chi\in\Delta([\underline{v},\overline{v}])} \int G(v-k^\star)d\chi(v)$$

$$\text{s.t.} \quad \max\{2F(v)-1,0\} \leqslant X(v) \leqslant \min\{2F(v),1\}, \text{ for all } v \in [\underline{v},\overline{v}].$$

Since the problem is linear in χ, it suffices to focus attention on the extreme points of the set of feasible distributions of medians. Again, Yang and Zentefis (2024) characterize such a set in Theorem 1. In my case, since G is increasing, the solution is $X^\star = \max\{2F - 1, 0\}$, which dominates all other feasible distributions of medians, in the first-order-stochastic-dominance sense. We can now recover all solutions to (RPEx) as all those redistricting plans that induce $X^\star$. It is easy to see that there are many such plans. The following proposition, already proven with a different argument by Kolotilin and Wolitzky (2024), describes such plans and summarizes the result for exogenous policies.

Proposition 0. *A feasible redistricting plan $\mathcal{H} \in \Delta(\Delta([\underline{v},\overline{v}]))$ is a solution to the redistricting problem with exogenous policies (RPEx) if and only if, for all $\pi \in \operatorname{supp}\mathcal{H}$, except for at most a zero-measure subset, there exists $v_\pi \geqslant 0$ such that $\pi(\{v_\pi\}) = \pi(\{v : v \leqslant 0\}) = \frac{1}{2}$.*

In the case of endogenous policies, it is not possible to reduce the object of interest to a distribution over a uni-dimensional space, because the objective function depends both on district medians and on candidates' positions. Nevertheless, in the next subsection, I show that the problem can still be simplified and solved using a different set of tools. In particular, I show that the object of interest can be reduced to a joint distribution over two uni-dimensional spaces, thus invoking the literature on the so-called optimal transport.

2.2 The Designer's Problem as an Optimal Transport Problem

While the redistricter's problem (RP) appears challenging, it can be simplified and restated as an optimal transport problem. The following result provides necessary conditions for optimality and constitutes the main ingredient for this transformation.

Proposition 1. *A feasible redistricting plan $\mathcal{H} \in \Delta(\Delta([\underline{v},\overline{v}]))$ is optimal only if, for all $\pi \in \operatorname{supp}(\mathcal{H})$, except for at most a zero-measure subset, there exist $v' \geqslant 0$ and $v'' \leqslant 0$ such that $\pi(\{v'\}) = \pi(\{v''\}) = \frac{1}{2}$.*

Proposition 1 constrains the set of feasible plans: optimality requires that the population of any district has at most two types. Moreover, each district in an optimal plan must place half the mass on a voter type above the median of F and the rest on a voter type below the median of F. Effectively, this proposition states that (RP) is a matching problem

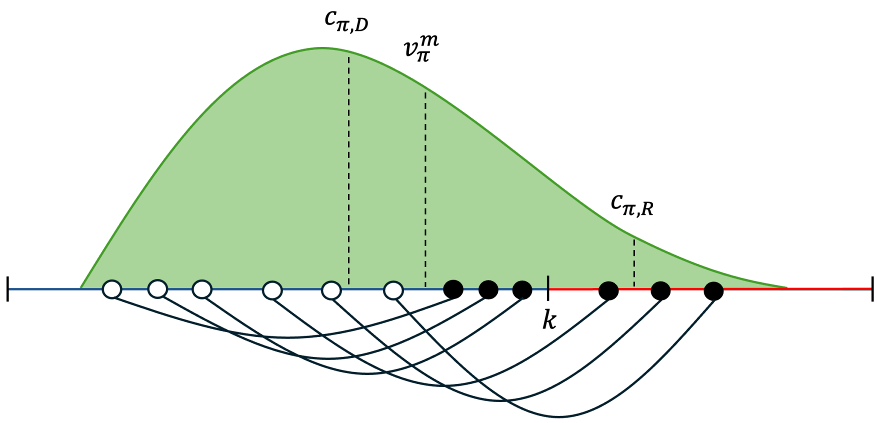
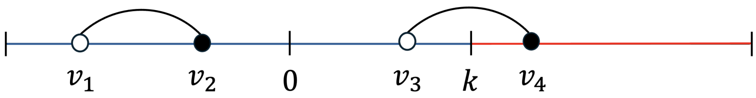


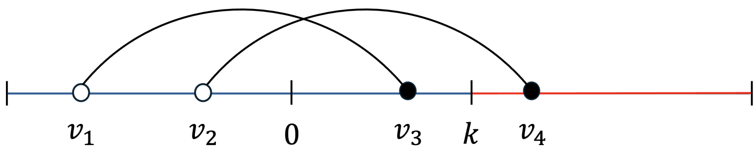
Figure 4: district π (in green) can be replaced by a continuum of binary districts. Each such district matches types above the median of π (●) to types below it (○) in a positive assortative manner.

of voter types across the median of their population distribution. The proof proceeds in two steps, for which I now provide an intuition.

First, it can be shown that any district in an optimal plan must be a *binary district*. That is, it must contain at most two voter types (in equal proportions). Consider a plan $\mathcal{H}$ and any district π in its support. As it turns out, $\mathcal{H}$ can be improved in the following way. Imagine replacing π with multiple smaller districts (possibly infinitely many), where each new district contains at most two voter types in equal proportions: a high type from above the median v_π^m and a low type from below v_π^m. Further, imagine that types above v_π^m and types below v_π^m are allocated to districts in a positive assortative manner. Specifically, suppose that any two districts contain types (h,l) and (h',l'), respectively. If the first district has a higher high type, $h > h'$, then it must also have a higher low type, $l \geqslant l'$. Such construction is illustrated in Figure 4. For illustration purposes, suppose that the shock realization is $\omega = 0$. Then, district π is lost, since v_π^m is closer to the Democratic candidate, $c_{\pi,D}$, than it is to the Republican candidate, $c_{\pi,R}$. Nevertheless, each of the binary districts depicted in Figure 4 is won. Indeed, consider any of the districts $\hat{\pi}$ where the highest type (●) is above k. Then, the Republican candidate $c_{\hat{\pi},R}$ sits exactly at the median $v_{\hat{\pi}}^m$, so that $c_{\hat{\pi},R} = v_{\hat{\pi}}^m =$ ●, and the district is won. Now consider any of the districts $\tilde{\pi}$ where the highest type is below k, so that $\tilde{\pi}$ contains only Democrats. Then, the Republican candidate $c_{\tilde{\pi},R}$ sits at k, the Democratic candidate $c_{\tilde{\pi},D}$ sits at the lowest type (○), and the highest type constitutes the median voter $v_{\tilde{\pi}}^m$. Note that $c_{\tilde{\pi},R}$ is more moderate than $c_{\pi,R}$,



(a) Two districts containing types from the same side of the population median.



(b) Alternative matching.

Figure 5: Districts must match voter types across the population median.

$c_{\tilde{\pi},D}$ is more extreme than $c_{\pi,D}$, and $v^m_{\tilde{\pi}}$ is higher than v^m_{π}, so that each $\tilde{\pi}$ is won.

Second, binary districts must contain exactly two voter types: the highest type drawn from above 0, the population median, and the lowest type drawn from below 0. Figure 5 illustrates this point. Suppose there are two districts, one containing types v_4 and v_3 and the other containing types v_2 and v_1, as shown in Figure 5a. Consider the alternative matching shown in Figure 5b, where v_1 is paired with v_3, while v_2 is paired with v_4. It can be shown that the pairing in 5b is superior to the one in 5a. For illustration purposes, suppose the realization of the shock is $\omega = 0$. Consider the district containing v_4. In both configurations, the Republican candidate sits exactly at v_4 and receives exactly half the votes, enough to win the district. Now consider the district containing v_1. In both cases, the Democratic candidate sits at v_1 and the Republican candidate sits at k. In 5a, v_2 is closer to v_1 than to k, so that all Democrats vote for the Democratic candidate and the district is lost. However, in 5b, type v_3 prefers voting for the Republican candidate, so that the district is won with exactly half the votes.

Proposition 1 justifies the definition of a relation between the set of redistricting plans and the set of joint distributions over $[\underline{v}, 0] \times [0, \overline{v}]$. Define $\phi' = \phi(\cdot | v \geqslant 0)$ and $\phi'' = \phi(\cdot | v < 0)$. In words, ϕ' is the distribution of voter types conditional on them being above the population median, while ϕ'' is the distribution of voter types conditional on them being below the population median. I call $T(\phi', \phi'') \subseteq \Delta([\underline{v}, 0] \times [0, \overline{v}])$ the set of joint distributions having marginals ϕ' and ϕ''.[13] Following the literature on optimal transport,

[13]Formally, $T(\phi', \phi'') = \left\{ \tau \in \Delta([\underline{v}, 0] \times [0, \overline{v}]) : \text{proj}_{[\underline{v},0]} \# \tau = \phi'', \text{proj}_{[0,\overline{v}]} \# \tau = \phi' \right\}$, where: $\text{proj}_{[\underline{v},0]}$ and

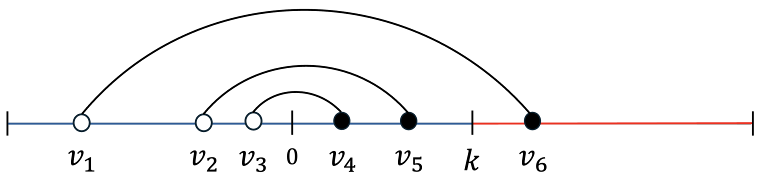


Figure 6: Of the three depicted districts, only the one containing voter types v_3 and v_4 is lost when $\omega = 0$.

I sometimes refer to $T(\phi', \phi'')$ as the set of *transport plans* from ϕ' to ϕ''.

Call $\Delta_2 \subseteq \Delta(\Delta([\underline{v}, \overline{v}]))$ the set of feasible plans satisfying Proposition 1. As it turns out, there is a one-to-one map between Δ_2 and $T(\phi', \phi'')$. Moreover, note that any district placing positive mass on $v' \geqslant 0$ and $v'' \leqslant 0$ can be won in one of two ways. For instance, after the realization of the location shock ω, it can be that $v' - \omega$ is above k, so that the district has at least a majority of Republican affiliates. For such a district, I have $v^m_{\pi^\omega} = c_{\pi^\omega,R} = v' - \omega$ and $c_{\pi^\omega,D} = v'' - \omega$. Alternatively, it can be that $v' - \omega$ is less than k, but still closer to k than it is to $v'' - \omega$, so that the district is split fifty-fifty between "extreme" and "moderate" Democrats, the latter choosing the default moderate Republican candidate. For such a district, I have $v^m_{\pi^\omega} = v' - \omega$, $c_{\pi^\omega,R} = k$, and $c_{\pi^\omega,D} = v'' - \omega$, with $k - (v' - \omega) \leqslant (v' - \omega) - (v'' - \omega)$. In Figure 6, three types of districts are shown. When $\omega = 0$, only the district containing voter types v_3 and v_4 is lost.

Given the above discussion, I can rewrite (RP) as:

$$\max_{\tau \in T(\phi', \phi'')} \int\int \mathbb{1}(v' - \omega \geqslant k) + \mathbb{1}(v' - \omega < k)\mathbb{1}\left(v' - \omega - \frac{v'' - \omega}{2} \geqslant \frac{k}{2}\right) d\tau(v', v'') d\gamma(\omega).$$

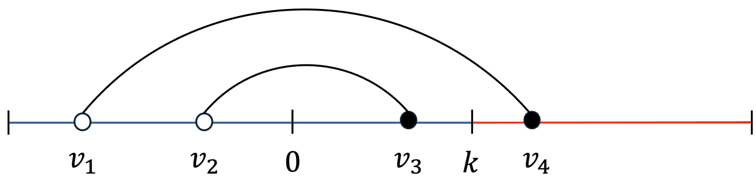
By switching the order of integration and further manipulating the above, I get the optimal transport problem:

$$\max_{\tau \in T(\phi', \phi'')} \int G\left(2v' - v'' - k\right) d\tau(v', v''). \tag{OTP}$$

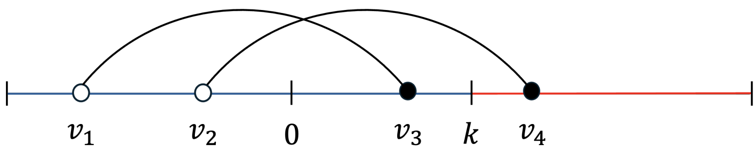
A transport plan τ in $T(\phi', \phi'')$ is *pure* whenever $\{v', v''\} \in \mathrm{supp}(\tau)$ implies $\{v', \tilde{v}''\}, \{\tilde{v}', v''\} \notin \mathrm{supp}(\tau)$ for $v' \neq \tilde{v}'$ and $v'' \neq \tilde{v}''$. Intuitively, purity requires that no "splitting of masses" occurs across voter types. A pure plan is sometimes referred to as a *transport map*.

