September 2006

Socially Optimal Districting:

A Theoretical and Empirical Exploration*

Abstract

This paper investigates the problem of optimal districting in the context of a simple model of legislative elections. In the model, districting matters because it determines the seat-vote curve, which describes the relationship between seats and votes. The paper first characterizes the optimal seat-vote curve, and shows that, under a weak condition, there exist districtings that generate this ideal relationship. The paper then develops an empirical methodology for computing seat-vote curves and measuring the welfare gains from implementing optimal districting. This is applied to analyze the districting plans used to elect U.S. state legislators during the 1990s.

Stephen Coate
Department of Economics
Cornell University
Ithaca NY 14853
sc163@cornell.edu

Brian Knight
Department of Economics
Brown University
Providence RI 02912
brian_knight@brown.edu

---

*This paper combines research reported in "Socially Optimal Districting" (NBER Working Paper #11462) and "Socially Optimal Districting: An Empirical Investigation" (NBER Working Paper #12313). We are greatly indebted to Jim Snyder for providing the data on state legislative elections used in this study. For helpful comments and encouragement we thank three anonymous referees, Alan Auerbach, Robert Barro, Tim Besley, Doug Bernheim, Allan Drazen, Andrew Gelman, Alan Gerber, Antonio Merlo, Tom Romer and seminar participants at NYU, UC Berkeley, Princeton, Rutgers, Stanford, Yale and the 2005 PIER conference on political economy. We thank the National Science Foundation (Grant SES-0452561) for financial support.

1 Introduction

The decennial redrawing of district lines used in electing candidates to federal, state, and local legislatures in the U.S. has generated intense interest among voters, politicians, and parties alike. The intensity of this interest in redistricting, as witnessed through both political and legal battles, should come as no surprise. Districting is a critical determinant of the representation of parties in legislatures. To illustrate, consider a legislature with 10 seats and suppose that 50% of the population are Democrats and 50% are Republicans. Then, while the "unbiased" seat share for the Democrats is 5, the actual share can vary from 1 to 9 depending upon the districting. For example, the Democrats can get 9 seats by creating 9 districts that are 51% Democrat.

The public interest in redistricting has given rise to a large empirical literature in political science analyzing redistricting plans and the redistricting process.[1] In this literature, redistricting plans are typically characterized by the implied relationship between seats in the legislature and support for parties among voters. In particular, the *seat-vote curve* relates the fraction of seats won by (say) the Democratic Party to their support among voters across all districts. Formally, the seat-vote curve is a function $S(V)$ where V is the aggregate fraction of votes received by the Democrats and S is the fraction of seats in the legislature that they hold. The key properties of a seat-vote curve are its *partisan bias* and *responsiveness*. Partisan bias - defined most simply as $S(1/2) - 1/2$ - measures how the districting advantages one or the other party.[2] Responsiveness - defined as $\Delta S/\Delta V$ - measures how the composition of the legislature changes in response to changes in citizens' voting behavior. Researchers have developed statistical methods to estimate seat-vote curves and to measure the associated responsiveness and partisan bias parameters. Basic questions are then how responsiveness and bias are altered following the redrawing of district lines and also how different institutional arrangements for redistricting influence these changes.

While this literature is certainly very interesting from a positive perspective, the normative lessons to be drawn from it are unclear. Is partisan bias necessarily a bad thing from the perspective of voters? What is the optimal degree of responsiveness? More generally, how should

[1] Important papers in this literature include Butler (1951), Gelman and King (1990 and 1994), King (1989), King and Browning (1987) and Tufte (1973). For general discussions of districting see Galderisi (2005) and Mann and Cain (2005).

[2] More generally, a seat-vote curve is said to display *partisan symmetry* if $S(V) = 1 - S(1 - V)$ for all V. A seat-vote curve is then defined to exhibit partisan bias if it deviates from partisan symmetry systematically in favor of one party.

voters be allocated across districts? Furthermore, how do the districting schemes that are used in practice compare with optimal schemes and what would be the welfare gains from optimal districting? Our purpose in this paper is to explore these questions.

We begin by developing a simple micro-founded model of legislative elections that provides a framework for investigating the districting problem which is consonant with the existing literature. There are three types of voters: Democrats, Republicans, and Independents. Democrats and Republicans have fixed ideologies, while Independents' are swing voters whose ideologies vary across elections. There are two political parties, one representing Democrats and the other Republicans. These parties field candidates in each district and the candidates with the most votes are elected to the legislature. Citizens care about the ideological make-up of the legislature which in turn depends upon the share of seats each party holds. The allocation of voter types across districts determines the seat-vote curve; i.e., the relationship between the Democratic seat share and their aggregate vote share.

In the context of this model, we analyze the problem of a districting authority who seeks to maximize expected social welfare. We assume that the authority can observe citizens' types (i.e., whether they are Democrats, Republicans, or Independents) and we abstract from geographical constraints in the ability to craft districts. We approach the problem by characterizing the *optimal seat-vote curve,* which describes the ideal relationship between seats and votes. We then develop a condition under which the optimal seat-vote curve is implementable in the sense that there exist districtings that generate this optimal relationship between seats and votes. These underlying districtings are *socially optimal districtings,* and we describe their composition.

We then use the model to develop an empirical methodology for assessing how actual districting plans compare with optimal plans and for computing the welfare gains from implementing optimal districting. This methodology has four distinct components: (i) a method for estimating actual seat-vote curves, (ii) a method for computing optimal seat-vote curves, (iii) a way of checking the condition for implementability and, (iv) a way of estimating the welfare gains from implementing optimal seat-vote curves. Finally, we apply this methodology to analyze the districting plans used to elect U.S. state legislators during the 1990s. We estimate actual and optimal seat-vote curves, check the condition for implementability, and then compute the welfare gains from optimal districting. We also compare actual and optimal seat-vote curves with those that would emerge under a system of proportional representation or with at-large elections.

Despite its obvious importance, the problem of socially optimal districting has attracted scant attention.[3] The bulk of existing theoretical work on districting focuses on the *partisan gerrymandering problem*: that is, how to craft political districts with the aim of maximizing a party's expected seat share.[4] The motivation is the purely positive one of shedding light on how partisan redistricting committees might further their political objectives.[5] The few normative papers that have been written work with very different underlying political models and objective functions than used in this paper. In work done independently, Gilligan and Matsusaka (2005) look at the optimal districting problem from a median voter perspective. In Downsian fashion, they assume that candidates adopt the ideology of the median voter in their districts and that policy outcomes depend upon the ideology of the median legislator. Their social objective is thus to minimize the distance between the median legislator's ideology and the median voter's ideology. Epstein and O'Hallaran (2004) focus on the *racial gerrymandering problem*, under which the planner attempts to maximize the welfare of minority groups.[6] More generally, we are not aware of either any normative work on districting that is consistent with the seat-vote curve approach of the empirical literature on districting or any work that evaluates the gains from optimal districting empirically.[7]

---

[3] This is despite the existence of a vibrant literature applying contemporary political economy modelling and welfare economic methods to explore the optimal design of political institutions. This literature includes efforts to understand the relative merits of different electoral systems (e.g., Lizzeri and Persico (2001) and Myerson (1999)); systems of campaign finance (e.g., Coate (2004) and Prat (2002)); and methods of choosing policy-makers (e.g., Maskin and Tirole (2004)). It also includes analyses of the desirability of citizens' initiatives (e.g., Matsusaka and McCarty (2001)); the optimal allocation of functions across layers of government (e.g., Lockwood (2002)); and the relative merits of presidential and parliamentary systems (e.g., Persson, Roland and Tabellini (2000)). The districting problem is somewhat different from these constitutional design questions in that it must be done on an on-going basis in any political system with geographically based districts. This makes the problem particularly salient.

[4] Papers in this tradition include Owen and Grofman (1988), Gilligan and Matsusaka (1999), Sherstyuk (1998) and Shotts (2001), (2002). More recently, Friedman and Holden (2006) revisit the partisan gerrymandering problem, providing an elegant and comprehensive analysis.

[5] In an interesting application of the theory, Shotts (2001) and (2002) uses his positive models of partisan gerrymandering to understand the policy implications of mandating that districting authorities form so-called *majority-minority districts*.

[6] Their analysis formalizes the intuition that there maybe a trade-off between *descriptive* and *substantive* representation. Descriptive representation is achieved by having districts elect black representatives, while substantive representation is achieved when the legislature chooses policies that favor black voters. Maximizing descriptive representation may require concentrating black voters into majority-minority districts, while maximizing substantive representation may require a more even spreading of black voters. The underlying structure of Epstein and O'Hallaran's model is simpler than the one presented in this paper in that it does not allow for Independents and there is no aggregate uncertainty in voters' preferences. On the other hand, to incorporate substantive representation, they model strategic policy choices on the part of politicians, whereas in our model parties' ideologies are fixed.

[7] In an interesting paper, Cameron, Epstein and O'Halloran (1996) develop a methodology for assessing the effect of different districting schemes on the substantive representation of minority interests as measured by U.S. House

Our paper also contributes to the empirical literature mentioned earlier on the estimation of seat-vote curves. Our key point of departure from this literature is the development of a micro-founded model that guides our empirical methodology. The only other work we are aware of that explores the micro foundations of seat-vote curves is the independent work of Besley and Preston (2006). These authors are interested in understanding how districting impacts the strategic platform choices of political parties. They develop a similar micro-founded model that generates an equilibrium relationship between seats and votes. Their main theoretical point is to show that the partisan bias of the seat-vote curve is a key determinant of parties' electoral incentives to be responsive to swing voters with respect to their platform choices. They provide empirical evidence in favor of their theory by showing that local government policy choice in the U.K. is related to the parameters of the local seat-vote curve in the way the theory predicts. Their work therefore identifies a theoretical mechanism by which the form of the seat-vote curve (and hence districting) matters for citizens' welfare and provides evidence for this. By contrast, our model, which assumes fixed party ideologies, reflects the conventional view that districting matters because it determines which party gets the most seats and hence the ideological composition of the legislature.

The paper proceeds as follows. Section 2 outlines the model, describes the optimal districting problem, and explains how we tackle it. Section 3 presents the key theoretical results. Section 4 develops our empirical methodology and Section 5 applies this methodology to U.S. state legislative elections. Section 6 concludes with a summary of the main results of the analysis.

2 The model

Consider a state in which there are three types of voters - Democrats, Republicans, and Independents. Voters differ in their political ideologies which are measured on a 0 to 1 scale. Democrats and Republicans have ideologies 0 and 1, respectively. Independents have ideologies that are uniformly distributed on the interval $[m-\tau, m+\tau]$ where $\tau > 0$. These voters are "swing voters" and so the ideology of the median Independent may vary across elections. Specifically, m is the realization of a random variable uniformly distributed on $[1/2-\varepsilon, 1/2+\varepsilon]$, where $\varepsilon \in (0, \tau)$ and $\varepsilon + \tau \leq 1/2$. The latter assumption guarantees that Independents are always between De-

members' roll-call voting scores on minority issues. This methodology is then applied to calculate the districting strategy that would maximize substantive black representation in the U.S. House.

mocrats and Republicans, while the former guarantees that some Independents lean Democrat and some lean Republican. The fractions of voters statewide who are Democrats, Republicans, and Independents are, respectively, π_D, π_R and π_I.

Policy choices in the state are determined by an n-seat legislature. Legislators' policy choices are influenced by their ideologies and, hence, citizens care about the ideological make up of their legislature. Specifically, if the average ideology of the legislature is x', a citizen with ideology x experiences a quadratic payoff given by $\beta - \gamma(x - x')^2$. The parameter β is the surplus a citizen would obtain from having a legislature that is perfectly congruent with his own ideology and the parameter γ measures the rate at which this surplus is dissipated as the ideology of the legislature diverges from that of the citizen. The ratio γ/β will play an important role in the welfare analysis and can be interpreted as the fraction of the surplus a partisan (i.e., a Democrat or Republican) obtains from having a perfectly congruent legislature that is dissipated by having a legislature composed entirely of the opposing ideology. This ratio is assumed to be bounded between zero and one ($\gamma/\beta \in [0, 1]$).[8]

There are two political parties in the state: the Democrats and the Republicans. These parties are made up of collections of citizens who share the same political ideology, so that the membership of the Democrat Party are Democrats and the membership of the Republican Party are Republicans. Legislators are all affiliated with one or the other party.

To select legislators, the state is divided into n equally-sized (in terms of population) political districts indexed by $i \in \{1/n, 2/n, ..., 1\}$. Each district then elects a representative to the legislature. Candidates are put forward by the two political parties. Following the *citizen-candidate approach* (Besley and Coate (1997) and Osborne and Slivinski (1996)), there is no commitment so that candidates cannot credibly promise to run on an ideology different from their true ideology. Accordingly, Democratic candidates are associated with the ideology 0 and Republican candidates with the ideology 1. Elections are held simultaneously in each of the n districts and the candidate with the most votes wins. In each district, every citizen votes sincerely for the representative whose ideology is closest to his own.

[8] As will be seen below, most of our results, with the exception of the measure of welfare gains, are independent of the exact values of the parameters β and γ.