Define $T^\star \subseteq T(\phi', \phi'')$ as the set of solutions to (OTP) and $\Delta_2^\star \subseteq \Delta_2$ as the set of solutions to (RP). The optimal transport problem (OTP) is *equivalent* to the redistricter's problem

$\mathrm{proj}_{[0,\overline{v}]}$ denote the projection functions of $[\underline{v}, 0] \times [0, \overline{v}]$ on $[\underline{v}, 0]$ and $[0, \overline{v}]$, respectively; $(\mathrm{proj}_{[\underline{v},0]} \# \tau)(A) = \tau(\mathrm{proj}^{-1}_{[\underline{v},0]}(A))$ and $(\mathrm{proj}_{[0,\overline{v}]} \# \tau)(A) = \tau(\mathrm{proj}^{-1}_{[0,\overline{v}]}(A))$ for all A measurable.



(a) The district containing voter types v_1, v_4 is considerably safer than the one containing types v_2, v_3.



(b) The two districts are of similar safety.

Figure 7: Two alternative configurations of districts.

(RP) if there exists a bijection from $T^\star$ to $\Delta_2^\star$ mapping each solution to (OTP) to a solution to (RP). The following theorem summarizes the discussion in this subsection.

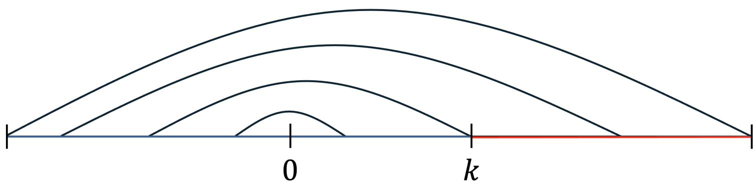
Theorem 1. *The optimal transport problem (OTP) is equivalent to the redistricter's problem (RP).*

3 Characterizing the Optimal Redistricting Plan

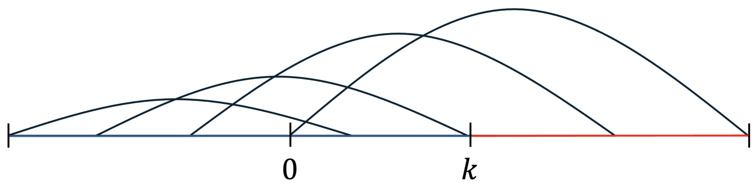
In this section, I characterize the solution(s) to the optimal transport problem (OTP) under different assumptions on the shock distribution G. As it turns out, the optimal redistricting plan depends heavily on the shape of G and, in most cases, on the distribution of voters F.

3.1 Benchmark Cases

I start by describing a few benchmark cases before focusing on the most realistic case of a symmetric, S-shaped shock. Suppose that G is strictly convex on its support. The location shock "shifts" the distribution of preferences. A negative shock shifts it to the right, increasing the fraction of Republican voters (hence it is a favorable shock), while a positive shock shifts the distribution to the left (so it is unfavorable). When G is convex, the marginal benefit of slightly improving a competitive district's safety is outweighed by



(a) When G is convex, the optimal transport plan maps types above 0 to types below 0 in a negative assortative manner.



(b) When G is concave, the optimal transport plan maps types above 0 to types below 0 in a positive assortative manner.

Figure 8: The unique optimal plan under G convex/concave.

the cost of marginally reducing the safety of an already secure district. Consider Figure 7. Panel 7a depicts two districts, the one containing voter types v_1, v_4 considerably safer than the one containing types v_2, v_3. One can think of constructing two alternative districts of intermediate safety, as in panel 7b, by matching v_1 to v_3 and v_2 to v_4. Because G is convex, the designer always prefers the matching in 7*a* to the one in 7*b*. In other words, the designer wants to create very safe districts consisting of extremists of both parties, alongside very unsafe districts consisting of more moderate voters. Based on this intuition, it can be shown that there exists a unique, pure solution to OTP and that such solution maps types above zero to types below zero in a *negative assortative manner*[14]. Figure 8a illustrates such negative assortative map.

Concave Shock. The case of a strictly concave G is analogous to that of a strictly convex G. In this case, the marginal benefit of slightly improving a competitive district's safety exceeds the cost of marginally reducing the safety of an already secure district. Hence, the designer prefers the configuration in Figure 7b to the one in Figure 7a and tries to create districts of similar safety. This preference for more balanced districts creates a natural tendency against negative assortative matching. Indeed, it can be shown that there exists a unique, pure solution to OTP that maps types above 0 to types below 0 in a *positive*

[14]Formally, a *negative assortative map* $\tau \in T(\phi', \phi'')$ is such that, for all $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau)$, $v' > \tilde{v}'$ implies $v'' \leqslant \tilde{v}''$.

assortative manner[15], as illustrated in Figure 8b.

Uniform Shock. For the sake of completeness, I analyze the case when G is affine on its support. It is easy to see that, given the linearity of G, problem OTP does not depend on T, therefore $T^\star = T(\phi', \phi'')$. In other words, any redistricting plan in Δ_2 is optimal.

The following result formalizes the findings for benchmark cases.

Proposition 2. *Consider the following cases:*

- *If G is strictly convex on its support, there exists a unique, pure solution $\tau \in T(\phi', \phi'')$ to OTP and it is a negative assortative map.*
- *If G is strictly concave on its support, there exists a unique, pure solution $\tau \in T(\phi', \phi'')$ to OTP and it is a positive assortative map.*
- *If G is affine on its support, any $\tau \in T(\phi', \phi'')$ is a solution to (OTP).*

In the next subsections, I will make use of the above benchmark results as building blocks to characterize the solutions to (OTP) when the shock is strictly convex below zero and strictly concave above zero, or S-shaped.

3.2 S-shaped and Symmetric Shock

Suppose that G is strictly S-shaped[16] and symmetric around zero. Any normal shock with mean zero falls under this category. For instance, a normal shock with sufficiently small variance is of particular interest because it approximates the solution to (RP) in the absence of an aggregate shock. As it turns out, (OTP) admits a unique solution. Moreover, one can capitalize on the benchmark cases studied in the previous subsection to show that such solution is the convex combination of a positive assortative map and a negative assortative map, each over an appropriate subset of $[0, \overline{v}] \times [\underline{v}, 0]$. Further specifics of G determine which exact subsets host negative or positive assortments, along with additional characteristics of the solution, like purity. Figure 9a depicts $[0, \overline{v}] \times [\underline{v}, 0]$ and partitions it into two subsets. For any (v', v'') belonging to the purple region, $2v' - v'' - k$ is less than

[15]Formally, a *positive assortative map* $\tau \in T(\phi', \phi'')$ is such that, for all $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau), v' > \tilde{v}'$ implies $v'' \geqslant \tilde{v}''$.

[16]Formally, G is strictly convex on $[\underline{v}, 0]$ and strictly concave on $[0, \overline{v}]$.

zero, while for any (v', v'') belonging to the green region, $2v' - v'' - k$ is greater than zero. Hence, G is convex on the purple subset of $[0, \overline{v}] \times [\underline{v}, 0]$ and concave on the green one. Remembering the discussion for benchmark cases, it should not be surprising that any two couples of voter types falling in the green subset of $[0, \overline{v}] \times [\underline{v}, 0]$ must constitute a positive assortment, in order to be part of a solution to (OTP). Similarly, any two couples of voter types falling in the purple subset of $[0, \overline{v}] \times [\underline{v}, 0]$ must constitute a negative assortment. The following result formalizes such intuition.

Proposition 3. *Suppose G is S-shaped and symmetric around 0. There exist τ^+, $\tau^- \in \Delta([0, \overline{v}] \times [\underline{v}, 0])$ and $\alpha \in [0, 1]$ such that:*

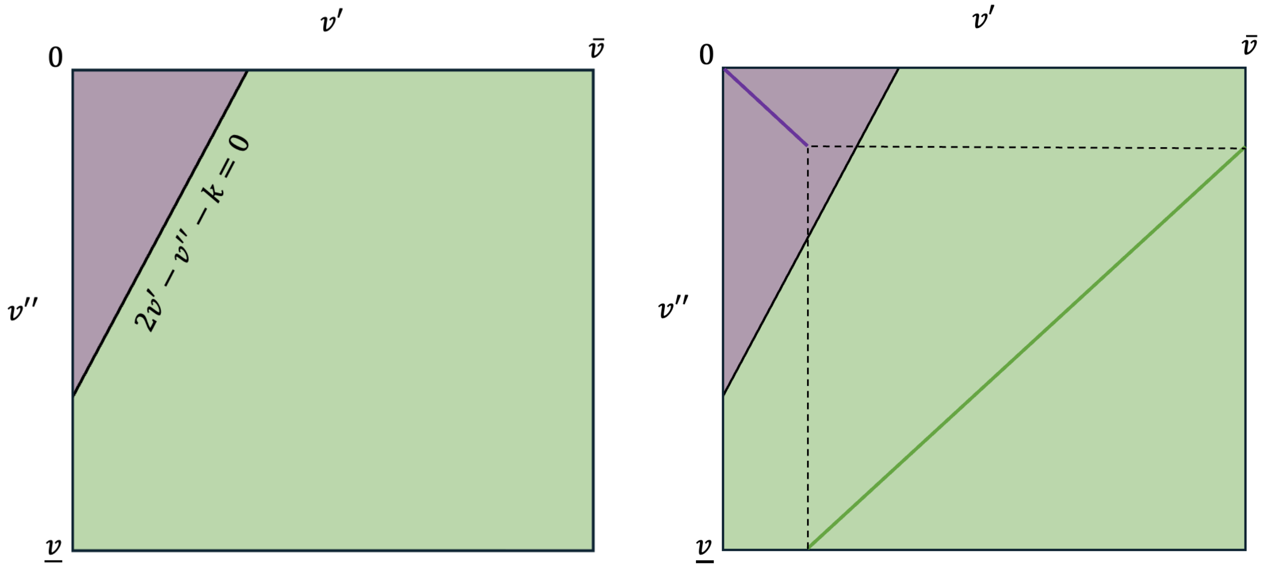
1. *τ^- is a negative assortative map with $\operatorname{supp}(\tau^-) \subseteq \{(v', v'') \in [0, \overline{v}] \times [\underline{v}, 0] : 2v' - v'' - k \geqslant 0\}$*
2. *τ^+ is a positive assortative map with $\operatorname{supp}(\tau^+) \subseteq \{(v', v'') \in [0, \overline{v}] \times [\underline{v}, 0] : 2v' - v'' - k \leqslant 0\}$*
3. *$(1 - \alpha)\tau^- + \alpha\tau^+$ is the unique solution to OTP.*

As an example, Figure 9b depicts the support of a pure candidate for optimality, when ϕ is uniform. In this case, the linearity of F results in a linear mapping of voter types above zero and below zero, with negative slope for matches in the purple region and positive slope for matches in the green region.

While Figure 9b depicts the support of a pure transport plan, the actual solution to (OTP) may very well not be pure, meaning that some voter types might belong to a positive assortative map and a negative assortative map, simultaneously. As I show next, this is indeed the most likely case under a symmetric S-shaped shock, whenever the optimum requires $\alpha < 1$.

Proposition 3 justifies the definition of set $T^{\pm}$ as the set of all plans $\tau \in T(\phi', \phi'')$ for which there exist τ^+, $\tau^- \in \Delta([0, \overline{v}] \times [\underline{v}, 0])$ and $\alpha \in [0, 1]$ such that:

1. τ^- is a negative assortative map with
$$\operatorname{supp}(\tau^-) \subseteq \{(v', v'') \in [0, \overline{v}] \times [\underline{v}, 0] : 2v' - v'' - k \geqslant 0\}$$
2. τ^+ is a positive assortative map with
$$\operatorname{supp}(\tau^+) \subseteq \{(v', v'') \in [0, \overline{v}] \times [\underline{v}, 0] : 2v' - v'' - k \leqslant 0\}$$



(a) The purple subset of $[\underline{v}, 0] \times [0, \overline{v}]$ hosts negative assortments, while the green subset hosts positive assortments.

(b) Example of $\tau = \alpha\tau^- + (1-\alpha)\tau^+$ under ϕ uniform.

Figure 9: The unique solution to (OTP) is the convex combination of a negative assortative map and a positive assortative map.

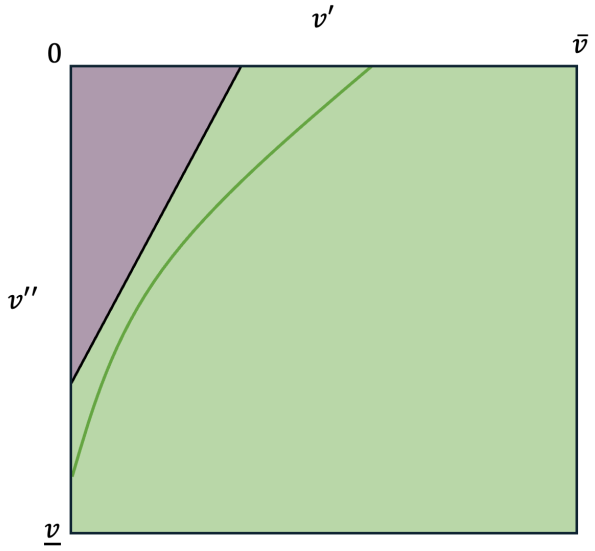
3. $\tau = (1-\alpha)\tau^- + \alpha\tau^+$.

The symmetry assumption allows me to derive a first-order condition that further characterizes the solution within $T^{\pm}$. The following result states that, at the optimum, either τ is fully positive assortative, or the support of its positive assortative part, τ^+, is tangent to the line $2v' - v'' - k = 0$.

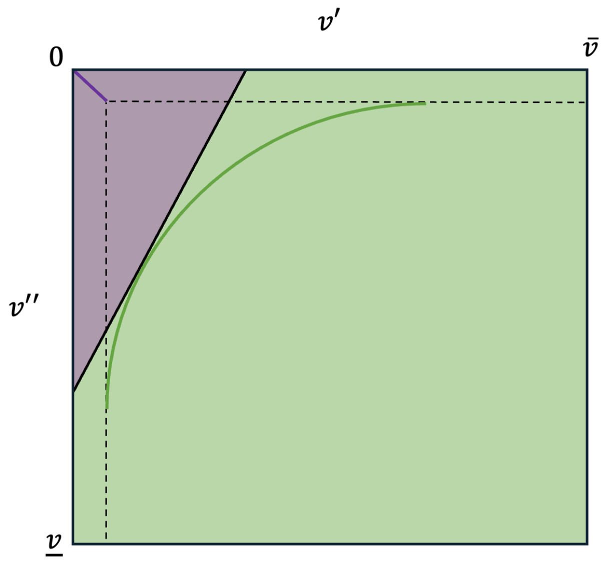
Proposition 4. *Suppose G is symmetric and S-shaped around 0. If $\tau = \alpha\tau^+ + (1-\alpha)\tau^- \in T^{\pm}$ is the solution to (OTP), there exist $v', v'' \in \mathrm{supp}(\tau^+)$ such that:*

$$(1-\alpha)(2v' - v'' - k) = 0.$$

Figure 10 illustrates Proposition 4, while Figure 11 demonstrates the intuition behind its proof. In Figure 11a, we see a plan in $T^{\pm}$. Here, voter types v_1, v_2, v_7, and v_8 are matched positively assortatively, while v_3, v_4, v_5, and v_6 are matched negatively assortatively. Now, consider breaking the match between v_3 and v_6 and including them in the positive assortative match, as shown in Figure 11b. If this new configuration belongs to $T^{\pm}$, it proves superior to the original arrangement. The key insight is that Figure 11b



(a) The solution is fully positive assortative.

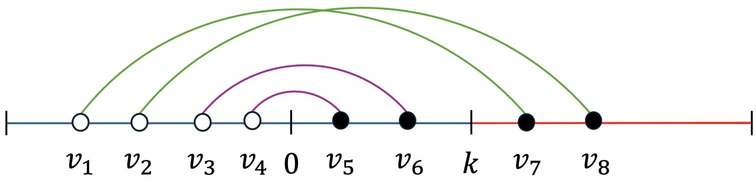


(b) Tangency condition.

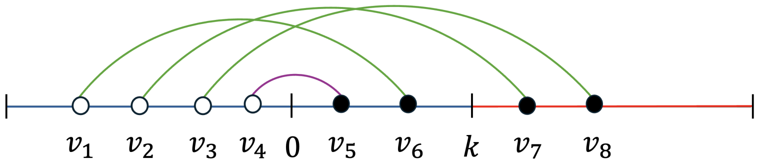
Figure 10: Under a symmetric shock, the solution is either fully positive assortative, or the positive assortative map touches $2v' - v'' - k = 0$.

introduces an additional match where $2v' - v'' - k > 0$. On the other hand, it reduces the magnitude by which the latter inequality is satisfied for the other matches in green. However, because the designer's objective function is linear in τ, the symmetry of the shock distribution makes the problem independent of shock variance. This means the redistricter's optimal strategy is straightforward: create the maximum possible number of districts where $2v' - v'' - k > 0$, regardless of the amount by which this inequality is satisfied.

Figure 12 shows the simulated solution when both F and G are normal distributions, for different values of k. Note that as k increases, and the fraction of Republicans decreases, more and more voters are matched in a negative assortative manner. This comparative static aligns with what Proposition 4 suggests. Specifically, when k is low, the redistricter can create many districts satisfying $2v' - v'' - k > 0$ through positive assortative matching. However, as k increases, this becomes harder to achieve, forcing the redistricter to resort to negative assortative matching to maintain as many winning districts as possible.

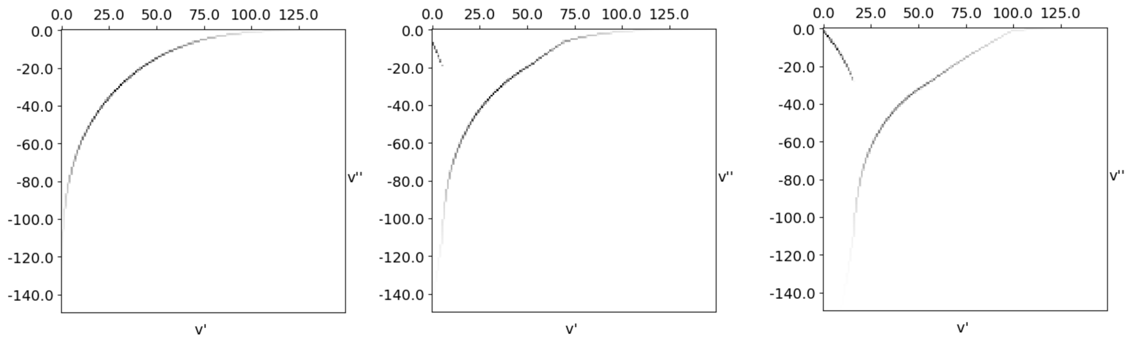


(a) Two positive assortative matched (in green) and two negative ones (in purple).



(b) Three positive assortative matched (in green) and one negative (in purple).

Figure 11: Under a symmetric shock, the configuration in the bottom panel is preferred to the one in the top panel.



(a) $k = 60$ (b) $k = 75$ (c) $k = 100$

Figure 12: Solution for simulated F normal, G normal, for $k = 60,\ 75,\ 100$

4 Implications

Gerrymandering has drawn intense scrutiny for decades, not only for its blatant manipulation of electoral boundaries but also for its far-reaching impacts on democratic representation and political polarization. In this section, I discuss two implications of my model that speak to the current political and legal debate.

4.1 Gerrymandering and Congress Polarization

The relationship between gerrymandering and partisan polarization in the U.S. House of Representatives is often oversimplified in public discourse. While frequently cited as a primary driver of heightened partisanship and legislative gridlock, the actual dynamics are far more nuanced and multifaceted. Empirical research on gerrymandering's impact remains divided, with scholars like McCarty et al. (2009) questioning its significance, while others identify measurable effects (Kenny et al., 2023).

Existing theoretical models, predominantly based on exogenous candidate regimes, have yielded conflicting predictions depending on assumptions about district composition. Models forecasting homogeneous districts, which effectively segregate extreme opponents (Gul and Pesendorfer, 2010; Kolotilin and Wolitzky, 2024), predict a distinct "gap" in the ideological distribution of elected representatives. This gap manifests as a polarized landscape, with moderate supporters of the redistricting party at one end and extreme opponents at the other. In contrast, models predicting district heterogeneity (Friedman and Holden, 2008) anticipate a more continuous spectrum of political representation, without such a pronounced ideological chasm.

By incorporating endogenous candidate emergence, my model predicts the formation of a significant gap in the distribution of district representatives without relying on the creation of homogeneous districts. Let $Q^{\omega}_{\mathcal{H}}$ denote the distribution of district representatives given a redistricting plan $\mathcal{H}$ and a shock realization ω. Formally:

Proposition 5. *For any optimal plan $\mathcal{H}$ and shock realization $\omega < k$, $Q^{\omega}_{\mathcal{H}}((-\omega, k)) = 0$.*

Consistent with existing literature, optimal plans create numerous right-leaning districts that elect moderate Republican candidates. However, they differ from previous models in their treatment of opposition voters. Instead of isolating Democratic voters in homogeneous districts, these plans strategically distribute them across heterogeneous districts, each containing a carefully calibrated mix of moderate and extreme Democratic voters. This nuance yields two critical implications. First, in these heterogeneous "packed" districts, extreme Democratic candidates consistently prevail as the consolidated bloc of extreme Democratic voters outweighs moderate Democrats. Second, these districts effectively disenfranchise moderate Democratic voters, who find themselves without representatives reflecting their political stances.

The analysis in this paper bridges competing theories. It predicts both discontinuities in the distribution of representatives, characteristics of voter-segregating models, and within-district polarization, typically associated with models yielding continuous representative distributions. While existing literature has primarily examined voter segregation as a driver of polarization, this work explores an alternative mechanism: strategic distribution of heterogeneous opposition voters. This approach offers a new perspective on the relationship between redistricting strategies and political polarization, which is worth exploring in future research.

4.2 Legislative implications for "majority-minority" districts

American federal legislation, including the 1965 Voting Rights Act, mandates that electoral district lines cannot be drawn in such a manner as to improperly dilute minorities' voting power. Since the 1986 Supreme Court decision in Thornburg v. Gingles, such laws have been interpreted as actively requiring the creation of districts where racial and ethnic minorities have the concrete opportunity to elect their own representatives. As a consequence, in the 118th Congress there are 26 congressional districts where Black people constitute a strict majority, and 37 are majority Hispanic or Latino (Klein, 2023). Such districts are called "majority-minority."

While such legislation aims to increase minority representation in Congress, it has sparked debate over its broader political implications. Some argue these districts inadvertently segregate Democratic voters, potentially mirroring aspects of an optimal Republican redistricting strategy. The impact of such legislation on partisan outcomes remains complex and unresolved.

As for the case of Congress polarization, the economic theory literature is divided. One school of thought predicts that segregating extreme opponents is optimal, suggesting majority-minority districts benefit Republicans by forcing Democratic redistricters to deviate from their optimal strategy. The opposing view argues for more heterogeneous districts, implying that majority-minority districts disrupt Republicans' desired optimum. In my model, imposing a clear homogeneous majority could prevent both Democrats and Republicans from exploiting the endogeneity of electoral incentives.

In some districts, racial or ethnic minorities may constitute a plurality rather than a majority. These areas, known as "minority opportunity" or "non-majority minority" districts, provide these groups the chance to elect their preferred representatives through coalitions with White voters or other minority groups.

Some political analysts argue that Black candidates can win in constituencies where Black voters comprise less than $40-50\%$ of the population. They contend that creating Black-majority districts is unnecessary and may inadvertently limit potential Black political gains (Canon, 2022).

Recent trends, particularly in the South, show people of color winning more seats in majority-white districts (Lublin et al., 2020). However, the success of this strategy largely depends on White Democrats' willingness to support candidates of color. If they are reluctant to do so, a phenomenon known as "White backlash" may occur, potentially benefiting Republican candidates as described in this paper.

5 Extension: Quantile Redistricting

In the previous sections, I relied on party candidates whose positioning rule is a function of party medians. The implication is that the designer can win a district even if there are no supporters inhabiting it. I now consider a scenario where candidates' positions are determined by a quantile, not necessarily the median, of the distribution of preferences within each party. Let $\frac{1}{2} < q < 1$ be the quantile parameter. Under this setting, the Democratic candidate locates at the q-quantile of the preference distribution conditional on being affiliated with the Democratic party, while the Republican candidate locates at the $(1-q)$-quantile of the preference distribution conditional on being affiliated with the Republican party. In this setting, as it will become clear soon, the designer always needs a non-zero measure of supporters in order to win a district. Therefore, it becomes much harder to characterize the solution in general[17]. Thus, for illustrative purposes, I assume F to be uniform on $[\underline{v}, \overline{v}]$.