2.1 Districtings

A *districting* is a division of the population into n districts. Formally, a districting is described by $(\pi_D(i), \pi_I(i), \pi_R(i))_{i=1/n}^{1}$ where $\pi_D(i)$ represents the fraction of Democrats in district i, $\pi_R(i)$ the fraction of Republicans, and $\pi_I(i)$ the fraction of Independents. The districting is chosen by a districting authority that knows the group membership of citizens and faces no geographic constraints in terms of how it can group citizens. Thus, any districting $(\pi_D(i), \pi_I(i), \pi_R(i))_{i=1/n}^{1}$ such that the average fractions of voter types equal the actual is feasible. Let Ω denote the set of feasible districtings and, to simplify notation, let $\delta \in \Omega$ denote a generic feasible districting.

2.2 Seat-vote curves

Any feasible districting $\delta = (\pi_D(i), \pi_I(i), \pi_R(i))_{i=1/n}^{1}$ implies a relationship between the Democratic seat share in the legislature and their statewide vote share. Note first that if the median independent has ideology m, the fraction of voters in district i voting for the Democrat candidate is

$$V(i; m) = \pi_D(i) + \pi_I(i)[\frac{1/2 - (m - \tau)}{2\tau}]. \tag{1}$$

This group consists of all the Democrats and those Independents whose ideologies are less than 1/2. The statewide vote share of the Democrat Party is therefore

$$V(m) = \pi_D + \pi_I[\frac{1/2 - (m - \tau)}{2\tau}]. \tag{2}$$

Let $\overline{V}$ and $\underline{V}$ denote, respectively, the maximum and minimum statewide Democrat vote shares; i.e., $\overline{V} = V(1/2 - \varepsilon)$ and $\underline{V} = V(1/2 + \varepsilon)$.

Now, for any feasible statewide vote share $V \in [\underline{V}, \overline{V}]$, let $m(V)$ denote the ideology of the median Independent that would generate the vote share V; i.e., $m(V) = V^{-1}(V)$. From (2), we have that

$$m(V) = \frac{1}{2} + \tau[\frac{\pi_I + 2\pi_D - 2V}{\pi_I}]. \tag{3}$$

Substituting this into (1), we obtain

$$V(i; m(V)) = \pi_D(i) + \pi_I(i)[\frac{V - \pi_D}{\pi_I}]. \tag{4}$$

District i elects a Democrat if $V(i; m(V)) \geq 1/2$, or, equivalently, if

$$V \geq V^*(i) = \pi_D + \pi_I[\frac{1/2 - \pi_D(i)}{\pi_I(i)}], \tag{5}$$

where $V^*(i)$ is the critical statewide vote threshold above which district i elects a Democrat. District i is a *safe Democrat* (*safe Republican*) seat if $V^*(i) \leq \underline{V}$ ($V^*(i) \geq \overline{V}$). A seat which is not safe is *competitive*.

Without loss of generality, order the districts so that $V^*(1/n) \leq V^*(2/n) \leq .... \leq V^*(1)$. Then, the fraction of seats the Democrats receive when they have vote share V is

$$S(V\,|\delta\,) = \max\{i : V^*(i) \leq V\}. \tag{6}$$

This is the *seat-vote curve* associated with the districting δ.

2.3 Socially optimal districtings

We are interested in the problem of a districting authority who desires to maximize expected aggregate utility.[9] Aggregate utility when the median Independent has ideology m and the Democrats have seat share S is given by:

$$W(S,m) = \beta - \gamma[\pi_D(1-S)^2 + \pi_R S^2 + \pi_I \int_{m-\tau}^{m+\tau} (1-S-x)^2 \frac{dx}{2\tau}]. \tag{7}$$

If the Democratic vote share is V, the median Independent has ideology $m(V)$ and hence expected aggregate utility under the districting δ is given by

$$\int_{\underline{V}}^{\overline{V}} W(S(V\,|\delta\,), m(V)) \frac{dV}{\overline{V}-\underline{V}}. \tag{8}$$

The districting authority's problem is therefore to choose a districting δ to solve the problem

$$\begin{array}{c} \max \int_{\underline{V}}^{\overline{V}} W(S(V\,|\delta\,), m(V)) \frac{dV}{\overline{V}-\underline{V}} \\ \text{s.t.} \quad \delta \in \Omega. \end{array} \tag{9}$$

A districting δ that solves this problem is a *socially optimal districting*.

The authority's problem is complicated by the fact that there is not a one-to-one mapping between districtings and seat-vote curves. The seat-vote curve is determined by the pattern of critical vote thresholds across districts. As is clear from (5), the same pattern of critical vote thresholds could in principle be achieved by many different districtings. To solve the problem, therefore, it is simpler to think of the districting authority as directly choosing a seat-vote curve

[9] This objective function is equivalent to aggregate expected utility under the assumption that each Independent voter is ex ante identical. Under this assumption, for a given draw of m, each Independent is equally likely to have any ideology on $[m-\tau, m+\tau]$. The expected payoff of any Independent voter is then just the payoff of the average Independent voter.

but subject to the implementability constraint that there exists a districting that generates it. Thus, we recast the districting authority's problem as choosing a seat-vote curve $S(V)$ to solve the problem

$$\begin{aligned} &\max \int_{\underline{V}}^{\overline{V}} W(S(V), m(V)) \tfrac{dV}{\overline{V}-\underline{V}} \\ &\text{s.t.} \quad S(V) = S(V\,|\delta\,) \text{ for some } \delta \in \Omega. \end{aligned} \tag{10}$$

The socially optimal districtings will then be those associated with the constrained optimal seat-vote curve.

The difficulties of handling the constraint that a seat-vote curve be generated by some feasible districting make this a challenging problem. Accordingly, we begin the analysis by characterizing the optimal relationship between seats and aggregate votes - the optimal seat-vote curve - ignoring the constraint that it be generated by some feasible districting. We then investigate whether there exist districtings that generate this optimal seat-vote curve. If there do exist such districtings, these will clearly be optimal. This two-stage procedure will not totally eliminate the need to consider the grand constrained optimization, but appears to do so in the empirically relevant cases.

2.4 Discussion of the model

In developing a model in which to analyze optimal districting, we faced many different modelling choices. In order for the reader to understand the reasons for the choices we have made, we briefly discuss the key ones here.

A first question was how to assume that the ideological composition of the legislature impacted citizens' payoffs. Tractability obviously dictated the use of a simple summary statistic of the distribution of legislator ideologies. The *median voter theorem* not withstanding, we choose the average rather than the median. This choice was motivated by two main considerations. First, the median assumption corresponds to the idea that the party with the majority of seats is completely decisive and minority party members have no influence at all. We feel this is unrealistic. A real world state legislature makes numerous decisions on many different areas of policy. Many of these decisions will be made by small sub-committees of legislators. This gives a legislator influence even if he is not in the majority party. Reflecting this, we seriously doubt that voters would be indifferent between a state legislature that is 51% Democrat and one that is 100% Democrat. In an empirical analysis of policy outcomes in U.S. states, Besley and Case (2003) provide evidence

in support of this general view; in particular, conditional on the Democrats controlling the state legislature, a 10% increase in the fraction Democrat in both the Lower and Upper House leads to a $10 per-capita increase in government spending (in 1982 dollars). Second, we wanted to be consistent with the existing literature which, given its focus on the properties of seat-vote curves, clearly distinguishes between state legislatures with different sized majorities. Thus, we needed the form of the seat-vote curve over its entire domain to matter for citizens' welfare. But, under the median assumption, the properties of the seat-vote curve are irrelevant for welfare over almost all of its domain. All that matters is the vote share at which the Democrats become the majority party.[10]

A second key modelling choice concerned the strategic choices faced by parties and candidates. We abstract from such considerations by assuming: (i) that each party simply puts up candidates who are party members and (ii) that members all share the same ideology. Moreover, we employ the citizen-candidate assumption, so that candidates have no choice but to effectively run on their true ideologies.[11] While it would certainly be interesting to extend the model to allow parties some flexibility in candidate choice, we believe that assuming away strategic choices is the natural place to start.[12] The first reason is that, under a Downsian vision of political competition in which candidates adopt the ideology that makes them most likely to win, it is not possible to consider the problem in terms of seat-vote curves.[13] Both parties' candidates in each district would adopt the position of the expected median voter and which candidate won would have no significance for welfare. Thus, the seat-vote curve and the ideas of partisan bias or responsiveness would cease to have much meaning. The second point to note is that, once again, we feel that it is implicit in the literature on estimating seat-vote curves that legislators from the same party have similar

[10] For example, suppose that the aggregate vote for Democrats increases from 30% to 40% and suppose their initial seat share is 30%. Then whether their seat share increases to 35% or 45% has no impact on policy because in either situation the median legislator remains a Republican. Thus, the responsiveness of the seat-vote curve over this part of the domain is irrelevant.

[11] Recent empirical evidence is consistent with this no commitment assumption. In particular, Lee, Moretti, and Butler (2004) find that exogenous changes in electoral support for one party does not induce changes in campaign platforms, as measured in post-election roll call voting data.

[12] As noted in the introduction, Besley and Preston (2006) study how districting impacts the platform choices of competing political parties in a model with a very similar flavor to ours. Their work could be a starting place for analyzing the problem of optimal districting taking into account its impact on parties' strategic choices.

[13] As noted in the introduction, Gilligan and Matsusaka (2005) study the optimal districting problem from this Downsian perspective under the assumption that the objective function is to minimize the distance between the ideology of the median legislator and the median voter. Their main finding is that identical districts are optimal. With identical districts, each candidate adopts the position of the population median voter and all candidates have a homogeneous (and optimal) ideology.

ideologies. If this were not the case, the notions of responsiveness and bias would be much more complicated. To illustrate, suppose that Democratic candidates came in two types - moderate and extreme. Then, whether a given aggregate Democratic vote increase led to the election of more moderate or more extreme Democrats would be very relevant for the true responsiveness of the system. Accordingly, responsiveness could not simply be measured by the slope of the seat-vote curve. Similarly, if Democrats get more seats when voters are evenly divided between the parties, then the true bias would naturally depend upon how Democratic seats are divided between moderates and extremists.

A third modelling choice concerned what to assume about citizens' voting behavior. It is important to note that an Independent voter who leans Democrat may be better off when his district elects a Republican if other districts disproportionately elect Democrats. For if his district elected a Republican, the average legislator ideology would be closer to his ideal point. We decided nonetheless to assume sincere voting (i.e., voting for the ideologically closest candidate) for two reasons. First, assuming sophisticated voting would substantially complicate the analysis because voters' optimal decisions would be strategic and determined as part of a statewide voting equilibrium. This equilibrium (which need not be unique) would of course be influenced by the districting. Second, as an empirical matter, it is not clear that most voters are this sophisticated. Similar incentives to diverge from voting for the candidate closest to one's own ideology arise when voters are electing Congressional and Presidential candidates and policy outcomes depend upon the ideologies of both Congress and the President (Alesina and Rosenthal (1995) and Fiorina (1992)). However, using a data set on voting behavior in these elections, Degan and Merlo (2006) show that sincere voting can explain virtually all individual-level observations.

A fourth issue was how to introduce uncertainty into citizens' voting behavior. Our model incorporates *individual uncertainty*, which is defined as variation in voting behavior across Independents in a given election and is captured by the parameter τ, as well as *aggregate uncertainty*, which is defined as variation in voting behavior of a given Independent across elections and is captured by the parameter ε. It is clear that aggregate uncertainty is required in order to generate a seat-vote curve.[14] But rather than assume that all voters are subject to a uniform ideological swing, we choose to assume that only Independents' preferences were uncertain. We took this

[14] When $\varepsilon = 0$, the fraction of Independents supporting the Democrats is certain and the seat-vote curve is degenerate.

approach for two reasons. First, with a uniform swing impacting all voters, our citizen-candidate framework would imply that candidate ideologies would also shift and thus votes and seats would be unchanged; this would clearly thwart our objective of generating a seat-vote curve. Second, in our empirical analysis, as will be clear below, the year-to-year variation in voting returns is used to identify the degree of variation in the ideology of Independent voters. Changes in partisan ideology, by contrast, are not reflected in voting patterns if, for example, left-leaning Republicans nonetheless vote for the Republican over the Democrat and thus it is not clear how the variation in partisan ideology would be identified empirically.

A fifth choice was to assume that the districting authority could both observe group membership and could form districts freely with no geographic constraints. The former assumption seems reasonable as a first approximation, since information about voters' ideological attachments is available through voter registration data or the study of past voting patterns (see the discussion in Altman, Mac Donald and McDonald (2005)).[15] The latter assumption is more difficult to defend on realism grounds and the neglect of the requirement that districts be connected subsets of some geographic space is certainly a weakness of the analysis. However, we feel that given the difficulty of knowing how to model geographic constraints, it makes sense to first understand what optimal districtings look like without them. Moreover, as we will see below, when the optimal seat-vote curve is implementable it can typically be implemented by a large class of districtings, some of which look quite "straightforward". Hence geographic constraints may actually be easily accommodated.