[17]For starters, the connection to optimal transport requires considering joint distributions with three fixed marginals, rather than just two. Hence, the optimal transport problem itself becomes less tractable.

Call $c^q_{\pi,D}$ and $c^q_{\pi,R}$ the positions taken by D and R in district π. Then:

$$(c^q_{\pi,D}, c^q_{\pi,R}) = \begin{cases} (v^q_{\pi,D}, k) & if \;\; \text{supp}(\pi) \subseteq (-\infty, k] \\ (k, v^q_{\pi,R}) & if \;\; \text{supp}(\pi) \subseteq [k, \infty) \\ (v^q_{\pi,D}, v^q_{\pi,R}) & otherwise \end{cases}$$

where:

$$v^q_{\pi,D} = \inf_a \quad \{a : \; \pi([a, \overline{v}] | v < k) \geqslant q\} \cap \{a : \; \pi([\underline{v}, a] | v < k) \geqslant q\}$$
$$v^q_{\pi,R} = \sup_a \quad \{a : \; \pi([a, \overline{v}] | v \geqslant k) \geqslant 1 - q\} \cap \{a : \; \pi([\underline{v}, a] | v \geqslant k) \geqslant 1 - q\}.$$

Every other detail of the model is the same as in Section 2. The following proposition characterizes the optimal measure of districts won by the designer for each realization of the shock ω, denoted by $V^\omega(q)$.

Proposition 6. *There are two cases:*

1. *Suppose* $F(k+\omega) \leqslant \frac{1}{2q}$. *Then* $V^\omega(q) = 1$.
2. *Suppose* $F(k+\omega) > \frac{1}{2q}$. *Then* $V^\omega(q) = \frac{2q}{2q-1}(1 - F(k+\omega))$.

In the first case, where the fraction of Democratic voters is less than or equal to $\frac{1}{2q}$, all districts are won by the Republican party. In the second case, where the fraction of Democratic voters is greater than $\frac{1}{2q}$, the designer wins $\frac{2q}{2q-1}(1 - F(k+\omega))$ district, which is strictly more than $2(1 - F(k+\omega))$, the districts she would win under exogenous policies. The intuition for the optimal plan is similar to the one in the previous sections. In each district, the redistricter needs to create a wedge between moderate and extreme Democrats, encouraging the emergence of an extreme Democratic candidate and counting on moderate Democrats to vote for a moderate Republican candidate. The following result characterizes optimal redistricting plans:

Proposition 7. *Define* $v^q = F^{-1}\left(\frac{1}{2q}\right)$. *A feasible redistricting plan* $\mathcal{H}$ *is optimal if and only if, for all* $\pi \in supp\ (\mathcal{H})$, *there exists* $v'_\pi \geqslant v^q$ *such that:*

1. $\pi(\{v'_\pi\}) = 1 - \pi(\{v : v < v^q\}) = \frac{2q}{2q-1}$, *and*
2. $|v^q_{\pi,D} - v^m_\pi| \geqslant |v^m_\pi - v'_\pi|$.

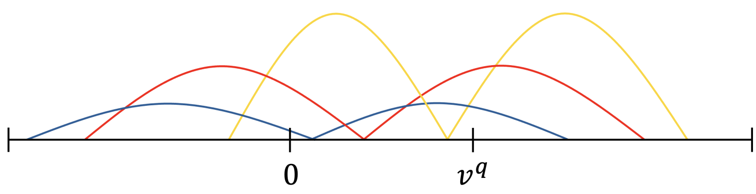


Figure 13: Three-wise positive assortative matching with probability masses of $\frac{1}{2}$, $\frac{1}{2q}$, and $\frac{2q}{2q-1}$.

This proposition states that an optimal redistricting plan under quantile redistricting must satisfy two conditions for each district π in the support of the plan. First, the mass of voters with type exactly equal to v'_π should be $\frac{2q}{2q-1}$, and the mass of voters with type below v_q should be $1 - \frac{2q}{2q-1}$. Second, the distance between $v^q_{P,D}$, which is the lowest median of the district, and the highest median of the district should be greater than or equal to the distance between the highest median and v'_π. As it turns out, a type of redistricting plan called "three-wise positive assortative" plan complies with the requirements of Proposition 7. Figure 13 illustrates such a plan. Each district contains exactly three voter types, one type below 0, one type between 0 and v^q, and another type above v^q, with probability masses of $\frac{1}{2}$, $\frac{1}{2q}$, and $\frac{2q}{2q-1}$, respectively. Given shock ω, the designer wins district π if and only if $v'_\pi - \omega \geqslant k$. Note that the intuition for the distribution of district representatives works exactly as in Proposition 5. Even under quantile redistricting, the winning candidate is either a Republican with position above k, or a Democrat with position below $-\omega$, so that the distribution of district representatives has a gap in $(-\omega, k)$.

6 Conclusion

This paper examines how partisan gerrymandering affects electoral outcomes when candidates' policy positions respond endogenously to district composition. The analysis reveals that optimal gerrymandering strategies differ fundamentally from traditional "pack-and-crack" approaches. While previous literature either predicts extreme opposition voters to be segregated or matched with extreme supporters, my model shows that optimal gerrymandering creates "mismatched slices"—pairing extreme opposition voters with moderate opposition voters. This strategy exploits primary election dynamics to drive opposition candidates toward extreme positions, ultimately allowing the redistricting party to win districts by appealing to moderate voters. These findings suggest that when candidate positions respond to district composition, gerrymandering becomes an even more potent tool than previously understood. By leveraging these electoral dynamics, redistricters can secure victories in more districts than would be possible under traditional approaches that assume fixed voter preferences. This enhanced effectiveness stems from the ability to turn voter diversity into an electoral weakness through strategic district design—a mechanism distinct from the uncertainty-based explanations in existing work.

These gerrymandering strategies have important implications for political representation and polarization. The optimal redistricting plan leads to a notable ideological gap in Congress, with districts electing either moderate Republicans or extreme Democrats. Notably, this polarization emerges even in heterogeneous districts, challenging the conventional wisdom that political extremism primarily results from voter segregation. The findings also have implications for minority representation, particularly in "minority opportunity" districts where success depends critically on coalition-building with white voters. Several important directions for future research emerge from this analysis. First, incorporating idiosyncratic uncertainty about voter preferences could reveal how local shocks affect optimal redistricting strategies and potentially limit gerrymandering's effectiveness. Second, exploring different objective functions for the redistricting party—such as maximizing policy influence rather than just seat share—could yield new insights into the relationship between gerrymandering and legislative outcomes. Finally, future work should focus on developing mechanisms to make redistricting both fair and representative. Such research could help inform practical electoral reforms that preserve democratic representation while limiting partisan manipulation.

Bibliography

Aliprantis, C. D. and K. c. Border (1994): *Infinite dimensional analysis*, Springer.

Besley, T. and I. Preston (2007): "Electoral bias and policy choice: theory and evidence," *The Quarterly Journal of Economics*, 122, 1473–1510.

Bouton, L., G. Genicot, M. Castanheira, and A. L. Stashko (2023): "Pack-Crack-Pack: Gerrymandering with Differential Turnout," Tech. rep., National Bureau of Economic Research.

Canon, D. T. (2022): "Race and redistricting," *Annual Review of Political Science*, 25, 509–528.

Chiappori, P.-A., R. J. McCann, and L. P. Nesheim (2010): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42, 317–354.

Daley, D. (2020): "Inside the Republican Plot for Permanent Minority Rule," *The New Republic*, accessed: September 21, 2024.

Flaccus, G. (2022): "GOP's Chavez-DeRemer flips Oregon 5th Congressional District," *Associated Press*, accessed: September 21, 2024.

Friedman, J. N. and R. T. Holden (2008): "Optimal gerrymandering: sometimes pack, but never crack," *American Economic Review*, 98, 113–44.

Glueck, K. (2022): "A left-leaning challenger takes down a moderate Democratic congressman in Oregon," *The New York Times*, accessed: September 21, 2024.

Gomberg, A., R. Pancs, and T. Sharma (2023): "Electoral maldistricting," *International Economic Review*, 64, 1223–1264.

Gul, F. and W. Pesendorfer (2010): "Strategic redistricting," *American Economic Review*, 100, 1616–41.

Kamenica, E. and M. Gentzkow (2011): "Bayesian persuasion," *American Economic Review*, 101, 2590–2615.

KANTOROVICH, L. V. (1942): "On the translocation of masses," in *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, 199–201.

KENNY, C. T., C. MCCARTAN, T. SIMKO, S. KURIWAKI, AND K. IMAI (2023): "Widespread partisan gerrymandering mostly cancels nationally, but reduces electoral competition," *Proceedings of the National Academy of Sciences*, 120, e2217322120.

KLEIN, M. (2023): "Diving into District Data: What the Latest Census Numbers Tell Us," *The Cook Political Report*, accessed: October 10, 2024.

KOLOTILIN, A., R. CORRAO, AND A. WOLITZKY (2023): "Persuasion with non-linear preferences," .

KOLOTILIN, A. AND A. WOLITZKY (2024): "The economics of partisan gerrymandering," .

LUBLIN, D., L. HANDLEY, T. L. BRUNELL, AND B. GROFMAN (2020): "Minority success in non-majority minority districts: Finding the "sweet spot"," *Journal of Race, Ethnicity, and Politics*, 5, 275–298.

MCCARTY, N., K. T. POOLE, AND H. ROSENTHAL (2009): "Does gerrymandering cause polarization?" *American Journal of Political Science*, 53, 666–680.

MONGE, G. (1781): "Mémoire sur la théorie des déblais et des remblais," *Mem. Math. Phys. Acad. Royale Sci.*, 666–704.

OWEN, G. AND B. GROFMAN (1988): "Optimal partisan gerrymandering," *Political Geography Quarterly*, 7, 5–22.

——— (2006): "Two-stage electoral competition in two-party contests: persistent divergence of party positions," *Social Choice and Welfare*, 26, 547–569.

PUPPE, C. AND A. TASNÁDI (2009): "Optimal redistricting under geographical constraints: Why "pack and crack" does not work," *Economics Letters*, 105, 93–96.

SANTAMBROGIO, F. (2015): "Optimal transport for applied mathematicians," *Birkäuser, NY*, 55, 94.

SCOTT, E. AND D. WEIGEL (2022): "Rep. Kurt Schrader loses primary in Oregon's 5th District," *The Washington Post*, accessed: September 21, 2024.

Shotts, K. W. (2002): "Gerrymandering, legislative composition, and national policy outcomes," *American Journal of Political Science*, 398–414.

Yang, K. H. and A. K. Zentefis (2024): "Monotone Function Intervals: Theory and Applications," *American Economic Review*, 114, 2239–2270.

A A Model of Two-Stage District Elections

In this appendix, I develop a model of probabilistic voting at the district level, providing a foundation for the dependence of district candidates' positions on the conditional medians of their respective vote base. While equilibrium policy divergence can be supported under a variety of models, the dependence of such policies on conditional medians is typical of two-stage election mechanisms. I propose one such model in which, like in many others, two forces work in opposing directions to determine equilibrium policies. One such force, a centripetal force, drives positions towards the district median, fueled by the concern voters have with the ability of their first-stage candidate to win general elections. Another force, a centrifugal force, moves positions away from each other, driven by first-stage voters' uncertainty about the exact position of the district median. Similarly to what was suggested, among others, by Owen and Grofman (2006), equilibrium policies tend to be driven towards party medians.

Consider a continuum of voters in a district. Each voter i has policy preferences given by single-peaked utility $u(\cdot, v_i)$, with $v_i \in [\underline{v}, \overline{v}]$ being the voter's unique ideal point. Further, assume that the distribution π of v_i admits a unique median, conditional median given $v_i \geqslant k$, and conditional median given $v_i < k$, denoted by v^m_π, $v^m_{\pi,R}$, and $v^m_{\pi,D}$.

Two-stage elections are held in the district. In the first, primaries stage, there are two Republican and two Democratic candidates. Voters with $v_i \geqslant k$ elect one of the Republican candidates, while voters with $v_i < k$ elect one of the Democratic candidates, by simple majority, with ties broken uniformly at random. In the second stage, general elections, all voters elect a district representative among the two first-stage winners, by simple majority, with ties broken uniformly at random.

Before the first-stage elections, candidates simultaneously announce a policy position and commit to it. They receive a payoff of 1 if they win the second-stage elections and 0 otherwise.

Suppose that voters and candidates do not know the position of the overall district median v^m_π, and believe it is distributed according to $H \in \Delta([\underline{v}, \overline{v}])$. In addition, candidates know the position of the conditional medians $v^m_{\pi,R}$ and $v^m_{\pi,D}$.

Given voters' single-peaked preferences, the winner of the general elections is the candidate closer to the district median. In the first stage, voters best respond to the anticipated position of the opposing nominee by voting for the candidate maximizing their expected utility. More formally, given the position of the opposing nominee y, a voter with ideal point t votes for the candidate whose position maximizes the following:

$$U(x;y,t) = u(x,t)p(x,y) + u(y,t)(1 - p(x,y)),$$

where $p(x,y)$ is the probability of x winning against y in general elections. As already stated, equilibrium policies in such a model are driven by party medians, but their exact value depends on the functional form of u and H. The literare analyses the model various such specifics. I suggest a *linear setting*, where voter preferences are linear and uncertainty is uniform. I prove the following.

Proposition 8. *Suppose $u(x,v_i) = -|x - v_i|$ and H is uniform on $[\underline{v}, \overline{v}]$. There exists a unique Nash equilibrium where the Republican and Democratic candidates set positions $c_{\pi,R} = v^m_{\pi,R}$ and $c_{\pi,D} = v^m_{\pi,D}$, respectively.*

The advantage of the linear setting is that voters' expected utilities in the primaries turn out to be single-peaked, with maximum reached at each voter's ideal point. Then, it can be argued that any candidate positioning at the party median will win first-stage elections against any other candidate at a different position, irrespective of the opposing party's behavior. Alternatively, Owen and Grofman (2006) consider a model where $u(x,v_i) = e^{-\alpha|x-v_i|}$, for some parameter $\alpha > 0$, and uncertainty is normal around the median, with standard deviation σ. In their model, voters' expected utility during primary elections is not necessarily single-peaked. Hence, they need to explicitly rule out a particular strategic behavior in which some first-stage voters anticipate that they will prefer the opponents' candidate at general elections and purposefully sabotage their own primaries by voting for an extremist[18]. Under this credible assumption, they prove the following:

Proposition 9. (Owen and Grofman, 2006). If $\sigma \geqslant \dfrac{\max\left\{1-e^{1-\alpha(v^m_\pi - v^m_{\pi,D})}, 1-e^{\alpha(v^m_{\pi,R} - v^m_\pi)}\right\}}{\alpha\sqrt{2\pi}}$, there exists a unique Nash equilibrium where the Republican and Democratic candidates set positions $c_{\pi,R} = v^m_{\pi,R}$ and $c_{\pi,D} = v^m_{\pi,D}$, respectively.

[18]Under uniform uncertainty, extreme events have a sufficiently high probability of happening, so that voters are deterred from this kind of "political gamble". Indeed, if they sabotage primaries by voting for an extremist, they run the risk of such candidate winning general elections as well.

B Proofs

B.1 Proofs of Section 2

Proof of Proposition 0. Since the designer wins district $\pi \in \Delta([\underline{v}, \overline{v}])$ whenever $v_\pi^m \geqslant k^\star + \omega$, a redistricting plan $\mathcal{H} \in \Delta(\Delta([\underline{v}, \overline{v}]))$ can be described by a distribution $\chi \in \Delta([\underline{v}, \overline{v}])$ over $v = v_\pi^m$. Using Theorem 2 in Yang and Zentefis (2024), (RPEx) can be stated as:

$$\max_{\chi \in \Delta([\underline{v}, \overline{v}])} \int\int \mathbb{1}\left(v - \omega \geqslant k^\star\right) d\chi(v) d\gamma(\omega)$$

$$s.t. \;\; \max\{2F(v) - 1, 0\} \leqslant X(v) \leqslant \min\{2F(v), 1\}, \text{ for all } v \in [\underline{v}, \overline{v}].$$

Switching the order of integration, it can be rewritten as:

$$\max_{\chi \in \Delta([\underline{v}, \overline{v}])} \int G(v - k^\star) d\chi(v)$$

$$s.t. \;\; \max\{2F(v) - 1, 0\} \leqslant X(v) \leqslant \min\{2F(v), 1\}, \text{ for all } v \in [\underline{v}, \overline{v}].$$

By definition of first order stochastic dominance, and since G is strictly increasing, I have:

$$\int G(v - k^\star) d\chi(v) < \int G(v - k^\star) d\overline{\chi}$$

for any $\chi > \overline{\chi}$, where $\max\{2F(v) - 1, 0\}$ is the CDF associated with $\overline{\chi}$. Hence, the optimal distribution of medians is $X^\star = \max\{2F - 1, 0\}$.

■

Proof of Proposition 1. Consider the following definition:

Definition 1. *A redistricting plan $\mathcal{H}$ is pairwise if $|\operatorname{supp} \pi| \leqslant 2$ for all $\pi \in \operatorname{supp}(\mathcal{H})$.*

The proof of this proposition relies on the following lemma, which states that for any redistricting plan, there exists a pairwise plan that achieves the same value, for each realization of the shock $\omega \in \operatorname{supp}(\gamma)$.

Lemma 1. *For any feasible plan $\mathcal{H}$, there exists a feasible pairwise plan $\hat{\mathcal{H}}$ such that, for all $\omega \in \operatorname{supp}(\gamma)$:*

$$\int \mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) d\hat{\mathcal{H}}(\hat{\pi}) \geqslant \int \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right) d\mathcal{H}(\pi)$$

Proof. Take any plan $\mathcal{H} \in \Delta(\Delta([\underline{v}, \overline{v}]]))$ such that $\int \pi d\mathcal{H}(\pi) = \phi$. First, for any $\pi \in \text{supp}(\mathcal{H})$, construct a measure $\hat{\mathcal{P}}_\pi \in \Delta(\Delta([\underline{v}, \overline{v}]))$ such that $\int \hat{\pi} d\hat{\mathcal{P}}_\pi(\hat{\pi}) = \pi$, and, for all $\hat{\pi} \in \text{supp}(\hat{\mathcal{P}}_\pi)$, $\text{supp}(\hat{\pi}) = \{v'_{\hat{\pi}}, v''_{\hat{\pi}}\}$, for some $v'_{\hat{\pi}}, v''_{\hat{\pi}} \in \text{supp}(\pi)$ with:

$$v'_{\hat{\pi}} \geqslant v^m_\pi,$$
$$v''_{\hat{\pi}} \leqslant v^m_\pi,$$
$$\hat{\pi}(\{v'_{\hat{\pi}}\}) = \hat{\pi}(\{v''_{\hat{\pi}}\}).$$

Moreover, for any $\hat{\pi}$, $\hat{\rho} \in \text{supp}(\hat{\mathcal{P}}_\pi)$, let $v'_{\hat{\pi}} > v'_{\hat{\rho}} \implies v''_{\hat{\pi}} \geqslant v''_{\hat{\rho}}$.

Second, construct alternative plan $\hat{\mathcal{H}} \in \Delta(\Delta([\underline{v}, \overline{v}]))$ such that, for any measurable set $A \subseteq \Delta([\underline{v}, \overline{v}])$, $\hat{\mathcal{H}}(A) = \int \hat{\mathcal{P}}_\pi(A) d\mathcal{H}(\pi)$. By construction, $\hat{\mathcal{H}}$ is feasible and pairwise. I now show that:

$$\int \mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) d\hat{\mathcal{H}}(\hat{\pi}) \geqslant \int \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right) d\mathcal{H}(\pi).$$

Specifically, I show that, for any $\pi \in \text{supp}(\mathcal{H})$, $\hat{\pi} \in \text{supp}(\hat{\mathcal{P}}_\pi)$, $\omega \in \text{supp}(\gamma)$, I have that $\mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) \geqslant \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right)$.

Consider the following three cases:

1. If $v^m_{\pi^\omega} \geqslant k$, then $c_{\hat{\pi}^\omega, R} = v^m_{\hat{\pi}^\omega} \geqslant k$, which means that $\mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) = 1 \geqslant \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right)$.

2. If $v^m_{\pi^\omega} < k$ and $\text{supp}(\hat{\pi}^\omega) = \{v^m_{\pi^\omega}\}$, it must be that, for all $\hat{\rho} \in \text{supp}(\hat{\mathcal{P}}_\pi)$:

$$v''_{\hat{\rho}} < v''_{\hat{\pi}} = v^m_\pi \implies v'_{\hat{\rho}} \leqslant v'_{\hat{\pi}} = v^m_\pi,$$

   which means that $v''_{\hat{\rho}} < v^m_\pi \implies v'_{\hat{\rho}} = v^m_\pi$. Then, it must be that $v^m_{\pi^\omega} = c_{\pi^\omega, D}$ and thus $\mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) = \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right) = 0$.

3. If $v^m_{\pi^\omega} < k$, $|\text{supp}(\hat{\pi}^\omega)| = 2$, and $v^m_{\hat{\pi}^\omega} = c_{\hat{\pi}^\omega, R} \geqslant k$, then $\mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) = 1 \geqslant \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right)$.

4. If $v^m_{\pi^\omega} < k$, $|\text{supp}(\hat{\pi}^\omega)| = 2$, and $v^m_{\hat{\pi}^\omega} < k$, it must be that $c_{\hat{\pi}^\omega, D} \leqslant c_{\pi^\omega, D}$, $c_{\hat{\pi}^\omega, R} = k \leqslant c_{\pi^\omega, R}$, and $v^m_{\hat{\pi}^\omega} \geqslant v^m_{\pi^\omega}$, so that $\mathbb{1}\left(c_{\hat{\pi}^\omega} \geqslant k\right) \geqslant \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right)$.

□

Consider plan $\mathcal{H} \in \Delta(\Delta([\underline{v}, \overline{v}]))$ such that $\int d\pi\mathcal{H}(\pi) = \phi$. The proof proceeds in two steps.

1. First, I show that if $\mathcal{H}$ is optimal it must be that, for all $\pi \in \text{supp}(\mathcal{H})$ (except at most for a zero-measure subset), either $|\text{supp}(\pi)| = 1$, or there exist $v' \neq v''$ such that $\pi(\{v'\}) = \pi(\{v''\}) = \frac{1}{2}$.

   Consider the pairwise feasible plan $\hat{\mathcal{H}}$ constructed in Lemma 1. By construction, $\hat{\mathcal{H}}$ is such that for all $\hat{\pi} \in \text{supp}(\hat{\mathcal{H}})$, either $|\text{supp}(\hat{\pi})| = 1$, or there exist $v' \neq v''$ such that $\hat{\pi}(\{v'\}) = \hat{\pi}(\{v''\}) = \frac{1}{2}$. Morevoer, $\hat{\mathcal{H}}$ is such that $\int \mathbb{1}(c_{\hat{\pi}^\omega} \geqslant 0)\, d\hat{\mathcal{H}}(\hat{\pi}) \geqslant \int \mathbb{1}(c_{\hat{\pi}^\omega} \geqslant 0)\, d\mathcal{H}(\hat{\pi})$ for all $\omega \in \text{supp}(\gamma)$.

   Suppose that, for a positive-measure subset $S \subseteq \text{supp}(\mathcal{H})$, either $|\text{supp}(\pi)| > 2$ or $\text{supp}(\pi) = \{v', v''\}$ with $v' \neq v''$ and $\pi(\{v'\}) \neq \pi(\{v''\})$. I now show that there exists measurable $\Omega \subseteq \text{supp}(\gamma)$ such that $\gamma(\Omega) > 0$ and:

$$\int \mathbb{1}(c_{\hat{\pi}^\omega} \geqslant k)\, d\hat{\mathcal{H}}(\hat{\pi}) > \int \mathbb{1}(c_{\pi^\omega} \geqslant k)\, d\mathcal{H}(\pi),$$

   for all $\omega \in \Omega$.