The final choice involved working with very specific assumptions on citizens' political preferences. Specifically, we assume that citizens have quadratic loss functions and that the distribution of Independents' ideologies is uniform across its support. These assumptions are made to both keep the theoretical problem tractable and to facilitate the development of an empirical methodology tightly tied to the theory. A key role of these assumptions is to ensure that the welfare function defined in (7) has a quadratic form. As we will see in the next section, this results in the optimal seat-vote curve having a simple linear form. This linearity is of great help when deriving the condition under which there exist districtings that make the seat-vote curve optimal. Furthermore, the quadratic form of the welfare function is key to enabling us to solve for optimal

[15] That said, it would certainly be interesting to extend the analysis here to the case in which the districting authority only observes a signal of each voter's ideology (as in Friedman and Holden (2006)).

districtings when this condition is violated. Obviously, it would be desirable to work with a more general model, but both our assumptions (i.e., quadratic preferences and uniformly distributed ideologies) are common in contemporary political economy models. Moreover, our formulation of the problem and the concepts we introduce are general and the considerations that our analysis identifies will be present in more general models.

3 Some theoretical results

This section explores the theory of socially optimal districting. It begins by characterizing the optimal seat-vote curve. It then derives a necessary and sufficient condition for this seat-vote curve to be implementable. Next it describes what the districtings that generate the optimal seat-vote curve look like when this condition is satisfied. It then briefly characterizes optimal districting when the implementability condition is not satisfied. Finally, it derives a useful formula for the welfare gains from optimal districting.

3.1 The optimal seat-vote curve

The *optimal seat-vote curve* $S^o(V)$ describes the ideal relationship between Democratic seats and aggregate votes ignoring the constraint that this relationship be generated by some feasible districting. To avoid tedious integer concerns, we assume that the number of districts is very large, so that we can treat S as a continuous variable defined on the unit interval $[0,1]$. Then, we obtain the following simple characterization of the optimal seat-vote curve:

Proposition 1: *The optimal seat-vote curve* $S^o : [\underline{V}, \overline{V}] \to [0,1]$ *is given by*

$$S^o(V) = 1/2 + (\pi_D - \pi_R)(1/2 - \tau) + 2\tau(V - 1/2). \tag{11}$$

Proof: The optimal seat-vote curve $S^o(V)$ is such that for all $V \in [\underline{V}, \overline{V}]$

$$S^o(V) = \arg \max_{S \in [0,1]} W(S, m(V)).$$

Assuming an interior solution, $S^o(V)$ satisfies the first order condition: $\partial W(S^o, m(V))/\partial S = 0$. Differentiating (7) yields

$$\partial W(S,m)/\partial S = 2\{\pi_D + \pi_I(1-m) - S\}.$$

Thus, $\partial W(S^o, m(V))/\partial S = 0$ if and only if $S^o = \pi_D + \pi_I(1 - m(V))$. In addition, note that $\partial^2 W(S, m)/\partial S^2 < 0$ so that the first order condition is sufficient for S^o to be optimal. Substituting in the expression for $m(V)$ from (3) yields the result. ■

Proposition 1 tells us that the optimal seat-vote curve is of the same linear form that has been estimated in some of the early empirical literature; that is, $S(V) = 1/2 + b + r(V - 1/2)$ where b measures partisan bias and r measures responsiveness (see, for example, Tufte (1973)).[16] The bias of the optimal curve is $(\pi_D - \pi_R)(1/2 - \tau)$ and its responsiveness is 2τ. In sharp contrast to what is implicitly assumed in the districting literature, the optimal curve exhibits partisan bias, and this bias is in favor of the party with the largest partisan base. The responsiveness of the optimal curve depends on the degree of variation in the preferences of swing voters.

The optimal seat-vote curve is illustrated in Figure 1. The horizontal axis measures the aggregate Democratic vote and the vertical the Democrats' share of seats. Since $\tau < 1/2$, the slope of the optimal seat-vote curve is less than 1 meaning that the fraction of Democrat seats increases at a constant but less than proportional rate as the aggregate Democrat vote increases. The seat-vote curve intersects the 45^o line when the aggregate vote is $\pi_D + \pi_I/2$. Thus, when exactly half the Independents lean Democrat, the optimal share of Democratic seats is $\pi_D + \pi_I/2$. Notice also that $S^o(\underline{V}) > 0$ and $S^o(\overline{V}) < 1$ so that there are safe seats for both parties.

To understand why the optimal responsiveness is 2τ, note first that the welfare maximizing Democratic seat share must be such that the social gains from increasing it marginally just equal the social losses. With the quadratic preferences, this marginal condition implies that the Democratic seat share must be such as to make the ideology of the average legislator equal the average ideology in the population. Thus, when the mean (which equals the median) Independent has ideology m, the optimal Democrat seat share should be $\pi_D + \pi_I(1 - m)$ because this would make the average ideology in the legislature equal to the population average - which is $\pi_R + \pi_I m$. When the aggregate Democrat vote share increases marginally, the change in the mean Independent's ideology is $dm/dV = -2\tau/\pi_I$ (see (3)) and hence the increase in the optimal Democrat seat share is just 2τ. Recall that τ measures the diversity of views among Independents, so that respon-

[16] As noted in section 2.4, the linearity of the optimal seat-vote curve follows from our assumptions that citizens have quadratic loss functions and that the distribution of Independents' ideologies is uniform across its support. In Coate and Knight (2005) we explore the implications of more general assumptions for the optimal seat-vote curve. We show that the factors identified in the basic model (i.e., $(\pi_D - \pi_R)$ and 2τ) remain key determinants of its partisan bias and responsiveness.

siveness is positively correlated with this diversity. This is because the greater the diversity of Independent views, the greater the change in mean Independent ideology signalled by any given increase in vote share.

To understand why the optimal seat-vote curve is biased, consider the case when the Democrats get exactly half the aggregate vote ($V = 1/2$). If the optimal seat-vote curve were unbiased then the Democrats should get half the seats ($S^o(1/2) = 1/2$). This would indeed be optimal if the average ideology in the population were 1/2. However, while the *median* voter in the population must have ideology 1/2 in this case, the *average* voter's ideology will only equal 1/2 when the fractions of Democrats and Republicans are equal. To see this, note from (3) that when $V = 1/2$, the median Independent's ideology must be $m(1/2) = 1/2 + \tau(\pi_D - \pi_R)/\pi_I$ which implies that the average ideology in the population is $1/2 + (\pi_R - \pi_D)(1/2 - \tau)$. Thus, to make the average legislator's ideology equal to the population average it will be necessary to have the Democratic seat share greater than 1/2 if π_D exceeds π_R. Fundamentally, then, the bias in the optimal seat-vote curve stems from the fact that the ideology of the median voter will typically differ from that of the average voter. This in turn reflects the fact that partisans feel more intensely about ideology than do swing voters.

3.2 When is the optimal seat-vote curve implementable?

Having understood the nature of the optimal seat-vote curve, we now turn to the question of implementability; that is, whether there exist districtings δ such that $S(V\,|\delta\,) = S^o(V)$. Such a districting would make the composition of the legislature such that average legislator ideology always equals the population average. Clearly, this cannot be achieved by making each district a microcosm of the community as a whole, because then all districts would vote in the same way and the legislature would be either all Democrat or all Republican. However, with appropriate district level heterogeneity, implementability seems possible. While the conditions that might guarantee it are by no means obvious, it is apparent that the fraction of Independents must matter. For, if there were no Independents, then the optimal seat-vote curve would be a single point and could be implemented, for example, by creating a fraction π_R districts majority Republican and a fraction π_D districts majority Democrat. On the other hand, if the entire population were Independents, then all districts would necessarily be identical and the optimal seat-vote curve is clearly not

implementable.[17]

Reflecting the importance of the fraction of Independents, we have the following key result:

Proposition 2: *The optimal seat-vote curve is implementable if and only if*

$$\pi_I\left(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau+\varepsilon)\ln(1+\frac{\varepsilon}{\tau})\right) \leq \min\{\pi_D, \pi_R\}. \tag{12}$$

Proof: See Appendix.

The condition in Proposition 2 is that there be "enough" Republicans and Democrats relative to Independents. There are several points to note about the condition. First, for all τ, the coefficient multiplying the fraction of Independents (i.e., $\frac{\varepsilon}{2\tau} + \varepsilon - (\tau+\varepsilon)\ln(1+\frac{\varepsilon}{\tau})$) is increasing in ε and converges to zero as ε converges to zero. Thus, the optimal seat-vote curve is necessarily implementable when the degree of aggregate uncertainty is sufficiently small. Intuitively, even though the districting authority cannot predict how any *individual* Independent voter will vote, it can predict the *fraction* of Independents who will vote for each party. This enables it to achieve the optimal Democratic seat share.[18] Second, for given ε, the coefficient is decreasing in τ and hence the optimal seat-vote curve is more likely to be implementable when there is more individual uncertainty. Intuitively, a larger τ makes the voting behavior of Independents as a group less volatile for a given shift in the ideology of the median Independent.

How permissive is the condition in Proposition 2? It is worth noting here that for any values of ε and τ satisfying our assumptions, $\varepsilon \leq (\tau+\varepsilon)\ln(1+\frac{\varepsilon}{\tau})$ and hence the coefficient multiplying the fraction of Independents is less than $\varepsilon/2\tau$. This in turn is less than $1/2$ and hence a sufficient condition for the optimal seat-vote curve to be implementable is that $\pi_I \leq 2\min\{\pi_D, \pi_R\}$. While the empirical application to follow will provide a more complete test for implementability, we simply note here that this requirement appears permissive. For example, according to data from Erikson, Wright and McGuiver (1993), this sufficient condition is violated in just four U.S. states.

[17] In this case, the optimal seat-vote curve is $S^o(V) = 1/2 + 2\tau(V - 1/2)$, while the only feasible seat-vote curve is $S(V) = 0$ if $V < 1/2$ and $S(V) = 1$ if $V > 1/2$.

[18] It is easy to see how. When $\varepsilon = 0$ exactly $1/2$ of the Independents will vote Democrat and $1/2$ will vote Republican. The optimal seat share for the Democrats is $\pi_D + \pi_I/2$. Consider a districting that grouped all the Democrats together, all Independents together and all Republicans together. This would generate $\pi_D + \pi_I/2$ Democratic seats if the ties in the Independent districts were resolved by the toss of a fair coin. The problem of ties can be avoided by transferring a small number of Democrats into $1/2$ the Independent districts and a small number of Republicans into the other half.

3.3 The optimal districtings

What do the districtings that generate the optimal seat-vote curve look like? When the condition of Proposition 2 is satisfied, we can use arguments used in the proof of Proposition 2 to show that the optimal seat-vote curve can always be implemented by a districting of the general form

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\underline{\pi}_D, \underline{\pi}_I) & \text{if } i \in [0, S^o(\underline{V})) \\ (\frac{\pi_D + \frac{\pi_I}{2} - i}{\pi_D + \frac{\pi_I}{2} - i + \pi_I \tau}, \frac{\pi_I \tau}{\pi_D + \frac{\pi_I}{2} - i + \pi_I \tau}) & \text{if } i \in [S^o(\underline{V}), \pi_D + \frac{\pi_I}{2}) \\ (0, \frac{\pi_I \tau}{i - (\pi_D + \frac{\pi_I}{2}) + \pi_I \tau}) & \text{if } i \in [\pi_D + \frac{\pi_I}{2}, S^o(\overline{V})] \\ (\overline{\pi}_D, \overline{\pi}_I) & \text{if } i \in (S^o(\overline{V}), 1]. \end{cases} \tag{13}$$

While this does not specify the fractions of Republicans $\pi_R(i)$, these can be readily obtained from the equality $\pi_R(i) = 1 - \pi_D(i) - \pi_I(i)$.

Districts $i \in [0, S^o(\underline{V}))$ are the safe Democratic seats. The fractions of Democrats and Independents in these seats $(\underline{\pi}_D, \underline{\pi}_I)$ must satisfy the inequality

$$\underline{\pi}_D + \underline{\pi}_I [\frac{\tau - \varepsilon}{2\tau}] \geq \frac{1}{2}. \tag{14}$$

This reflects the fact that the minimum fraction of Independents voting Democrat is $\frac{\tau - \varepsilon}{2\tau}$. Districts $i \in (S^o(\overline{V}), 1]$ are the safe Republican seats and, because the maximum fraction of Independents voting Democrat is $\frac{\tau + \varepsilon}{2\tau}$, the voter allocations in these seats $(\overline{\pi}_D, \overline{\pi}_I)$ satisfy the inequality

$$\overline{\pi}_D + \overline{\pi}_I [\frac{\tau + \varepsilon}{2\tau}] \leq \frac{1}{2}. \tag{15}$$

Districts $i \in [S^o(\underline{V}), S^o(\overline{V})]$ are the competitive districts. They are divided into *Democrat-leaning districts* ($i \in [S^o(\underline{V}), \pi_D + \frac{\pi_I}{2})$) and *Republican-leaning districts* ($i \in [\pi_D + \frac{\pi_I}{2}, S^o(\overline{V})]$). The Democrat-leaning districts are populated by only Democrats and Independents, with the fraction of Independents varying from $\tau/(\tau + \varepsilon)$ to 1. These districts all elect a Democrat candidate when the majority of Independents prefer the Democrats; i.e., when $V \geq \pi_D + \frac{\pi_I}{2}$. However, they differ in their critical vote thresholds because they contain different fractions of Independents. Thus, the fraction of these districts electing Democrats varies smoothly as the aggregate Democrat vote share increases from $\underline{V}$ to $\pi_D + \frac{\pi_I}{2}$. Similarly, the Republican-leaning districts are populated by only Republicans and Independents, with the fraction of Independents varying from 1 to $\tau/(\tau + \varepsilon)$. These districts all elect Republicans when the majority of Independents prefer Republicans, but

the fraction electing a Republican varies smoothly as the aggregate vote share increases from $\pi_D + \frac{\pi_I}{2}$ to $\overline{V}$.

Districtings of the form described in (13) are extreme in the sense that the competitive districts have no voters of one type.[19] It is reasonable to object that such districts are unlikely to be practically feasible when account is taken of geographic constraints.[20] In this regard, it is important to note that the optimal seat-vote curve can typically be implemented with much more "straightforward" districtings. To illustrate, consider the class of districtings in which the fraction of Independents is constant across districts. In this class, all that varies across districts is the fraction of Democrats and Republicans. Then, we have the following result:

Proposition 3: *The optimal seat-vote curve is implementable with a districting of the form*

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\underline{\pi}_D, \pi_I) & \text{if } i \in [0, S^o(\underline{V})) \\ (\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau}, \pi_I) & \text{if } i \in [S^o(\underline{V}), S^o(\overline{V})] \\ (\overline{\pi}_D, \pi_I) & \text{if } i \in (S^o(\overline{V}), 1] \end{cases} \tag{16}$$

if and only if

$$\frac{\pi_I \varepsilon (1 - \pi_I) + (\frac{1}{2} - \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau}))\pi_I(\frac{1}{2} - \varepsilon)}{\frac{1}{2} + \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau})} \leq \min\{\pi_D, \pi_R\}. \tag{17}$$

Proof: See Appendix

The competitive districts in districtings of the form described in Proposition 3 can still be divided into Democrat-leaning districts ($i \in [S^o(\underline{V}), \pi_D + \frac{\pi_I}{2}]$) and Republican-leaning districts ($i \in [\pi_D + \frac{\pi_I}{2}, S^o(\overline{V})]$). However, all districts contain all three types of voters. The Democrat-leaning districts just have a greater fraction of Democrats than Republicans, with the ratio of Democrats to Republicans varying from $[1 - \pi_I(\frac{\tau - \varepsilon}{\tau})]/[1 - \pi_I(\frac{\tau + \varepsilon}{\tau})]$ to 1. The Republican-leaning districts have a greater fraction of Republicans, with the ratio of Democrats to Republicans varying from 1 to $[1 - \pi_I(\frac{\tau + \varepsilon}{\tau})]/[1 - \pi_I(\frac{\tau - \varepsilon}{\tau})]$.

The important point to note is that the condition of Proposition 3 is not that much more restrictive than that of Proposition 2. As an illustration of this point, Figure 2 plots the sets of (π_D, π_I) pairs that satisfy the conditions of Propositions 2 and 3 under the assumption that

[19] It may also be noted that the safe Democrat and safe Republican seats are identical. However, this is inconsequential. All that matters is that these seats be safe. The seat-vote curve is unaffected by the exact margin of victory in these seats.

[20] Indeed, Shotts (2002) argues that geographic constraints necessitate that each district contain at least some minimum fraction of each type of voter.

$\varepsilon = 0.1$ and $\tau = 0.2$. The horizontal axis measures π_I and the vertical axis measures π_D. The entire triangular area represents the set of (π_D, π_I) pairs satisfying the condition of Proposition 2. The shaded area represents the set of pairs that satisfy the condition of Proposition 2 but not that of Proposition 3.

3.4 What happens when the optimal seat-vote curve is not implementable?

While the condition of Proposition 2 is permissive, it is worth understanding what the constrained optimal seat-vote curve looks like when it is not satisfied. In this case, the implementability constraint in problem (10) must bind. Solving for the constrained optimal seat-vote curve is an intricate problem because of the difficulty of accounting for the implementability constraint. Thus, we will simply provide a flavor of the solution and refer the reader to our working paper (Coate and Knight (2005)) for the complete details.

When the condition of Proposition 2 is not satisfied, there are three possibilities. First, the condition is violated for π_D and there are *not enough Democrats*. Second, it is violated for π_R and there are *not enough Republicans*. Finally, it is violated for both π_D and π_R and there are *not enough Democrats or Republicans*. The principles involved in the first and second case are identical and, in the third case, the constrained optimal seat-vote curve is just a combination of those emerging in the first and second cases. Thus, it will suffice to discuss the first case.

Figure 3 illustrates the constrained optimal seat-vote curve - denoted $S^*(V)$ - when there are not enough Democrats. Panel (a) illustrates the case in which π_D is less than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$ and Panel (b) the case in which π_D exceeds $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$. In either case, the constrained optimal seat-vote curve lies below the optimal seat-vote curve on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2})$ and equals it thereafter. What this tells us is when the median Independent favors the Republicans, it is not possible to elect enough Democrats to make the average ideology of the legislature equal to the population average. However, when the median Independent favors the Democrats there is no longer a problem, because Democrats can be elected from districts that are populated solely (or largely) by Independents.

In the case illustrated in Panel (a) the logic of the constrained optimum is to allocate the available Democrats to create as many safe Democrat districts as possible. These safe seats contain only Democrats and Independents and the fraction of Democrats is the minimal necessary to always give the Democrat candidate a majority. The jump in the seat-vote curve is created by

the presence of a group of Independent only districts. These districts elect a Democrat if and only if the majority of Independents favor the Democrats which is why the jump occurs at the vote share $\pi_D + \frac{\pi_I}{2}$.

In the case illustrated in Panel (b) the constrained optimum involves safe Democrat seats but some Democrats are also grouped with Independents to form competitive districts. This generates the convex portion of the seat-vote curve on the interval $[\underline{V}, \widehat{V})$. In addition, there are Independent only districts so that the seat-vote curve again has a jump at the vote share $\pi_D + \frac{\pi_I}{2}$. As π_D gets larger (holding constant π_I) the point at which the seat-vote curve flattens ($\widehat{V}$) moves to the right and, for sufficiently large π_D, equals $\pi_D + \frac{\pi_I}{2}$ and the flat spot disappears. Moreover, the convex portion of the seat-vote curve straightens out. The complex shape of the seat-vote curve in this case stems from an inherent non-convexity created by the implementability constraint in problem (10).

3.5 The welfare gain to socially optimal districting

Returning to the case in which the optimal seat-vote curve is implementable, what would be the welfare gain associated with implementation? Conveniently, the gain turns out to be proportional to the squared distance between the baseline and optimal seat-vote curves. To see this, note first that expected citizen welfare can be expressed as a function of the baseline and optimal seat-vote curves. Letting $EW^*(S(V))$ denote expected social welfare under the seat vote curve $S(V)$, we have:[21]

Lemma: *It is the case that*

$$EW^*(S(V)) = \beta - \gamma\{c + E[S(V)^2] - 2E[S(V)S^o(V)]\} \tag{18}$$

where c is a constant given by $c = \pi_D + \pi_I[1/4 + \varepsilon^2/3 + \tau^2/3]$.

Proof: See Coate and Knight (2006).

Using this formula to compute the welfare gain immediately establishes:

Proposition 4: *The welfare gain from socially optimal districting can be written as*

$$G = EW^*(S^o(V)) - EW^*(S(V)) = \gamma E[(S^o(V) - S(V))^2]. \tag{19}$$

[21] That is, $EW^*(S(V)) = \int_{\underline{V}}^{\overline{V}} W(S(V), m(V)) \frac{dV}{\overline{V} - \underline{V}}$.

Intuitively, the larger the distance from the optimal seat-vote curves, the larger are the welfare gains associated with socially optimal districting.

4 Empirical methodology

From an empirical perspective, we would like to know how seat-vote curves generated by actual legislative districting plans differ from optimal seat-vote curves. We would also like to know if the condition for implementability is satisfied. Finally, if this condition is satisfied, we would like to know the magnitude of the welfare gains from implementing the optimal seat-vote curve. This section presents our methodology for doing all this. In developing this, we assume that the researcher has estimates at the district level of the mean and standard deviation of the Democratic vote share under the districting plan in question. In addition, we assume that the analyst knows the statewide fraction of voters who identify as Independents.[22]

4.1 Estimating seat-vote curves

As explained in Section 2.3, seat-vote curves are determined by the range of possible statewide Democratic vote shares $[\underline{V}, \overline{V}]$ and the pattern of district-specific threshold vote levels ($V^*(i)$). We will show that both the range of vote levels and the vote thresholds can be expressed solely as a function of the means and standard deviations of the district-level and statewide Democratic vote share. The first step in establishing this is to provide expressions for these moments. Beginning with the district-specific moments, the mean and standard deviation of Democratic votes in district i, as expressed in equation (1), are

$$\begin{array}{c} \mu_i = E(V(i)) = \pi_D(i) + \frac{1}{2}\pi_I(i) = \frac{1}{2}[1 + \pi_D(i) - \pi_R(i)] \\ \sigma_i = \sqrt{Var(V(i))} = \sqrt{\frac{\pi_I(i)^2 var(m)}{4\tau^2}} = \frac{\pi_I(i)\varepsilon}{2\sqrt{3}\tau} \end{array} . \tag{20}$$

Observe that the standard deviation of the Democratic vote share is proportional to the fraction of Independents. It is increasing in the degree to which the median Independent shifts support between the two candidates from election to election (ε) but is decreasing in the diversity of preferences among Independents (τ).

[22] In the empirical application to follow, we estimate the moments by running panel data regressions that relate district voting returns to voter characteristics. Data on the statewide fraction of Independents is taken from survey data.

Turning to the statewide moments, we take cross-district averages of the district-specific means and standard deviations to obtain:

$$\begin{aligned} \mu = E(V) &= \pi_D + \tfrac{1}{2}\pi_I = \tfrac{1}{2}[1 + \pi_D - \pi_R] \\ \sigma = \sqrt{Var(V)} &= \sqrt{\tfrac{\pi_I^2 var(m)}{4\tau^2}} = \tfrac{\pi_I \varepsilon}{2\sqrt{3}\tau} \end{aligned} . \tag{21}$$

Using these state-wide moments, we can now write the maximum and minimum statewide Democratic vote shares as

$$\begin{aligned} \underline{V} &= V(1/2 + \varepsilon) = \mu - \sqrt{3}\sigma \\ \overline{V} &= V(1/2 - \varepsilon) = \mu + \sqrt{3}\sigma \end{aligned} . \tag{22}$$

Moreover, using the district-specific and statewide moments and the definition of $V^*(i)$ in equation (3), we can write the vote threshold for electing a Democratic candidate in district i as follows:

$$V^*(i) = \mu + \frac{\sigma}{\sigma_i}(1/2 - \mu_i). \tag{23}$$

Relabelling the districts so that

$$\mu + \frac{\sigma}{\sigma_{1/n}}(1/2 - \mu_{1/n}) \leq \mu + \frac{\sigma}{\sigma_{2/n}}(1/2 - \mu_{2/n}) \leq .... \leq \mu + \frac{\sigma}{\sigma_1}(1/2 - \mu_1),$$

the seat-vote curve is the function defined on the interval $[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]$ given by:

$$S(V) = \max\{i : \mu + \frac{\sigma}{\sigma_i}(1/2 - \mu_i) \leq V\}. \tag{24}$$

4.2 Estimating the optimal seat-vote curve

Using the fact that $\pi_D - \pi_R = 2\mu - 1$, we can re-write the optimal seat-vote curve described in Proposition 1 as follows:

$$S^o(V) = 1/2 + (2\mu - 1)(1/2 - \tau) + 2\tau(V - 1/2). \tag{25}$$

Observe that the optimal seat-vote curve and, in particular, its responsiveness and partisan bias parameters depend critically upon the diversity of preferences among Independents (τ). Even with information on the statewide standard deviation, this parameter τ cannot be identified separately from the underlying parameters π_I and ε (see equation (14)).[23] However, it is possible to identify

[23] Intuitively, large swings in the Democratic vote share within a state could be due to a large fraction of Independents (π_I), large swings in the preferences of the median Independent (ε), or a tight distribution of ideology among Independents (τ), in which case relatively small swings in the preferences of the median Independent translate into relatively large swings in voting outcomes.

the ratio $\alpha = \varepsilon/\tau$, with data on the statewide fraction of Independents (π_I) and the statewide standard deviation of the Democratic vote share (σ) as follows:

$$\alpha \equiv \frac{\varepsilon}{\tau} = \frac{2\sqrt{3}\sigma}{\pi_I}. \tag{26}$$

Further, using the theoretical restriction on the sum of these preference parameters $\varepsilon + \tau \leq 1/2$, we can place an upper bound on optimal responsiveness:

$$2\tau \leq \frac{1}{(1+\alpha)}. \tag{27}$$

In the baseline analysis of the empirical application to follow, we assume that optimal responsiveness equals this upper bound. In addition, as a robustness check, we allow the optimal responsiveness to fall in a range below this upper bound.

4.3 Verifying the condition for implementation

The condition for implementability presented in Proposition 2 cannot be verified directly without information on the underlying preference parameters (ε, τ). As just noted however, with outside information on the fraction of Independents, we can identify the ratio $\alpha = \varepsilon/\tau$. We can use information on this ratio to place an upper bound on the coefficient associated with the implementability of the optimal seat-vote curve. In particular, as noted in Section 3.2, we know that

$$(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})) \leq \alpha/2, \tag{28}$$

and this implies that a sufficient condition for implementability is:

$$\pi_I \leq \frac{2\min(\pi_D, \pi_R)}{\alpha}. \tag{29}$$

Substituting in the expression for α from equation (26) and using the fact that $\mu = \pi_D + \pi_I/2$, this sufficient condition can be re-written as

$$\pi_I \leq 2\min(\mu - \sqrt{3}\sigma, 1 - \mu - \sqrt{3}\sigma). \tag{30}$$

Thus, the fraction of Independents must be below a critical value, the calculation of which only requires information on the statewide mean and standard deviation of the Democratic vote share.