   Take $\pi \in S$ and suppose $\text{supp}(\pi) \subseteq (-\infty, 0]$. Take the measure $\hat{\mathcal{P}}_\pi$ constructed in the proof of Lemma 1. Excluding the cases $|\text{supp}(\pi)| = 1$ and $|\text{supp}(\pi)| = \{v', v''\}$ with $v' \neq v''$, $\pi(\{v'\}) = \pi(\{v''\})$, it must be that $\hat{\mathcal{P}}_\pi\left(\hat{\pi} : v^m_{\hat{\pi}} - c_{\hat{\pi},D} > v^m_\pi - c_{\pi,D}\right) > 0$. For any $\hat{\pi}$ such that $v^m_{\hat{\pi}} - c_{\hat{\pi},D} > v^m_\pi - c_{\pi,D}$, there exist $\underline{\omega} < \overline{\omega}$ such that $2v^m_{\hat{\pi}^\omega} - c_{\hat{\pi}^\omega,D} > 0 > 2v^m_{\pi^\omega} - c_{\pi^\omega,D}$ for all $\omega \in (\underline{\omega}, \overline{\omega})$. Indeed, $2v^m_{\hat{\pi}^\omega} - c_{\hat{\pi}^\omega,D} = 2v^m_{\hat{\pi}} - c_{\hat{\pi},D} - \omega > 0$ for all $\omega \in (v^m_{\hat{\pi}}, 2v^m_{\hat{\pi}} - c_{\hat{\pi},D})$ and $2v^m_{\pi^\omega} - c_{\pi^\omega,D} = 2v^m_\pi - c_{\pi,D} - \omega < 0$ for all $\omega \in (2v_\pi - c_{\pi,D}, +\infty)$. Since $v^m_{\hat{\pi}} - c_{\hat{\pi},D} > v^m_\pi - c_{\pi,D}$, it suffices to take $\underline{\omega} = 2v^m_\pi - c_{\pi,D}$ and $\overline{\omega} = 2v^m_{\hat{\pi}} - c_{\hat{\pi},D}$. Now, suppose $\text{supp}(\pi) \nsubseteq (-\infty, 0]$. There exists $\omega^\star$ such that, for all $\omega \geqslant \omega^\star$, $\text{supp}(\pi^\omega) \subseteq (-\infty, 0]$. Then, the reasoning for $\text{supp}(\pi) \subseteq (-\infty, 0]$ applies.

2. I showed that if $\mathcal{H}$ is optimal it must be that, for all $\pi \in \text{supp}(\mathcal{H})$, either $|\text{supp}(\pi)| = 1$, or there exist v', v'' such that $\pi(\{v'\}) = \pi(\{v''\}) = \frac{1}{2}$. Now I show that it must be that $v' \geqslant v^m$ and $v'' \leqslant v^m$ for all $\pi \in \text{supp}(\mathcal{H})$ (except for at most a zero-measure subset). Suppose there exist $\overline{S} \subseteq \text{supp}(\mathcal{H})$ such that $\mathcal{H}(\overline{S}) > 0$ and for all $\overline{\pi} \in \overline{S}$, $\text{supp}(\overline{\pi}) = \{\overline{v}'_{\overline{\pi}}, \overline{v}''_{\overline{\pi}}\}$ with $\overline{v}'_{\overline{\pi}} \geqslant \overline{v}''_{\overline{\pi}} \geqslant v^m$. Since $\int \pi d\mathcal{H}(\pi) = \phi$, there must exist $\underline{S} \subseteq \text{supp}(\mathcal{H})$ such that $\mathcal{H}(\underline{S}) = \mathcal{H}(\overline{S}) > 0$ and for all $\underline{\pi} \in \underline{S}$, $\text{supp}(\underline{\pi}) = \{\underline{v}'_{\underline{\pi}}, \underline{v}''_{\underline{\pi}}\}$ with $\underline{v}''_{\underline{\pi}} \leqslant \underline{v}'_{\underline{\pi}} \leqslant v^m$. Consider the measurable set $\hat{S} \subseteq \Delta([\underline{v}, \overline{v}])$ and suppose that, for all $\hat{\pi}' \in \hat{S}$, there exist $\hat{\pi}'' \in \hat{S}$, $\overline{\pi} \in \overline{S}$, and $\underline{\pi} \in \underline{S}$ such that $\hat{\pi}'(\overline{v}'_{\overline{\pi}}) = \hat{\pi}'(\underline{v}'_{\underline{\pi}}) = \frac{1}{2}$, $\hat{\pi}''(\overline{v}''_{\overline{\pi}}) = \hat{\pi}''(\underline{v}''_{\underline{\pi}}) = \frac{1}{2}$. Consider the alternative plan $\hat{\mathcal{H}}$, identical to $\mathcal{H}$ but such that $\hat{\mathcal{H}}(\hat{S}) = \mathcal{H}(\overline{S}) + \mathcal{H}(\underline{S})$ and $\hat{\mathcal{H}}(\overline{S}) = \hat{\mathcal{H}}(\underline{S}) = 0$. Similarly to the first part of this proof, there

exist $\Omega \subseteq \text{supp}(\gamma)$ such that $G(\Omega) > 0$ and:

$$\int \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right) d\hat{\mathcal{H}}(\pi) > \int \mathbb{1}\left(c_{\pi^\omega} \geqslant k\right) d\mathcal{H}(\pi),$$

for all $\omega \in \Omega$.

■

Proof of Theorem 1. The proof of this result rests on the following lemma.

Lemma 2. *There exists a bijection $t : T(\phi', \phi'') \to \Delta_2$.*

Proof. Call $\Delta_1 \subseteq \Delta([\underline{v}, \overline{v}])$ the set of district distributions $\pi \in \Delta([\underline{v}, \overline{v}])$ such that there exist $v'_\pi \geqslant 0$, $v''_\pi \leqslant 0$ with $\pi(\{v'_\pi\}) = \pi(\{v''_\pi\}) = \frac{1}{2}$. Define the function $s : \Delta_1 \to [\underline{v}, 0] \times [0, \overline{v}]$ such that, for each $\pi \in \Delta_1$, $s(\pi) = (v'_\pi, v''_\pi)$. Note that s is a bijection.

Define the function $t : T(\phi', \phi'') \to \Delta(\Delta_1)$ such that, for all $\tau \in T(\phi', \phi'')$, for all measurable $B \subseteq \Delta_1$, $t(\tau)(B) = \tau(s(B))$. It is easy to see that t is well defined.

First, I show that $t(T(\phi', \phi'')) = \Delta_2$. Indeed, take any $\tau \in T(\phi', \phi'')$. I need to show that $t(\tau)$ is feasible. That is, I show that, for all measurable $A \subseteq [\underline{v}, \overline{v}]$:

$$\int_{\Delta_1} \pi(A) dt(\tau)(\pi) = \phi(A).$$

Note that $\int_{\Delta_1} \pi(A) dt(\tau)(\pi) = \frac{1}{2} \int_{\Delta_1} \pi(A|v \leqslant 0) dt(\tau)(\pi) + \frac{1}{2} \int_{\Delta_1} \pi(A|v \geqslant 0) dt(\tau)(\pi)$. Hence, I show that $\int_{\Delta_1} \pi(A|v \geqslant 0) dt(\tau)(\pi) = \phi'(A)$. The reasoning for $\int_{\Delta_1} \pi(A|v \leqslant 0) dt(\tau)(\pi) = \phi''(A)$ is analogous. By definition of t I have:

$$\int_{\Delta_1} \pi(A|v \leqslant 0) dt(\tau)(\pi) = \int_{\Delta_1} \pi(A|v \leqslant 0) d\tau(s(\pi)).$$

With a change of variable I get:

$$\int_{\Delta_1} \pi(A|v \leqslant 0) d\tau(s(\pi)) = \int_{[0,\overline{v}] \times [\underline{v}, 0]} \mathbb{1}(v' \in A) d\tau(v', v'').$$

Finally:

$$\int_{[0,\overline{v}] \times [\underline{v}, 0]} \mathbb{1}(v' \in A) d\tau(v', v'') = \left(\text{proj}_{[0,\overline{v}]} \# \tau\right)(A) = \phi'(A),$$

where the last equality follows from the definition of transport plan.

Second, I show that $\pi \neq \hat{\pi}$ implies $t(\pi) \neq t(\hat{\pi})$. Since $\pi \neq \hat{\pi}$, there exists measurable $B \subseteq [\underline{v}, 0] \times [0, \overline{v}]$ such that $\pi(B) \neq \hat{\pi}(B)$. Then:

$$t(\pi)(B) = \pi(s(B)) \neq \hat{\pi}(s(B)) = t(\hat{\pi})(B)$$

where the second inequality follows from the fact that s is a bijection. □

I now show that $\tau^\star \in T^\star$ if and only if $\mathcal{H}^\star = t(\tau^\star) \in \Delta_2^\star$.

For any $\mathcal{H} \in \Delta_2$, I have:

$$\int\int \mathbb{1}(c_{\pi^\omega} \geqslant k) d\mathcal{H}(\pi) d\gamma(\omega)$$

$$=$$

$$\int\int \mathbb{1}(v'_\pi - \omega \geqslant k) + \mathbb{1}(v'_\pi - \omega < k)\mathbb{1}\left(v'_\pi - \omega - \frac{v''_\pi - \omega + k}{2} \geqslant k\right) d\mathcal{H}(\pi) d\gamma(\omega).$$

By switching the order of integration and further manipulating, I get:

$$\int\int \mathbb{1}(v'_\pi - \omega \geqslant k) + \mathbb{1}(v'_\pi - \omega < k)\mathbb{1}\left(v'_\pi - \omega - \frac{v''_\pi - \omega + k}{2} \geqslant k\right) d\gamma(\omega) d\mathcal{H}(\pi)$$

$$=$$

$$\int\int \mathbb{1}(v'_\pi - \omega \geqslant k) d\gamma(\omega) + \int \mathbb{1}(v'_\pi - \omega < k)\mathbb{1}\left(v'_\pi - \omega - \frac{v''_\pi - \omega + k}{2} \geqslant k\right) d\gamma(\omega) d\mathcal{H}(\pi)$$

$$=$$

$$\int G(v'_\pi - k) + G\left(2v'_\pi - v''_\pi - k\right) - G(v'_\pi - k) d\mathcal{H}(\pi)$$

$$=$$

$$\int G\left(2v'_\pi - v''_\pi - k\right) d\mathcal{H}(\pi).$$

Using Proposition 1, $\mathcal{H}^\star \in \Delta_2$ is optimal if and only if:

$$\int G\left(2v'_\pi - v''_\pi - k\right) d\mathcal{H}(\pi) \leqslant \int G\left(2v'_\pi - v''_\pi - k\right) d\mathcal{H}^\star(P) \ \text{ for all} \mathcal{H} \in \Delta_2.$$

With a change of variable:

$$\int G\left(2v'_\pi - v''_\pi - k\right) dt(\tau)(P) = \int G\left(2v'_\pi - v''_\pi - k\right) d\tau(s(P)) = \int G\left(2v'_\pi - v''_\pi - k\right) d\tau(v'_\pi, v''_\pi)$$

$$\leqslant$$

$$\int G\left(2v'_\pi - v''_\pi - k\right) dt(\tau^\star)(\pi) = \int G\left(2v'_\pi - v''_\pi - k\right) d\tau^\star(s(\pi)) = \int G\left(2v'_\pi - v''_\pi - k\right) d\tau^\star(v'_\pi, v''_\pi),$$

for all $\tau \in T(\phi', \phi'')$ and for $\tau^\star = t^{-1}(\mathcal{H}^\star)$. Hence, $\mathcal{H}^\star$ is optimal if and only if $\tau^\star$ is optimal.

■

B.2 Proofs of Section 3

In this subsection, I borrow results in Santambrogio (2015) and Chiappori et al. (2010) as building blocks to characterize the solution to (OTP). In particular, Lemma 3, Lemma

4, and Lemma 5 are adapted for my context from the above references. Given a function $f : X \to \mathbb{R}$ locally Lipschitz, I define its *superderdifferential* at $x_0 \in X$, $\partial f(x_0)$, to consist of the set of real numbers β such that:

$$f(x) \leqslant f(x_0) + \beta(x - x_0) + o(|x - x_0|) \text{ as } x \to x_0,$$

with the error term being allowed to depend on the x_0. Note that if the function is differentiable at x_0, I have that $\partial f(x_0) = \{f'(x_0)\}$. I provide the following definitions.

Definition 2. *G satisfies the twist condition whenever G is locally Lipschitz and $\partial G(a) \cap \partial G(b) = \varnothing$ for all $a \neq b$.*

Definition 3. *G satisfies the sub-twist condition whenever G is locally Lipschitz and, for all $a \in \mathrm{supp}(\gamma)$:*

$$|\{b \in \mathrm{supp}(\gamma) : b \neq a,\ \partial G(a) \cap \partial G(b) \neq \varnothing\}| \leqslant 1.$$

Definition 4. *For any $\tau \in T(\phi', \phi'')$, $\mathrm{supp}(\tau) \subseteq [0, \overline{v}] \times [\underline{v}, 0]$ is cyclically monotone (CM) if, for every $n \in \mathbb{N}$, every permutation σ, and every finite family of points $(v_1', v_1''), \ldots, (v_n', v_n'') \in \mathrm{supp}(\tau)$:*

$$\sum_{i=1}^{k} G(2v_i' - v_i'' - k) \geqslant \sum_{i=1}^{k} G(2v_i' - v_{\sigma(i)}'' - k).$$

Then, the following lemmas hold.