4.4 Estimating the welfare gains from socially optimal districting

Given that we only observe voting outcomes, which do not reveal the intensity of voter preferences for one party over another, the parameters of the surplus expression (β, γ) in equation (7) are not identified in the empirical analysis. However, we can use the theoretical restriction on the ratio of these parameters in order to calculate the range of proportionate welfare gains. To this end, first note that the percentage increase in aggregate welfare from socially optimal districting can be written as follows:

$$\Delta G = \frac{EW^*(S^o(V)) - EW^*(S(V))}{EW^*(S(V))}. \tag{31}$$

Then, using the expressions from equations (18) and (19) and dividing through by β, we have that:

$$\Delta G(\frac{\gamma}{\beta}) = \frac{\frac{\gamma}{\beta} E[(S^o(V) - S(V))^2]}{1 - \frac{\gamma}{\beta}\{c + E[S(V)^2] - 2E[S(V)S^o(V)]\}}. \tag{32}$$

Recall that the ratio γ/β is the fraction of the surplus a partisan obtains from having a perfectly congruent legislature that is dissipated by having a legislature composed entirely of the opposition party. When parties are not that polarized in terms of their underlying ideologies or when the legislature is responsible for choosing only policies on which there is little disagreement across ideologies (for example, spending on public safety and highway maintenance) this ratio may be close to zero. When parties are polarized and are choosing policies on which there is strong ideological disagreement (such as the level of transfer payments for the poor or the regulation of abortion), this ratio may be close to one. In the former case, districting is not very important, while in the latter case it is crucial to citizen welfare. Using the restriction that $\gamma/\beta \in [0, 1]$, we can thus bound these proportionate welfare gains as follows:

$$0 \leq \Delta G(\frac{\gamma}{\beta}) \leq \Delta G(1). \tag{33}$$

Because the upper bound will only be relevant for legislatures in states in which parties are polarized and which choose policies on which there is strong disagreement, we will provide welfare calculations for different values of this key ratio (γ/β) in the empirical application to follow.

We next turn to the measurement of this welfare gain. Inserting equation (25) into equation (32), we have that

$$\Delta G(\frac{\gamma}{\beta}) = \frac{(\frac{\gamma}{\beta})\{(\zeta^o)^2 + 2\zeta^o r^o \mu + (r^o)^2(\mu^2+\sigma^2) - 2E[(\zeta^o + r^o V)S(V)] + E[S(V)^2]\}}{1 - (\frac{\gamma}{\beta})\{c + E[S(V)^2] - 2\zeta^o E[S(V)] - 2r^o E[VS(V)]\}} \quad (34)$$

where $r^o = 2\tau$ represents optimal responsiveness and $\zeta^o = \mu(1-2\tau)$ represents the vertical intercept of the optimal seat-vote curve. Notice that we can express the constant c as $c = \mu + \frac{2\sqrt{3}\tau\sigma}{\varepsilon}[\varepsilon^2/3 + \tau^2/3 - 1/4]$, and hence, given a particular value of the ratio γ/β and the parameter τ, this expression can be evaluated by computing three moments associated with the seat-vote curve: $E[S(V)]$, $E[S(V)^2]$, and $E[VS(V)]$. Finally, we have developed expressions relating these moments to the district-specific vote thresholds ($V^*(i)$), which, as noted above, can in turn be related to the moments of the Democratic vote share; we refer readers to Coate and Knight (2006) for the exact form of these expressions.

5 Application to U.S. state legislatures

In this section, we apply our methodology to analyze the districting plans used to elect U.S. state legislators during the 1990s districting period.[24] For consistency with the theoretical framework, we focus on states with single-member districts.[25] In addition, given the bicameral nature of state legislatures, we follow the existing empirical literature on redistricting and focus on elections to the Lower House. As shown in Table 1, we have complete data for 28 states, most of which adopted redistricting plans in 1992 and then again in 2002. For these states, there were five elections held under the 1990s districting plan: 1992, 1994, 1996, 1998, and 2000.[26] To estimate the moments of the Democratic vote shares under these districting plans, we use data from Ansolabehere and Snyder (2002) on state legislative election returns together with Census data on voter characteristics by state legislative district.[27] We also use state-level estimates of

[24] We prefer this setting over the U.S. House because, in federal redistricting, state officials control the redistricting process and each redistricting plan thus only partially contributes to the resulting allocation of national seats across parties in Congress. Redistricting plans for State legislatures are also controlled by State officials, and redistricting plans thus perfectly correspond to changes in seats in State legislatures.

[25] Some States elect multiple members from each district to State legislatures.

[26] States deviating from this pattern of elections include Virginia, which has elections in odd years and adopted redistricting plans in 1991 and 2001, and Colorado, whose district lines were redrawn in 1998 following litigation over the representation of minority groups in the State legislature.

[27] These data, which are published in Barone et al (1998), include the fraction of residents living in urban areas, the fraction living in suburban areas, household income, percent of residents with a college degree, percent over age 65, percent African-American, and percent Hispanic.

the fraction of Independent voters derived from annual New York Times surveys in which voters are asked to self-identify as Republican, Democrat, or Independent.[28]

5.1 Estimation of moments

There are a number of possible methods for estimating the moments of the voting distribution under a districting plan, the appropriateness of which may vary from application to application. With a sufficiently long panel, for example, moments in the key expressions above could simply be replaced with their analogous sample moments. However, because redistricting typically occurs every ten years in the U.S. and elections every two years, in our application we have five observations per district at most. In addition, this approach is problematic given that sample moments cannot be calculated for districts with uncontested elections, which occur frequently in state legislative elections. As an alternative, we use an econometric model for estimating the moments as a function of the characteristics of voters residing in the district. This approach circumvents the short panel problem by modeling the moments in terms of a small number of parameters and thus does not require the estimation of two parameters, or moments, per district. It also circumvents the problem of uncontested districts as it allows the analyst to use characteristics of voters residing in the districts in order to predict these two moments. Finally, as will be seen below, the use of an econometric model allows the researcher to compute confidence intervals around key measures of our welfare analysis, such as the gains to socially optimal districting.

Before providing a specific econometric formulation for these moments, it is instructive to note that, using equations (1) and (20), the Democratic vote share in district i can be written as a linearly separable function of the district-specific mean and variance along with a shock to the preferences of Independent voters:

$$V(i) = \mu_i + \sigma_i w \tag{35}$$

where $w = \frac{\sqrt{3}}{\epsilon}(1/2 - m)$ is distributed uniformly with mean zero and variance equal to 1. This formulation then naturally leads to the following specification, in which the two moments are related to voter characteristics:

[28] These data were downloaded from the website http://php.indiana.edu/~wright1/cbs7603_pct.zip. In order to compute the time-invariant fraction of Independents for each State, we take averages across the years listed in Table 1. We choose this data source over others, such as the National Election Survey, due to the large sample size, an important consideration when computing state-specific statistics. During the 1992-2000 period as a whole, these surveys included over 200,000 respondents nationally.

$$\begin{aligned} \mu_i &= X_i'\theta_1 + \sigma_1\xi_i \\ \sigma_i^2 &= \exp(X_i'\theta_2 + \sigma_2\upsilon_i) \end{aligned}, \tag{36}$$

where X_i denotes a vector of observed voter characteristics, (θ_1, θ_2) denotes parameters associated with these observed characteristics, (ξ_i,υ_i) denote district-specific random effects, assumed to be distributed standard normal, and (σ_1,σ_2) are parameters associated with the random effects. Reflecting our use of panel data, we next introduce a time dimension (t) and, inserting the above parameterizations for the two moments into equation (35), we obtain the following random effects model with heteroskedasticity:

$$\begin{aligned} V_t(i) &= X_i'\theta_1 + \sigma_1\xi_i + u_{it} \\ \ln(u_{it}^2) &= X_i'\theta_2 + \sigma_2\upsilon_i + \ln(w_t^2) \end{aligned}. \tag{37}$$

Before describing the estimation procedure associated with the econometric model in equation (37), it is important to note two sources of uncertainty facing the analyst in estimating the moments in equation (36). First, the analyst does not observe the true parameters $(\theta_1, \theta_2,\sigma_1,\sigma_2)$ and must instead use estimated parameters. Second, the random effects (ξ_i,υ_i) are unobserved by the econometrician. We deal with both of these issues of uncertainty via a bootstrapping simulation procedure. In particular, for each replication $r = 1, 2, ..., 100$, an associated sample of size N is drawn with replacement from our dataset of N districts, where N is the number of districts with at least one contested election in our dataset.[29] Given that we draw with replacement, a given district in our dataset may be not represented, represented once, or represented multiple times in the sample associated with a given replication r. For each of these 100 samples, we then estimate the parameters of equation (37) via a standard two-step approach. First, we estimate a subset of the parameters (θ_1^r, σ_1^r) using a random-effects panel data regression of votes in district i $(V_t(i))$ on observed voter characteristics in district i (X_i). In the second step, we regress the log of the squared residual obtained from the first step on observed district characteristics and obtains estimates of the remaining parameters of interest (θ_2^r, σ_2^r).[30] Taken across replications, we then have an entire distribution of parameter estimates $(\theta_1^r, \theta_2^r, \sigma_1^r, \sigma_2^r)_{r=1}^{r=100}$.

[29] Our dataset consists of 2,973 state legislative districts, of which 2,703 had at least one contested election during the relevant time period.

[30] This bootstrap approach also corrects for the additional uncertainty associated with the second step regression including a generated variable on the left-hand side.

Table 2 provides our estimates of the parameters θ_1 and θ_2 using the original dataset along with the bootstrap standard errors.[31] In addition to the district characteristics reported in Table 2, we also included a set of state dummy variables in both equations, thereby allowing two districts with identical observable characteristics but in different states to have different voting patterns. As shown in the first column, the mean vote share for the Democratic Party (μ_i), is increasing in the percent urban and suburban (both of these are relative to the omitted category - percent rural), percent with a college degree, percent over age 65, percent African-American, and percent Hispanic (both of these are relative to the omitted category - percent white) but is decreasing in household income. As shown in the second column, the variance is decreasing in household income and in the percent African-American.

The parameter estimates $(\theta_1^r, \theta_2^r, \sigma_1^r, \sigma_2^r)$ associated with replication r are then used to compute the key moments of the voting distribution for every district in our original dataset, including those districts with only uncontested elections. Note that even with the parameter estimates from a given replication r and data on district characteristics (X_i), we do not learn the district-specific mean and variance of the voting distribution, as expressed in equation (36), because the random effects (ξ_i,υ_i) are unobserved. To get around this problem, random effects (ξ_i^r,υ_i^r) are drawn from the standard normal distribution for each replication r and for each district i in our dataset.[32] The district-specific moments (μ_i^r, σ_i^r) associated with replication r are then calculated as follows:

$$\begin{aligned} \mu_i^r &= X_i'\theta_1^r + \sigma_1^r \xi_i^r \\ \sigma_i^r &= \sqrt{\exp(X_i'\theta_2^r + \sigma_2^r \upsilon_i^r)} \end{aligned} . \tag{38}$$

As should be clear from equation (38), our simulation approach accounts for two forms of uncertainty described above: the use of estimated parameters along with the inability of the econometrician to observe the exact realization of the random effects.

[31] Alternatively, one could report the mean of the parameters across replications. These estimates are very similar to those reported in Table 2 and are available from the authors upon request.

[32] An alternative solution to this problem would be to simply "shut down" this unobserved component and use only the observed component as our estimate of the moments. This is, after all, the central tendency (the mean and median) of the distribution of moments. The difficulty with this approach, however, is that it will tend to understate the degree of cross-district heterogeneity, which in turn may tend to overstate the responsiveness of the seat-vote curve. To see this, consider the extreme case in which the true moments are heterogeneous across districts but in which the observed voter characteristics have no explanatory power in the regressions. Then, if the unobserved component is shut down, the estimated seat-vote curve will be that associated with identical districting and will thus jump from 0 to 1 at V=1/2, and all districts will thus be considered competitive. Of course, our voter characteristics do have explanatory power but the more general lesson still holds: ignoring the unobserved component tends to understate the degree of heterogeneity across districts.

With estimates of district-specific moments in hand, corresponding statewide moments (μ^r, σ^r) associated with replication r are then calculated by averaging across the district-specific moments. Key objects of interest, such as the vote threshold for electing a Democratic candidate, as expressed in equation (23), can then be calculated for each district i and replication r, and it follows that a state-specific seat-vote curve $S^r(V)$ can be calculated for each replication r. Finally, aggregating across all replications $r = 1, 2, ..100$, key aspects of the distribution of seat-vote curves $(S^r(V))_{r=1}^{r=100}$, such as average responsiveness, median responsiveness, and the 90-percent confidence interval for responsiveness, are calculated for each state.[33]

5.2 Seat-vote curves

Given that a plot of all replications would be cumbersome, we begin by presenting the seat-vote curves from a particular replication, that associated with the median welfare gain from socially optimal districting (as expressed in equation (34)). Note that these seat-vote curves are presented primarily for illustration, and we will subsequently present statistics pertaining to the entire distribution of seat-vote curves across replications. As shown in Figure 4, the range of possible statewide support for the Democratic Party seems reasonable, at least in this particular replication. For example, in New York, a heavily Democratic state, the Party receives support on the range [0.56,0.70], while in Utah, a heavily Republican state, the Democrats receive support on the range [0.38,0.50]. Notice also that the seat-vote curve is close to linear in some states, such as Maine and Pennsylvania, while it has important non-linearities in other states, such as Delaware and Virginia.

As shown in column 1 of Table 3, the responsiveness associated with these estimated seat-vote curves exceeds two in every state, and, in some states, exceeds three.[34] Findings of significant

[33] This simulation approach is similar to methods used in Gelman and King (1990, 1994). In Gelman and King (1994), for example, votes (V_i) are assumed to be related to observable candidate characteristics (X_i), such as incumbency, and the authors estimate the parameters of the regression equation $V_i = X_i\beta + u_i$, where u_i is unobserved and normally distributed; i.e. $u_i \sim N(0, \sigma^2)$. Importantly, the variance is assumed to be constant across districts. With the estimated parameters in hand, the authors then simulate the model, and the implied seat-vote curve, by drawing from the distribution of district-specific votes for the Democrat (i.e., the u_i's). This similarity not withstanding, there are three key differences between our approach and that of Gelman and King. First, as noted in the introduction, we begin with a theoretical model and focus on providing micro foundations for the measurement of seat-vote curves. Second, their approach assumes that this variance of the vote generating process is constant across districts; in the context of our model, this assumption would require that all districts have an equal fraction of Independent voters, and we thus allow this variance to be heterogeneous across districts. Third, in explaining differences in voting patterns across districts, Gelman and King (1994) rely on observable candidate characteristics whereas, given our theoretical assumption of homogeneous candidates, we rely on observable voter characteristics.

[34] This responsiveness measure is obtained by computing the slope of the linear seat-vote curve that best ap-

responsiveness are quite common in the existing literature, which has focused on a responsiveness of three, a finding that has become known as the "Cube Law" (King and Browning, 1987). Given its prominence in the existing literature, we next report the partisan bias associated with the estimated seat-vote curve, as measured by $S(1/2) - 1/2$. Notice that in the four states for which $V = 1/2$ does not lie in the range of possible Democratic vote shares, this measure cannot be computed. Although this measure is based upon only a single point on the seat-vote curve, it is nonetheless interesting, as shown in column 4, that the seat-vote curve is biased towards the Republicans in 18 states but towards the Democrats in only 4 states; the seat-vote curve is unbiased in Montana and Oregon. The cross-state average bias of -3% implies that, when voters are equally split, Republicans would secure 53% of the seats on average, relative to 47% for Democrats, and would thus hold a significant advantage of 6% in the legislature.

5.3 Optimal seat-vote curves

For comparison purposes, Figure 4 also includes the optimal seat-vote curves from the replication associated with the median welfare gain for each state. These are plotted under the assumption that optimal responsiveness, 2τ, is at its maximal level $1/(1+\alpha)$ in each state. It is apparent from Figure 4, that the actual seat-vote curves are overly responsive in all cases, suggesting that districting plans used to elect U.S. state legislators during the 1990s created too few safe seats. Column 2 of Table 3 reports the responsiveness of the optimal seat-vote curve. As noted in the theoretical section, optimal responsiveness is always below one, while actual responsiveness substantially exceeds one in all cases. That is, as shown in column 3, the difference between actual and optimal responsiveness is close to 2 in most states.

Column 5 of Table 3 reports the partisan bias associated with the optimal seat-vote curve, which is defined as $S^o(1/2) - 1/2$. While the estimated seat-vote curve was shown to be biased towards Republicans in this particular replication, the optimal seat-vote curve tends to be biased towards the Democrats. Thus, the bias towards the Republicans exhibited in the actual seat-vote curves cannot generally be justified as optimal. As noted above, an important objection to this comparison of actual and optimal partisan bias is that it just tells us about the properties of the curves at the vote share $V = 1/2$. For a more global comparison, we computed the expected

proximates the estimated seat-vote curve. To be more precise, define the *linearized estimated seat-vote curve* to be $S^l(V) = 1/2 + b + r(V - 1/2)$, where b and r are chosen in order to minimize the expected square distance between the estimated and the linearized estimated seat-vote curves. That is: $r = \text{cov}[V, S(V)]/\text{var}(V)$ and $b = E[S(V)] - 1/2 - r[E(V) - 1/2]$. Then our measure of responsiveness is r.

Democratic seat share under the estimated and optimal seat-vote curves ($E(S(V))$ and $E(S^o(V))$). Column 7 of Table 3 reports the difference in this expected seat-share. When this difference is positive, the expected Democratic seat share is higher than optimal. The interesting thing to note is that, in this expected seat sense, there appears to be no obvious bias towards Republicans. If anything, this alternative bias measure suggests that districting systems are overly biased towards Democrats as this measure is positive in over one-half of the states. This makes us hesitant to draw any strong conclusions concerning the general direction of bias.

The properties of seat-vote curves reported in Table 3 are based upon a single replication, that associated with the median welfare gain. Table 4 reports more general findings of our analysis via the properties of the entire distribution of replicated seat-vote curves. The first three columns report the mean difference in responsiveness between estimated and optimal seat-vote curves across replications, the median difference in responsiveness across replications, and the 90 percent confidence interval for the difference across replications. This confidence interval is computed by ranking our measures of responsiveness across replications and then choosing the 5th and 95th percentile of that distribution. As shown, the finding that the estimated seat-vote curve is overly responsive in Table 3 is a robust finding as the 90 percent confidence interval for the difference in responsiveness is positive in all states. The findings regarding partisan bias, by contrast, are more mixed. As shown in columns 4 and 5, the mean and median difference in bias is negative in most cases, reflecting the previous finding that estimated seat-vote curves tend to be overly biased in favor of the Republicans. The confidence interval, however, includes zero for all except nine states, and thus our finding that the actual seat-vote curve is overly biased in favor of Republicans should be interpreted with caution. Similarly, as shown in the final three columns, no definitive conclusions can be drawn in a statistical sense regarding the difference in expected seats.

5.4 Verifying the condition for implementation

As shown in Table 5, the condition for implementation is satisfied with probability one in every state. That is, in every replication, the fraction of Independents was below the maximal level described in equation (30). The state closest to not satisfying the requirements is Rhode Island, which is reported to have 51% Independents, just slightly below the cross-replication average maximal level of 58%. In no replications, however, did this maximum level fall below the reported 51% share of Independents. In summary, these results demonstrate that the condition for imple-

mentability of the optimal seat-vote curve is indeed permissive, being satisfied in all of 28 states included in our analysis and by a large margin in all cases except for Rhode Island.[35]

5.5 Welfare gains

Given that the optimal seat-vote curve is implementable in all states and across all replications, it is interesting to measure the welfare gains associated with socially optimal districting. To begin with, we compute the percentage welfare gains under the assumptions that the ratio γ/β is at its maximal level (i.e., 1) and that optimal responsiveness is at its maximal level in each state (i.e., $1/(1+\alpha)$). As shown in Table 6, the percentage welfare gains to socially optimal districting are relatively small; averaged across all states, these median and average gains are reported at 1.81% and 1.88%, respectively. There is, however, considerable variation across states with Rhode Island and South Carolina at the opposite extremes. A visual comparison of the seat-vote curves in Figure 4 supports these welfare calculations as the seat-vote curves are similar in states with low potential welfare gains but quite different in states with large potential welfare gains.[36] Regarding the precision of these estimates, the confidence intervals demonstrate that the upper bounds on these welfare gains are also quite low, ranging from 1.47% in South Carolina to 6.02% in Rhode Island.[37]

For a sense of how these results depend on the specific assumptions about the parameters γ/β and 2τ, Table 7 reports the average welfare gains, where the average is taken across both states and replications, associated with socially optimal districting arising under different parameter values. We allow the ratio γ/β to vary from 0.25 to 1 and allow 2τ to vary from $0.25/(1+\alpha)$ to $1/(1+\alpha)$

[35] Note that in evaluating the conditions for implementation, we have not used external information on the statewide fraction of Democrats and Republican. It is comforting to note, however, that the implied fraction of partisans from voting behavior are highly correlated with the fraction of voters self-reporting as partisans. Using the fact that the implied statewide fraction of Democrats and Republicans are given by $\pi_D = \mu - \pi_I/2$ and $\pi_R = 1 - \mu - \pi_I/2$, we calculate the cross-replication correlation between the implied and reported fraction of voters to be roughly 0.8 for both Republicans and Democrats.

[36] While a welfare comparison of the continuous optimal seat-vote curve and the discrete measured seat-vote curve is somewhat artificial, we use the continuous optimal seat-vote curve in order to apply the implementability condition from Proposition 2. To provide a sense of the error associated with this approximation, we have derived a discrete optimal seat-vote curve, which is a step-function approximation of the continuous optimal seat-vote curve. Welfare associated with this discrete optimal seat-vote curve is similar to welfare under the continuous optimal seat-vote curve, and the approximation error associated with the use of a continuous optimal seat-vote curve is thus small in practice. The small size of this error should not be surprising given that the discrete and continuous optimal seat-vote curve converge as the number of districts grows large. As shown in Table 1, states tend have a large number of legislative districts.

[37] Of course, if this surplus is itself very large, then these gains could be quite large in monetary terms. But without further assumptions on the underlying welfare parameters β and γ, these percentage gains in welfare cannot be converted into monetary terms.

in each state. The notation η refers to the numerator in the ratio $\eta/(1+\alpha)$, so that the case in which 2τ is set equal to $0.25/(1+\alpha)$ in each state corresponds to $\eta = 0.25$; the case in which 2τ is set equal to $0.5/(1+\alpha)$ in each state corresponds to $\eta = 0.5$; etc. As shown, holding the ratio γ/β constant, the welfare gains to socially optimal districting, averaged across all states, are uniformly increasing as the parameter η is reduced. This pattern reflects the fact that reductions in the parameter η are associated with reductions in optimal responsiveness, which, as shown previously, was already below the responsiveness associated with the estimated seat-vote curves. Holding the parameter η constant and reducing the ratio γ/β, we have that welfare is uniformly decreasing. As explained above, districting matters less for welfare as this ratio decreases, reflecting the fact that the conflict between citizens over the available policy choices is less severe.

The lesson to be drawn from Tables 6 and 7 is that the welfare gains from socially optimal districting are relatively small as a proportion of the total surplus generated by state legislatures. In principle, there may be two reasons for this. The first is that the districting plans that states actually implement are relatively close to optimal plans. The second is that, because of the diverse ideological make up of the U.S. states, aggregate welfare is relatively insensitive to varying districting plans. To get a feel for which of these views is correct, we computed the proportionate welfare gains that would arise from implementing the optimal seat-vote curve over the case of identical districting in which each district is a microcosm of the whole. The idea is that if these gains are large, then the second view cannot be correct. Table 8 reports the results under the same parametric assumptions that underlie Table 6 and they strongly suggest that the second view is not correct. Thus, it seems that the states are doing districting in a way that is generating seat-vote curves that are relatively close to optimal.[38]

A further interesting benchmark is the seat-vote curve that would be generated by a *proportional representation* electoral system (PR); that is, $S(V) = V$. As shown in Table 9, introducing PR would raise welfare in all states. Moreover, the gains from a movement from PR to the optimal seat-vote curve, are very small, ranging from an average of 0.02% in several states to 0.08% in Oklahoma under our baseline parameter values. These findings suggest that almost all the welfare gains associated with optimal districting could be achieved via PR. This result reflects several features of the seat-vote curves underlying these welfare calculations. First, the PR seat-vote

[38] Note that the seat-vote curve generated by identical districting equals 0 for $V < 1/2$ and 1 for $V \geq 1/2$. In the context of our model, this is also the seat-vote curve generated by at-large voting systems, providing an additional motivation for this benchmark.

curve is linear, a feature shared by the optimal, but not actual, seat-vote curves. Second, the PR seat-vote curve has responsiveness of 1, while responsiveness associated with the estimated seat-vote exceeds 1 in all cases. As noted previously, optimal responsiveness is less than 1.[39]

6 Conclusion

It will be clear to the reader that the exploration of socially optimal districting outlined in this paper is very much a first step. The simple model underlying the analysis ignores many factors that are important in real world discussions of districting. Nonetheless, the model provides a way of thinking about the problem that is consistent with a long tradition of empirical research on districting and both the theoretical and empirical analysis have yielded some interesting findings. We conclude the paper by summarizing what we see as the most significant of these.

The first notable result concerns the nature of the optimal relationship between seats and votes. Under the assumptions of our model, the optimal seat-vote curve is of the same linear form that has been estimated in some of the early empirical literature. In sharp contrast to what has been implicitly assumed in the districting literature, the optimal curve is biased and this bias favors the party with the largest partisan base. The responsiveness of the optimal curve depends on the degree of variation in the preferences of swing voters. This responsiveness is always less than one so that the change in seat share optimally damps the change in vote share.