Lemma 3. *If G satisfies the twist condition, there exists a unique, pure solution to (OTP).*

Proof. By Theorem 2 in Chiappori et al. (2010). □

Lemma 4. *If G satisfies the sub-twist condition, there exists a unique solution to (OTP).*

Proof. By Theorem 3 in Chiappori et al. (2010). □

Lemma 5. *If $\tau \in T(\phi', \phi'')$ is a solution to (OTP), then $\mathrm{supp}(\tau)$ is CM.*

Proof. By assumption, G is continuous. Hence, the result holds by Theorem 1.38 in Santambrogio (2015). □

I am now ready to prove the results in Section 3.

Proof of Proposition 2. Suppose G is strictly convex. Then, for all x, either $\partial G(x) = \emptyset$, or $\inf\ \partial G(x') > \sup\ \partial G(x)$ for all $x' > x$. Then, G satisfies the twist condition, and, by Lemma 3, there exists a unique, pure solution $\tau^\star$ to (OTP). Take any $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau^\star)$, such that $v' > \tilde{v}'$. By Lemma 5, it must be that:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) \geqslant G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Since G is strictly convex, the function $\Phi(v', v'') = G(2v' - v'' - k)$ is strictly submodular. Suppose that $v'' > \tilde{v}''$. By submodularity:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) < G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Hence, it must be $v'' \leqslant \tilde{v}''$.

Suppose G is strictly concave. Then, for all x, $\sup\ \partial G(x) < \inf\ \partial G(x')$ for all $x' > x$. Then, G satisfies the twist condition, and, by Lemma 3, there exists a unique, pure solution $\tau^\star$ to (OTP). Take any $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau^\star)$, such that $v' > \tilde{v}'$. By Lemma 5, it must be that:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) \geqslant G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Since G is strictly concave, the function $\Phi(v', v'') = G(2v' - v'' - k)$ is strictly supermodular. Suppose that $v'' < \tilde{v}''$. By supermodularity:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) < G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Hence, it must be $v'' \geqslant \tilde{v}''$.

Now, suppose G is affine. Then, for any $\tau \in T(\phi', \phi'')$:

$$\int G(2v' - v'' - k) d\tau(v', v'') = G\left(\int (2v' - v'' - k) d\tau(v', v'')\right) =$$
$$= G\left(\int 2v' d\phi' - \int v'' d\phi'' - k\right),$$

by definition of $T(\phi', \phi'')$. Since the objective function is constant over $T(\phi', \phi'')$, I have that $T^\star = T(\phi', \phi'')$.

■

Proof of Proposition 3. First, since G is strictly convex below 0 and strictly concave above 0, it satisfies the sub-twist condition, even if it does not necessarily satisfy the twist condition. Hence, by Lemma 4, a solution τ to (OTP) exists and is unique.

Take any $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau)$, such that $2v' - v'' - k < 0$ and $2\tilde{v}' - \tilde{v}'' - k < 0$. Suppose that $v' > \tilde{v}'$. By Lemma 5, it must be that:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) \geqslant G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Suppose that $v'' > \tilde{v}''$. Since G is strictly convex below 0 and symmetric around 0, I have:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) < G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Hence, it must be $v'' \leqslant \tilde{v}''$.

Take any $(v', v''), (\tilde{v}', \tilde{v}'') \in \text{supp}(\tau)$, such that $2v' - v'' - k \geqslant 0$ and $2\tilde{v}' - \tilde{v}'' - k \geqslant 0$. Suppose that $v' > \tilde{v}'$. By Lemma 5, it must be that:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) \geqslant G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Suppose that $v'' < \tilde{v}''$. Since G is strictly concave above 0, I have:

$$G(2v' - v'' - k) + G(2\tilde{v}' - \tilde{v}'' - k) < G(2v' - \tilde{v}'' - k) + G(2\tilde{v}' - v'' - k).$$

Hence, it must be $v'' \geqslant \tilde{v}''$.

Suppose $\tau(\{v', v'' :\ 2v' - v'' - k \geqslant 0\}) > 0$ and $\tau(\{v', v'' :\ 2v' - v'' - k < 0\}) > 0$. Define the measure τ^+ to be $\tau^+(\cdot) = \tau(\cdot | \{v', v'' :\ 2v' - v'' - k \geqslant 0\})$, and the measure τ^- to be $\tau^-(\cdot) = \tau(\cdot | \{v', v'' :\ 2v' - v'' - k < 0\})$. Note that τ^+ is positive assortative, while τ^- is negative assortative. Moreover:

$$\tau(\cdot) = \tau\left(\{v', v'' :\ 2v' - v'' - k \geqslant 0\}\right) \tau^+(\cdot) + \tau\left(\{v', v'' :\ 2v' - v'' - k < 0\}\right) \tau^-(\cdot).$$

Suppose $\tau(\{v', v'' :\ 2v' - v'' - k \geqslant 0\}) = 0$. Then define $\tau^-(\cdot) = \tau(\cdot)$ and note that it is negative assortative.

Suppose $\tau(\{v', v'' :\ 2v' - v'' - k < 0\}) = 0$. Then define $\tau^+(\cdot) = \tau(\cdot)$ and note that it is positive assortative.

■

Proof of Proposition 4. Consider plan $\tau = \alpha\tau^+ + (1-\alpha)\tau^-$, in $T^{\pm}$. Define $\Gamma^+ = \text{supp}(\tau^+)$, $\Gamma^- = \text{supp}(\tau^-)$, $\mu^+ = \text{proj}_{[0,\bar{v}]}\#\tau^+$, $\mu^- = \text{proj}_{[0,\bar{v}]}\#\tau^-$, $\nu^+ = \text{proj}_{[\underline{v},0]}\#\tau^+$, $\nu^- = \text{proj}_{[\underline{v},0]}\#\tau^-$. Consider any set $\hat{\Gamma} \subseteq \Gamma^-$ and small $0 < \epsilon \leqslant 1$. Construct the measure $\hat{\tau}^- = \tau^-(\cdot | \hat{\Gamma})$, with marginals $\hat{\mu}^- = \text{proj}_{[0,\bar{v}]}\#\hat{\tau}^-$ and $\hat{\nu}^- = \text{proj}_{[\underline{v},0]}\#\hat{\tau}^-$. Define:

$$\tilde{\tau}^- = \frac{\tau^- - \epsilon\tau^-(\hat{\Gamma})\hat{\tau}^-}{1 - \epsilon\tau^-(\hat{\Gamma})},$$

with marginals $\tilde{\mu}^- = \frac{\mu^- - \epsilon\tau^-(\hat{\Gamma})\hat{\mu}^-}{1-\epsilon\tau^-(\hat{\Gamma})}$ and $\tilde{\nu}^- = \frac{\nu^- - \epsilon\tau^-(\hat{\Gamma})\hat{\nu}^-}{1-\epsilon\tau^-(\hat{\Gamma})}$. Then, construct $\tilde{\tau}^+$ to be positive assortative with marginals $\tilde{\mu}^+ = \frac{\mu^+ + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\hat{\mu}^-}{1+\frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})}$ and $\tilde{\nu}^+ = \frac{\nu^+ + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\hat{\nu}^-}{1+\frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})}$. Finally, consider the measure:

$$\tilde{\tau} = \alpha\left(1 + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\right)\tilde{\tau}^+ + (1-\alpha)(1-\epsilon\tau^-(\hat{\Gamma}))\tilde{\tau}^-.$$

The proof relies on the following lemma.

Lemma 6. *If $\tilde{\tau}$ is in $T^{\pm}$, then:*

$$\int G(2v' - v'' - k)d\tilde{\tau} > \int G(2v' - v'' - k)d\tau.$$

Proof. By substitution:

$$\int G(2v' - v'' - k)d\tilde{\tau} =$$

$$= \alpha\left(1 + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\right)\int G(2v' - v'' - k)d\tilde{\tau}^+ + (1-\alpha)(1-\epsilon\tau^-(\hat{\Gamma}))\int G(2v' - v'' - k)d\tilde{\tau}^-,$$

where:

$$+(1-\alpha)(1-\epsilon\tau^-(\hat{\Gamma}))\int G(2v' - v'' - k)d\tilde{\tau}^- =$$

$$= (1-\alpha)\int G(2v' - v'' - k)d\tau^- - (1-\alpha)\epsilon\tau^-(\hat{\Gamma})\int G(2v' - v'' - k)d\hat{\tau}^-.$$

Hence, the inequality in the statement of the present lemma holds if and only if:

$$\alpha\left(1 + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\right)\int G(2v' - v'' - k)d\tilde{\tau}^+ - (1-\alpha)\epsilon\tau^-(\hat{\Gamma})\int G(2v' - v'' - k)d\hat{\tau}^- >$$

$$> \alpha\int G(2v' - v'' - k)d\tau^+.$$

Now, define $\hat{\tau}^+ = \tilde{\tau}^+(\cdot|\text{supp}(\hat{\mu}^-) \times [\underline{v}, 0])$, with marginals $\hat{\mu}^+ = \text{proj}_{[0,\overline{v}]}\#\hat{\tau}^+ = \hat{\mu}^-$ and $\hat{\nu}^+ = \text{proj}_{[\underline{v},0]}\#\hat{\tau}^+$. Morevoer, define $\overline{\tau}^+ = \tilde{\tau}^+(\cdot|\text{supp}(\mu^+) \times [\underline{v}, 0])$, with marginals $\bar{\mu}^+ = \text{proj}_{[0,\overline{v}]}\#\overline{\tau}^+ = \mu^+$ and $\overline{\nu}^+ = \text{proj}_{[\underline{v},0]}\#\overline{\tau}^+$. Note that:

$$\tilde{\tau}^+ = \frac{\overline{\tau}^+ + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})\hat{\tau}^+}{1 + \frac{1-\alpha}{\alpha}\epsilon\tau^-(\hat{\Gamma})}.$$

By substituting in the inequality:

$$(1-\alpha)\epsilon\tau^-(\hat{\Gamma})\int G(2v' - v'' - k)d\hat{\tau}^+ - (1-\alpha)\epsilon\tau^-(\hat{\Gamma})\int G(2v' - v'' - k)d\hat{\tau}^- >$$

$$> \alpha\int G(2v' - v'' - k)d\tau^+ - \alpha\int G(2v' - v'' - k)d\overline{\tau}^+.$$

By assumption $\tilde{\tau}$ is in $T^{\pm}$, so that $\int G(2v' - v'' - k)d\hat{\tau}^+ > 0$.

To show the inequality holds, I proceed in three steps:

1. Because G is S-shaped and symmetric around 0, it must be that:

$$\frac{\int G(2v'-v''-k)d\hat{\tau}^{+}-\int G(2v'-v''-k)d\hat{\tau}^{-}}{\int(2v'-v''-k)d\hat{\tau}^{+}-\int(2v'-v''-k)d\hat{\tau}^{-}} > \\ > \frac{\int G(2v'-v''-k)d\tau^{+}-\int G(2v'-v''-k)d\overline{\tau}^{+}}{\int(2v'-v''-k)d\tau^{+}-\int(2v'-v''-k)d\overline{\tau}^{+}}$$

2. I have that:

$$\int(2v'-v''-k)d\hat{\tau}^{+}-\int(2v'-v''-k)d\hat{\tau}^{-}= \\ =-\int v''d\hat{\nu}^{+}+\int v''d\hat{\nu}^{-},$$

and:

$$\int(2v'-v''-k)d\tau^{+}-\int(2v'-v''-k)d\overline{\tau}^{+}= \\ =-\int v''d\nu^{+}+\int v''d\overline{\nu}^{+}.$$

Note that it must be that:

$$\nu^{+}+\frac{1-\alpha}{\alpha}\epsilon\tau^{-}(\hat{\Gamma})\hat{\nu}^{-}=\overline{\nu}^{+}+\frac{1-\alpha}{\alpha}\epsilon\tau^{-}(\hat{\Gamma})\hat{\nu}^{+},$$

so that:

$$\int(2v'-v''-k)d\tau^{+}-\int(2v'-v''-k)d\overline{\tau}^{+}= \\ =\frac{1-\alpha}{\alpha}\epsilon\tau^{-}(\hat{\Gamma})\int(2v'-v''-k)d\hat{\pi}^{+}-\int(2v'-v''-k)d\hat{\tau}^{-}.$$

3. Putting together 1. and 2. delivers the desired inequality.

□

Suppose $(1-\alpha)\neq 0$ and $2v'-v''-k>0$ for all $(v',v'')\in \text{supp}(\tau^{+})$. It suffices to show that there exist $\hat{\Gamma}$ and ϵ so that $\tilde{\tau}$ is in $T^{\pm}$. Consider $v^{\star}=\sup_{\text{supp}(\Gamma^{-})}(v',-v'')$. Call $B_{\delta}(v^{\star})\subseteq\Gamma^{-}$ a neighborhood of $v^{\star}$ in Γ^{-} of radius δ.