The second interesting finding relates to the condition under which there exist districtings that generate the optimal seat-vote curve. This condition, which requires that the fraction of Independents be not "too large", is permissive and is satisfied in all states in our data set. Moreover, when this condition is satisfied, there will typically exist many different districtings that can generate the optimal seat-vote curve. These districtings involve safe seats for both parties and a range of competitive districts that vary smoothly in terms of their ideological mix. They look sufficiently straightforward to be achievable even when geographic constraints are accounted for.

[39] The reader may be concerned that our assumption that optimal responsiveness is at its upper bound may contribute in part to these small welfare gains associated with a movement from PR to optimal districting. Indeed, as the diversity parameter τ approaches 1/2, the optimal seat-vote curve converges to the PR seat-vote curve, as can be seen in equation (11). So there would obviously be no welfare gains to socially optimal districting in this case. To examine the role of this upper bound assumption, we experimented with alternative values of the parameter η. For example, when $\eta = 0.25$, optimal responsiveness equals $0.25/(1+\alpha)$ and is thus less than 1/4; even in this case, however, the welfare gains associated with a movement from PR to the optimal seat-vote curve are small, averaging 0.18% across states and replications. Thus, these results are robust to alternative parameter values and are not driven by the assumption that optimal responsiveness is at its upper bound.

This suggests that it may be reasonable to regard the optimal seat-vote curve as an attainable benchmark for districters in the U.S. states.

The third key result concerns the difference between optimal and actual seat-vote curves. We find that the seat-vote curves generated by the districting plans used in the 1990s by the states in our data set are overly responsive to changes in voter preferences. This result seems unlikely to be an artifact of our particular dataset. As we have previously noted, while optimal responsiveness must be less than 1, prior empirical evidence has typically estimated responsiveness well in excess of 1. In terms of partisan bias, we find no evidence of a direction of bias towards one party or the other when the measure of the difference in expected seats is employed.

The fourth finding relates to the potential welfare gains from socially optimal districting for the states in our data set. While there is significant variation across states, we find that these gains are quite small, at least relative to the overall surplus generated by state legislatures. We have argued that this is because states' districting plans were reasonably close to optimal rather than because aggregate welfare was insensitive to varying districting plans. We also find that almost all of the welfare gains from socially optimal districting could be realized by implementing a proportional representation system.

References

Alesina, Alberto and Howard Rosenthal, (1995), *Partisan Politics, Divided Government and the Economy*, Cambridge: Cambridge University Press.

Altman, Micah, Karin Mac Donald and Michael McDonald, (2005), "Pushbutton Gerrymanders? How Computing Has Changed Redistricting," in Thomas Mann and Bruce Cain (eds), *Party Lines*, Washington D.C.: Brookings Institution Press.

Ansolabehere, Stephen and James Snyder, (2002), "The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices: 1942-2000," *Election Law Journal*, 1, 315-338.

Barone, Michael, William Lilley III, and Laurance DeFranco, (1998), *State Legislative Elections: Voting Patterns and Demographics,* Washington D.C.: Congressional Quarterly.

Besley, Timothy and Anne Case, (2003), "Political Institutions and Policy Choices: Evidence from the United States," *Journal of Economic Literature*, 41, 7-73.

Besley, Timothy and Stephen Coate, (1998), "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 112(1), 85-114.

Besley, Timothy and Ian Preston, (2005), "Electoral Bias and Policy Choices: Theory and Evidence," mimeo, London School of Economics.

Butler, David, (1951), Appendix, in H. G. Nicholas (ed), *The British General Election of 1950*, London: Macmillan.

Cameron, Charles, David Epstein and Sharon O'Hallaran, (1996), "Do Majority-Minority Districts Maximize Substantive Black Representation in Congress," *American Political Science Review*, 90(4), 794-812.

Coate, Stephen, (2004), "Pareto-Improving Campaign Finance Policy," *American Economic Review*, 94, 628-655.

Coate, Stephen and Brian Knight, (2005), "Socially Optimal Districting," NBER Working Paper #11462.

Coate, Stephen and Brian Knight, (2006), "Socially Optimal Districting: An Empirical Investigation," NBER Working Paper #12313.

Degan, Arianna and Antonio Merlo, (2006), "Do Voters Vote Sincerely?" mimeo, University of Pennsylvania.

Epstein, David and Sharon O'Hallaran, (2004), "The 45% Solution: Racial Gerrymandering and Representative Democracy," mimeo, Columbia University.

Erikson, Robert, Gerald Wright and John McIver, (1993), *Statehouse Democracy: Public Opinion and Policy in the American States*, Cambridge: Cambridge University Press.

Fiorina, Morris, (1992), *Divided Government,* New York: Mcmillan.

Friedman, John and Richard Holden, (2006), "Optimal Gerrymandering: Sometimes Pack, But Never Crack," mimeo, Harvard University.

Galderisi, Peter, (ed), (2005), *Redistricting in the New Millennium*, Oxford: Lexington Books.

Gelman, Andrew and Gary King, (1990), "Estimating the Electoral Consequences of Legislative Redistricting," *Journal of the American Statistical Association*, 85, 274-282.

Gelman, Andrew and Gary King, (1994), "Enhancing Democracy through Legislative Redistricting," *American Political Science Review*, 88(3), 541-559.

Gilligan, Thomas and John Matsusaka, (1999), "Structural Constraints on Partisan Bias under the Efficient Gerrymander," *Public Choice*, 100, 65-84.

Gilligan, Thomas and John Matsusaka, (2005), "Public Choice Principles of Redistricting," mimeo, University of Southern California.

King, Gary, (1989), "Representation through Legislative Redistricting: A Stochastic Model," *American Journal of Political Science,* 33(4), 787-824.

King, Gary and Robert Browning, (1987), "Democratic Representation and Partisan Bias in Congressional Elections," *American Political Science Review*, 81(4), 1251-1273.

Lee, David, Enrico Moretti, and Matthew Butler, (2004), "Do Voters Affect or Elect Policies? Evidence from the U.S. House," *Quarterly Journal of Economics*, 119(3), 807-859.

Lizzeri, Alessandro and Nicola Persico, (2001), "The Provision of Public Goods under Alternative Electoral Incentives," *American Economic Review*, 91, 225-339.

Lockwood, Ben, (2002), "Distributive Politics and the Costs of Centralization," *Review of Economic Studies*, 69, 313-337.

Mann, Thomas and Bruce Cain, (eds), (2005), *Party Lines*, Washington D.C.: Brookings Institution Press.

Maskin, Eric and Jean Tirole, (2004), "The Politician and the Judge: Accountability in Government," *American Economic Review*, 94, 1034-1054.

Matsusaka, John and Nolan McCarty, (2001), "Political Resource Allocation: Benefits and Costs of Voter Initiatives," *Journal of Law, Economics and Organization*, 17, 413-448.

Myerson, Roger, (1999), "Theoretical Comparisons of Electoral Systems," *European Economic Review*, 43, 671-697.

Osborne, Martin and Al Slivinski, (1996), "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, 111(1), 65-96.

Owen, Guillermo and Bernard Grofman, (1988), "Optimal Partisan Gerrymandering," *Political Geography Quarterly*, 1988, 7(1), 5-22.

Persson, Torsten, Gerard Roland and Guido Tabellini, (2000), "Comparative Politics and Public Finance," *Journal of Political Economy*, 108, 1121-1161.

Persson, Torsten and Guido Tabellini, (2004), "Constitutions and Economic Policy," *Journal of Economic Perspectives*, 18, 75-98.

Prat, Andrea, (2002), "Campaign Advertising and Voter Welfare," *Review of Economic Studies*, 69, 999-1018.

Sherstyuk, Katerina, (1998), "How to Gerrymander: A Formal Analysis," *Public Choice*, 95, 27-49.

Shotts, Kenneth, (2001), "The Effect of Majority-Minority Mandates on Partisan Gerrymandering," *American Journal of Political Science*, 45, 120-135.

Shotts, Kenneth, (2002), "Gerrymandering, Legislative Composition, and National Policy Outcomes," *American Journal of Political Science*, 46, 398-414.

Tufte, Edward, (1973), "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review*, 67(2), 540-554.

7 Appendix

7.1 Proof of Proposition 2

In developing the conditions under which the optimal seat-vote curve is implementable it is more convenient to work with the inverse optimal seat-vote curve than the optimal seat-vote curve. In general, an *inverse seat-vote curve* is described by a triple $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ where $\underline{i}$ and $\overline{i}$ are scalars satisfying $0 \leq \underline{i} \leq \overline{i} \leq 1$ and $V^*(\cdot)$ is a non-decreasing function defined on $[\underline{i}, \overline{i}]$ with range $[\underline{V}, \overline{V}]$. The interpretation is that $\underline{i}$ is the fraction of districts that are safe Democrat; $1 - \overline{i}$ the fraction that are safe Republican; and $V^*(i)$ is the critical vote threshold for competitive district $i \in [\underline{i}, \overline{i}]$. Given a seat vote curve $S(V)$ we form its inverse in the following way: $\underline{i}$ is just $S(\underline{V})$; $\overline{i}$ is $S(\overline{V})$ and for all $i \in [\underline{i}, \overline{i}]$, $V^*(i)$ is such that $S(V) = i$. Thus, given the optimal seat-vote curve described in Proposition 1, the optimal inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(i)\}$ is given by

$$\underline{i}_o = \pi_D + \pi_I(1/2 - \varepsilon), \tag{39}$$

$$\overline{i}_o = \pi_D + \pi_I(1/2 + \varepsilon), \tag{40}$$

and for all $i \in [\underline{i}_o, \overline{i}_o]$

$$V_o^*(i) = \frac{[i - (\pi_D + \frac{\pi_I}{2})(1 - 2\tau)]}{2\tau}. \tag{41}$$

We will need the following definitions for the proof. A *districting* is a description of the fractions of voter types in each district $\{(\pi_D(i), \pi_I(i)) : i \in [0,1]\}$ where for all i, $(\pi_D(i), \pi_I(i))$ belongs to the two dimensional unit simplex Δ_+^2. (We omit the fraction of Republicans since $\pi_R(i) = 1 - \pi_D(i) - \pi_I(i)$.) A districting $\{(\pi_D(i), \pi_I(i)) : i \in [0,1]\}$ is *feasible* if $\int_0^1 \pi_I(i)di = \pi_I$ and $\int_0^1 \pi_D(i)di = \pi_D$. A districting $\{(\pi_D(i), \pi_I(i)) : i \in [0,1]\}$ *generates* the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ if: (i) $\pi_D + \pi_I[\frac{1/2 - \pi_D(i)}{\pi_I(i)}] \leq \underline{V}$ for all $i \in [0, \underline{i})$, (ii) $\pi_D + \pi_I[\frac{1/2 - \pi_D(i)}{\pi_I(i)}] \geq \overline{V}$ for all $i \in (\overline{i}, 1]$, and, (iii) $\pi_D + \pi_I[\frac{1/2 - \pi_D(i)}{\pi_I(i)}] = V^*(i)$ for all $i \in [\underline{i}, \overline{i}]$. Requirement (i) is that districts $i \in [0, \underline{i})$ are safe Democrat seats and requirement (ii) is that districts $i \in (\overline{i}, 1]$ are safe Republican seats. Requirement (iii) is that competitive district $i \in [\underline{i}, \overline{i}]$ has a critical vote threshold just equal to $V^*(i)$. A seat-vote curve is *implementable* if there exists a feasible districting that generates its associated inverse seat-vote curve.

We want to know if the optimal seat-vote curve $S^o(V)$ is implementable. We begin by describing the districtings that can generate the optimal inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(i)\}$. In

describing this set, there is no loss of generality in assuming that the safe Democrat and Republican districts are identical. Thus, we may assume that $(\pi_D(i), \pi_I(i)) = (\underline{\pi}_D, \underline{\pi}_I)$ for all $i \in [0, \underline{i}_o)$ and $(\pi_D(i), \pi_I(i)) = (\overline{\pi}_D, \overline{\pi}_I)$ for all $i \in (\overline{i}_o, 1]$ where $(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I) \in \Delta^2_+$.[40] Using the definitions of $\underline{V}$ and $\overline{V}$ (see (2)), requirements (i) and (ii) from above imply that

$$\underline{\pi}_D + \underline{\pi}_I(\frac{\tau - \varepsilon}{2\tau}) \geq \frac{1}{2} \tag{42}$$

and

$$\overline{\pi}_D + \overline{\pi}_I(\frac{\tau + \varepsilon}{2\tau}) \leq \frac{1}{2}. \tag{43}$$

These inequalities reflect the fact that the minimum and maximum fraction of Independents voting Democrat are, respectively, $\frac{\tau-\varepsilon}{2\tau}$ and $\frac{\tau+\varepsilon}{2\tau}$.