Because ϕ admits a continuous density, and $2v'-v''-k>0$ for all $v',v''\in\text{supp}(\tau^{+})$, there exist δ and ϵ, such that, for $\hat{\Gamma}=B_{\delta}(v^{\star})$, $2v'-v''-k\geqslant 0$ for all $v',v''\in\text{supp}(\hat{\tau}^{+})$ and for all $v',v''\in\text{supp}(\overline{\tau}^{+})$, so that $\tilde{\tau}$ is in $T^{\pm}$.

■

B.3 Proofs of Section 4

Proof of Proposition 5. By Proposition 1, any district π in an optimal plan $\mathcal{H}$ must be such that $\pi(\{v'\}) = \pi(\{v''\}) = \frac{1}{2}$ for $v' \geqslant 0$ and $v'' \leqslant 0$. Fix shock realization ω. Consider the following cases:

- $v' - \omega \geqslant k$. Then, $c_\pi = v' - \omega \geqslant k$.
- $v' - \omega < k$ and $v' - \omega - \frac{v''-\omega+k}{2} \geqslant k$. Then $c_\pi = k$.
- $v' - \omega < k$ and $v' - \omega - \frac{v''-\omega+k}{2} < k$. Then, $c_\pi = v'' - \omega \leqslant -\omega$.

Hence, $Q^{\omega}_{\mathcal{H}}((-\omega, k)) = 0$.

■

B.4 Proofs of Section 5

Proof of Proposition 6. First, note that a necessary, but not sufficient, condition for the designer to win district π when the realized shock is ω, is that $\pi^{\omega}(\{v : v \geqslant k\}) \geqslant \frac{2q-1}{2q}$. Suppose not. Then $\pi^{\omega}(\{v : v < k\}) > 1 - \frac{2q-1}{2q} = \frac{1}{2q}$. The equilibrium position of the Democratic candidate is $v^{q}_{\pi^{\omega},D}$, the leftmost q-quantile of π^{ω} conditional on $v < k$. Then, it must be that $\pi^{\omega}(\{v : v \leqslant v^{q}_{\pi^{\omega},D}\}) > \frac{1}{2}$, which means that the district is won by the Democrats. Now, for any redistricting plan $\mathcal{H}$ and any realization of the schok ω, define by $\chi^{\omega} \in \Delta([0,1])$ the distribution over $x^{\omega} = \pi^{\omega}(\{v : v \geqslant k\})$, that is the distribution of Republican voters across districts. Then, an upper bound on the designer utility is:

$$\int \mathbb{1}\left(x^{\omega} \geqslant \frac{2q-1}{2q}\right) d\chi^{\omega}(x^{\omega}) \leqslant \int \frac{2q}{2q-1} x^{\omega} d\chi^{\omega}(x^{\omega}) = \min\left\{1, \frac{2q}{2q-1}(1 - F(k+\omega))\right\},$$

where the equality holds by the law of iterated expectations. To finish the proof, it suffice to show that the above upper bound can always be achieved. Because F is uniform, it is easy to see that is achieved by the redistricting plan that matches any $v' \in [v^{q}, \overline{v}]$ to a $v'' = v^{q} - \frac{\overline{v}-v'}{2q-1} \in [0, v^{q}]$ and to a $v''' = -\frac{q}{2q-1}(\overline{v} - v) \in [\underline{v}, 0]$, with respective weights $\frac{2q-1}{2q}$, $\frac{1}{2q}$, and $\frac{1}{2}$.

■

Proof of Proposition 7. For any redistricting plan $\mathcal{H}$, call $H(\omega)$ the measure of districts won by the designer when the aggregate shock takes realization ω. By Proposition 6, at any realized shock ω, the designer can win at most measure $\min\left\{1, \frac{2q}{2q-1}(1-F(k+\omega))\right\}$ of districts, so $H(\omega) \leqslant \min\left\{1, \frac{2q}{2q-1}(1-F(k+\omega))\right\}$ at any ω. This implies that any feasible H must satisfy $H(\omega) \leqslant H^\star(\omega)$, where:

$$H^\star(\omega) = \begin{cases} 1 & if \;\; \omega \leqslant v^q - k \\ \frac{2q}{2q-1}(1-F(k+\omega)) & if \;\; \omega > v^q - k \end{cases}$$

The designer expected utility for any feasible H is then:

$$\int H(\omega) d\gamma(\omega) \leqslant \int H^\star(\omega) d\gamma(\omega),$$

with strict inequality if $H(\omega) \neq H^\star(\omega)$ for any ω. If $H^\star$ is attainable at every ω, $\mathcal{H}$ is optimal if and only if it induces $H^\star$. $H^\star$ is always attainable, as shown by the proof of Proposition 6. This means that $\mathcal{H}$-almost all the districts π the designer wins if and only if the shock is at most ω must satisfy $\pi(\{v : v = \omega\}) = 1 - \pi(\{v : v < v^q\}) = \frac{2q-1}{2q}$. However, since to win a district the designer needs at least $\frac{1}{2}$ of voters to vote for the Republican candidate and $\frac{2q-1}{q} \leqslant \frac{1}{2}$, some voters with $v < v^q$ must vote for R. Precisely $\frac{1}{2} - \frac{2q-1}{2q}$ additional voters must prefer $v = \omega$, the Republican candidate, to $v^q_{\pi^\omega, D}$, the Democratic candidate, with $v^q_{\pi^\omega, D}$ being the q - quantile of π^ω conditional on $v < k$. But given that $P(\{v : v < k\}) = \frac{1}{2q}$, the Democratic candidate sits at a median of π. For him not to win the district it must be that:

$$|v^q_{\pi^\omega, D} - v^m_{\pi^\omega}| \geqslant |v^m_{\pi^\omega} - \omega|.$$

■

B.5 Proofs of Appendix A

Proof of Proposition 8. Consider a Democratic voter with ideal point t (the reasoning is similar for a Republican voter). Given position y of the Republican candidate, the expected utility of electing a first-stage Democratic candidate with position x is:

$$U(x; y, t) = u(x, t)p(x, y) + u(y, t)(1 - p(x, y)),$$

where $p(x, y)$ is the probability of the Democratic candidate winning against the Republican candidate in the general elections. The proof of this result relies on the following lemma.

Lemma 7. *$U(x;y,t)$ is strictly increasing for $x<t$, strictly decreasing for $x>t$, and achieves its maximum at $x=t$.*

Proof. First, suppose $y>t$. Consider three cases:

Case $x>y$. First, note that $U(x;y,t)\leqslant U(y;y,t)$. Moreover, $U(x;y,t)$ is decreasing:

$$U(x;y,t)=-(x-t)\left(1-\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)-(y-t)\left(\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t)=-\left(1-\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)+\frac{x-t}{2(\overline{v}-\underline{v})}-\frac{y-t}{2(\overline{v}-\underline{v})}=$$

$$=-1+\frac{x+y-2\underline{v}}{2(\overline{v}-\underline{v})}+\frac{x-y}{2(\overline{v}-\underline{v})}=-1+\frac{x-2\underline{v}}{2(\overline{v}-\underline{v})}<0$$

Case $t<x<y$. $U(x;y,t)$ is decreasing:

$$U(x;y,t)=-(x-t)\left(\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)-(y-t)\left(1-\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t)=-\left(\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)-\frac{x-t}{2(\overline{v}-\underline{v})}+\frac{y-t}{2(\overline{v}-\underline{v})}=$$

$$=-\frac{x+y-2\underline{v}}{2(\overline{v}-\underline{v})}+\frac{y-x}{2(\overline{v}-\underline{v})}=\frac{-2x+2\underline{v}}{2(\overline{v}-\underline{v})}<0$$

Case $x<t$. $U(x;y,t)$ is increasing:

$$U(x;y,t)=-(t-x)\left(\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)-(y-t)\left(1-\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t)=\left(\frac{\frac{x+y}{2}-\underline{v}}{\overline{v}-\underline{v}}\right)+\frac{x-t}{2(\overline{v}-\underline{v})}+\frac{y-t}{2(\overline{v}-\underline{v})}=$$

$$=\frac{x+y-2\underline{v}}{2(\overline{v}-\underline{v})}+\frac{y+x-2t}{2(\overline{v}-\underline{v})}=\frac{2(x+y-t-\underline{v})}{2(\overline{v}-\underline{v})}>0$$

Second, suppose $y<t$. Consider three cases:

Case $x < y$. First, note that $U(x;y,t) < U(y;y,t)$. Moreover, $U(x;y,t)$ is increasing:

$$U(x;y,t) = -(t-x)\left(\frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) - (t-y)\left(1 - \frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t) = \left(\frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) + \frac{x-t}{2(\overline{v} - \underline{v})} - \frac{y-t}{2(\overline{v} - \underline{v})} =$$

$$= \frac{x+y-2\underline{v}}{2(\overline{v} - \underline{v})} + \frac{x-y}{2(\overline{v} - \underline{v})} = \frac{x-2\underline{v}}{2(\overline{v} - \underline{v})} > 0$$

Case $y < x < t$. $U(x;y,t)$ is increasing:

$$U(x;y,t) = -(t-x)\left(1 - \frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) - (t-y)\left(\frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t) = \left(1 - \frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) - \frac{x-t}{2(\overline{v} - \underline{v})} + \frac{y-t}{2(\overline{v} - \underline{v})} =$$

$$= 1 - \frac{x+y-2\underline{v}}{2(\overline{v} - \underline{v})} + \frac{y-x}{2(\overline{v} - \underline{v})} = 1 - \frac{-2x+2\underline{v}}{2(\overline{v} - \underline{v})} > 0$$

Case $x > t$. $U(x;y,t)$ is decreasing:

$$U(x;y,t) = -(x-t)\left(1 - \frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) - (t-y)\left(\frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right),$$

and by taking the first order derivative:

$$U'(x;y,t) = -\left(1 - \frac{\frac{x+y}{2} - \underline{v}}{\overline{v} - \underline{v}}\right) + \frac{x-t}{2(\overline{v} - \underline{v})} + \frac{y-t}{2(\overline{v} - \underline{v})} =$$

$$= -1 + \frac{x+y-2\underline{v}}{2(\overline{v} - \underline{v})} + \frac{y+x-2t}{2(\overline{v} - \underline{v})} = -1 + \frac{2(x+y-t-\underline{v})}{2(\overline{v} - \underline{v})} < 0.$$

Finally, note that $U(x;y,t)$ is continuous, so it is reaches its maximum at $x = t$, it is strictly decreasing for $x > t$, and strictly increasing for $x < t$.

$\square$

Given Lemma 7, a Democratic candidate at $v^m_{\pi,D}$ will win first-stage elections against any candidate at a different position (and similarly for Republicans). Indeed, Democratic voters to the right of $v^m_{\pi,D}$, accounting for half of all Democratic voters, prefer a candidate at $v^m_{\pi,D}$ to any other candidate to the left of $v^m_{\pi,D}$. Moreover, Democratic voters to the left of $v^m_{\pi,D}$, prefer a candidate at $v^m_{\pi,D}$ to any other candidate to the right of $v^m_{\pi,D}$. Importantly, this reasoning is independent of the Republican candidate's position y. Since positioning at $v^m_{\pi,D}$ ($v^m_{\pi,R}$, respectively) gives a Democratic (Republican) first-stage candidate a positive probability of winning second-stage elections, there exists an equilibrium where both Democratic candidates set at $v^m_{\pi,D}$ and both Republican candidates set at $v^m_{\pi,R}$.

Now, I show there can not be any other equilibrium. First, any situation where first-stage candidates do not tie can not be an equilibrium, because the losing candidate can move to $v^m_{\pi,D}$ ($v^m_{\pi,R}$) and have positive probability of winning second-statge elections. Second, any situation where first-stage candidates do not set the same position can not be an equilibrium. To see this, note that, in order to have different positions and tie at the same time, one of the Democratic (Republican) first-stage candidate needs to choose a position to the left (right) of $v^m_{\pi,D}$ ($v^m_{\pi,R}$). However, any such position is outside of the support of the median distribution, since candidates know the position of the conditional medians, and is therefore dominated by $v^m_{\pi,D}$ ($v^m_{\pi,R}$). Finally, suppose the two first-stage candidates set at the same position, different from $v^m_{\pi,D}$ ($v^m_{\pi,R}$). Then, there exists small $\epsilon > 0$ such that one of the candidates has a profitable deviation by moving closer to $v^m_{\pi,D}$ ($v^m_{\pi,R}$) by ϵ.

■