In the competitive districts $[\underline{i}_o, \overline{i}_o]$, requirement (iii) ties down what the function $\pi_D(i)$ must look like over the interval $[\underline{i}, \overline{i}]$ given any choice of the function $\pi_I(i)$. Specifically, $\pi_D(i) = f(\pi_I(i), V_o^*(i))$ where

$$f(x, y) = \frac{1}{2} - \frac{x}{\pi_I}(y - \pi_D). \tag{44}$$

In addition, we must have that $(\pi_I(i), f(\pi_I(i), V_o^*(i))) \in \Delta^2_+$ for all $i \in [\underline{i}_o, \overline{i}_o]$. This constraint amounts to the requirement that

$$\pi_I(i) \in [0, \min\{\frac{\pi_I}{2(V_o^*(i) - \pi_D)}; \frac{\pi_I}{2(\pi_I + \pi_D - V_o^*(i))}\}]. \tag{45}$$

Notice that $V_o^*(i) - \pi_D$ is less than $\pi_I + \pi_D - V_o^*(i)$ if and only if $V_o^*(i)$ is less than $\frac{\pi_I}{2} + \pi_D$. Letting $\widehat{i}_o = \frac{\pi_I}{2} + \pi_D$, it is the case that $V_o^*(i) < \frac{\pi_I}{2} + \pi_D$ for all $i \in [\underline{i}_o, \widehat{i}_o)$ and $V_o^*(i) \geq \frac{\pi_I}{2} + \pi_D$ for all $i \in (\widehat{i}_o, \overline{i}_o]$. Thus, we can write the constraint as

$$\pi_I(i) \in \begin{cases} [0, \frac{\pi_I}{2(\pi_I + \pi_D - V_o^*(i))}] & \text{if } i < \widehat{i}_o \\ [0, \frac{\pi_I}{2(V_o^*(i) - \pi_D)}] & \text{otherwise} \end{cases}. \tag{46}$$

We conclude from this that the districtings that generate the optimal inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(i)\}$ can be described by the set of all $\{(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I), \pi_I(i)\}$ such that $(\underline{\pi}_D, \underline{\pi}_I)$ and $(\overline{\pi}_D, \overline{\pi}_I)$ belong to Δ^2_+ and satisfy (42) and (43) and $\pi_I(i)$ satisfies (46) for all $i \in [\underline{i}_o, \overline{i}_o]$. We call this the *set of generating districtings* and denote it by Φ. The question of implementability is

[40] For example, if $(\pi_D(i), \pi_I(i))$ varied over the safe Democrat seats $i \in [0, \underline{i}_o)$, then we could create a districting with identical safe Democrat districts that used exactly the same fractions of voter types in the safe Democrat districts by letting $(\pi_D(i), \pi_I(i)) = (\int_0^{\underline{i}_o} \pi_D(i) \frac{di}{\underline{i}_o}, \int_0^{\underline{i}_o} \pi_I(i) \frac{di}{\underline{i}_o})$ for all $i \in [0, \underline{i}_o)$.

whether there exists a districting in this set which is feasible; i.e., which satisfies

$$\underline{i}_o\underline{\pi}_I + (1-\bar{i}_o)\overline{\pi}_I + \int_{\underline{i}_o}^{\bar{i}_o} \pi_I(i)di = \pi_I \tag{47}$$

and

$$\underline{i}_o\underline{\pi}_D + (1-\bar{i}_o)\overline{\pi}_D + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I(i), V_o^*(i))di = \pi_D. \tag{48}$$

How do we know when this is true? The following observation is key to the method that we use. Let Φ^* denote the subset of generating districtings that satisfy the feasibility requirement that the average fraction of Independents equals the actual fraction of the population (i.e., (47)). Then we have:

Lemma A.1: *Let* $\{(\underline{\pi}_D^o, \underline{\pi}_I^o), (\overline{\pi}_D^o, \overline{\pi}_I^o), \pi_I^o(i)\}$ *and* $\{(\underline{\pi}_D^1, \underline{\pi}_I^1), (\overline{\pi}_D^1, \overline{\pi}_I^1), \pi_I^1(i)\}$ *be two districtings in the set* Φ^* *such that*

$$\underline{i}_o\underline{\pi}_D^o + (1-\bar{i}_o)\overline{\pi}_D^o + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I^o(i), V_o^*(i))di \geq \pi_D \geq \underline{i}_o\underline{\pi}_D^1 + (1-\bar{i}_o)\overline{\pi}_D^1 + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I^1(i), V_o^*(i))di.$$

Then there exists a feasible districting in the set Φ.

Proof: Let

$$\Pi^o = \underline{i}_o\underline{\pi}_D^o + (1-\bar{i}_o)\overline{\pi}_D^o + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I^o(i), V_o^*(i))di$$

and

$$\Pi^1 = \underline{i}_o\underline{\pi}_D^1 + (1-\bar{i}_o)\overline{\pi}_D^1 + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I^1(i), V_o^*(i))di.$$

Choose $\lambda \in [0,1]$ such that

$$\lambda\Pi^o + (1-\lambda)\Pi^1 = \pi_D.$$

Then consider the districting $\{(\underline{\pi}_D^\lambda, \underline{\pi}_I^\lambda), (\overline{\pi}_D^\lambda, \overline{\pi}_I^\lambda), \pi_I^\lambda(i)\}$ that is the convex combination of $\{(\underline{\pi}_D^o, \underline{\pi}_I^o), (\overline{\pi}_D^o, \overline{\pi}_I^o), \pi_I^o(i)\}$ and $\{(\underline{\pi}_D^1, \underline{\pi}_I^1), (\overline{\pi}_D^1, \overline{\pi}_I^1), \pi_I^1(i)\}$ with weight λ. This districting is in the set Φ and is feasible. ■

Thus, if there exist two districtings in the set Φ^* one of which involves a higher fraction of Democrats than there are in the population and one a lower fraction, then there must exist a feasible districting in Φ^*.

Consider now the following pair of optimization problems:

$$\begin{array}{c} \min \underline{i}_o\underline{\pi}_D + (1-\bar{i}_o)\overline{\pi}_D + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I(i), V_o^*(i))di \\ s.t.\ \{(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I), \pi_I(i)\} \in \Phi^*, \end{array} \quad P_{\min}$$

and

$$\begin{array}{c} \max \underline{i}_o \underline{\pi}_D + (1-\bar{i}_o)\overline{\pi}_D + \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I(i), V_o^*(i))di \\ s.t.\ \{(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I), \pi_I(i)\} \in \Phi^*. \end{array} \quad P_{\max}$$

The minimization problem selects the districting in Φ^* that has the minimal fraction of Democrats, while the maximization problem selects the districting that has the maximal fraction of Democrats. Letting the values of these problems be $\underline{\Pi}$ and $\overline{\Pi}$ respectively, it follows from Lemma A.1 that there exists a feasible districting generating $\{\underline{i}_o, \bar{i}_o, V_o^*(i)\}$ if and only if $\underline{\Pi} \leq \pi_D \leq \overline{\Pi}$. Thus, the optimal seat-vote curve is implementable if and only if $\underline{\Pi} \leq \pi_D \leq \overline{\Pi}$.

7.1.1 The minimization problem $P_{\min}$

To progress further, we need to develop expressions for the values of the minimization and maximization problems $\underline{\Pi}$ and $\overline{\Pi}$. Consider first the minimization problem $P_{\min}$. To simplify the problem, note that in any solution it is clearly optimal to have no more Democrats than necessary in the safe Democrat seats. Thus, from (42), we have that

$$\underline{\pi}_D = \frac{1}{2} - \underline{\pi}_I(\frac{\tau-\varepsilon}{2\tau}). \tag{49}$$

Similarly, it is optimal to have no Democrats in the safe Republican seats and hence

$$\overline{\pi}_D = 0. \tag{50}$$

It follows from (50) that we can rewrite (43) as $\overline{\pi}_I \leq \frac{\tau}{\tau+\varepsilon}$. Similarly, (49) implies that the constraint that $\underline{\pi}_D + \underline{\pi}_I \leq 1$ amounts to $\underline{\pi}_I \leq \frac{\tau}{\tau+\varepsilon}$. Thus, we can rewrite the minimization problem as follows:

$$\begin{array}{c} \min_{\{\pi_I(i), \overline{\pi}_I, \underline{\pi}_I\}} \int_{\underline{i}_o}^{\bar{i}_o} f(\pi_I(i), V_o^*(i))di + \underline{i}_o[\frac{1}{2} - \underline{\pi}_I(\frac{\tau-\varepsilon}{2\tau})] \\ s.t.\ \ \underline{\pi}_I \in [0, \frac{\tau}{\tau+\varepsilon}];\ \ \overline{\pi}_I \in [0, \frac{\tau}{\tau+\varepsilon}];\ (46) \text{ and } (47). \end{array} \quad P_{\min}$$

In order for this problem to have a solution, it must be the case that the set Φ^* is non-empty. Thus, there must exist at least one generating districting which has the property that the average fraction of Independents equals the actual fraction in the population. A necessary and sufficient condition for this to be true is that

$$\pi_I \leq \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \int_{\underline{i}_o}^{\widehat{i}_o} \frac{\pi_I}{2(\pi_I + \pi_D - V_o^*(i))}di + \int_{\widehat{i}_o}^{\bar{i}_o} \frac{\pi_I}{2(V_o^*(i) - \pi_D)}di + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}.$$

The expression on the right hand side is the fraction of Independents associated with the generating districting that maximizes the use of Independents. Using the fact that

$$\int_{\underline{i}_o}^{\widehat{i}_o} \frac{\pi_I}{2(\pi_I + \pi_D - V_o^*(i))} di = \int_{\widehat{i}_o}^{\overline{i}_o} \frac{\pi_I}{2(V_o^*(i) - \pi_D)} di = \pi_I \tau \ln(1 + \frac{\varepsilon}{\tau}),$$

we can write this as

$$\pi_I \leq \underline{i}_o \frac{\tau}{\tau + \varepsilon} + \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}) + (1 - \overline{i}_o)\frac{\tau}{\tau + \varepsilon}. \tag{51}$$

We will assume that this inequality is satisfied in what follows.

We can now state the value of the minimization problem:

Lemma A.2: *(i) If* $\pi_I \in [\underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}) + (1 - \overline{i}_o)\frac{\tau}{\tau+\varepsilon}]$, *then*

$$\underline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1 + \frac{\varepsilon}{\tau}) + \underline{i}_o \frac{\varepsilon}{\tau + \varepsilon}.$$

(ii) If $\pi_I \in [\pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau})]$, *then*

$$\underline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1 + \frac{\varepsilon}{\tau}) + \underline{i}_o [\frac{1}{2} - (\frac{\pi_I - \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau})}{\underline{i}_o})(\frac{\tau - \varepsilon}{2\tau})].$$

Proof: Ignoring the inequality constraints on the choice variables, the Lagrangian for the problem is

$$\mathcal{L} = \int_{\underline{i}_o}^{\overline{i}_o} f(\pi_I(i), V_o^*(i)) di + \underline{i}_o [\frac{1}{2} - \underline{\pi}_I (\frac{\tau - \varepsilon}{2\tau})] + \lambda [\underline{i}_o \underline{\pi}_I + \int_{\underline{i}_o}^{\overline{i}_o} \pi_I(i) di + (1 - \overline{i}_o)\overline{\pi}_I]$$

where λ is the Lagrange multiplier on the aggregate constraint (47). Using the definition of the function $f(\cdot)$ we can write this as

$$\mathcal{L} = \int_{\underline{i}_o}^{\overline{i}_o} \pi_I(i)[\lambda - \frac{(V_o^*(i) - \pi_D)}{\pi_I}] di + \underline{\pi}_I \underline{i}_o [\lambda - (\frac{\tau - \varepsilon}{2\tau})] + \overline{\pi}_I (1 - \overline{i}_o)\lambda + \text{constant}$$

We can therefore minimize the Lagrangian pointwise with respect to $\pi_I(i)$, $\underline{\pi}_I$ and $\overline{\pi}_I$, respecting the inequality constraints on these variables. The value of the multiplier λ must be such that (47) is satisfied.

From the fact that $V_o^*(i) \in [\underline{V}, \overline{V}]$, we know that

$$\frac{\tau + \varepsilon}{2\tau} \geq \frac{(V_o^*(i) - \pi_D)}{\pi_I} \geq \frac{\tau - \varepsilon}{2\tau} \geq 0 \quad \text{for all } i \in [\underline{i}_o, \overline{i}_o].$$

It follows that $\lambda \leq \frac{\tau+\varepsilon}{2\tau}$, for if this were not the case, then the solution involves $\pi_I(i) = 0$ for all i, $\underline{\pi}_I = 0$ and $\overline{\pi}_I = 0$. This means that constraint (47) cannot be satisfied. In addition, note that if the multiplier lies in the interval 0 to $\frac{\tau-\varepsilon}{2\tau}$ this generates no more potential solutions than values

of the multiplier equal to 0. Thus, we can restrict attention to three possibilities: (i) $\lambda = 0$; (ii) $\lambda = \frac{\tau-\varepsilon}{2\tau}$; and (iii) $\lambda \in (\frac{\tau-\varepsilon}{2\tau}, \frac{\tau+\varepsilon}{2\tau})$.

Case 1: $\lambda = 0$

In this case, the solution involves setting the fraction of Independents in the safe Democrat seats and competitive seats equal to their maximal levels, so that $\underline{\pi}_I = \frac{\tau}{\tau+\varepsilon}$ and

$$\pi_I(i) \in \begin{cases} \frac{\pi_I}{2(\pi_I+\pi_D-V_o^*(i))} & \text{if } i \in [\underline{i}_o, \widehat{i}_o) \\ \frac{\pi_I}{2(V_o^*(i)-\pi_D)} & \text{if } \ i \in [\widehat{i}_o, \overline{i}_o] \end{cases}.$$

The fraction of Independents in the safe Republican seats does not affect the value of the Lagrangian and hence can be set equal to any level $x \in [0, \frac{\tau}{\tau+\varepsilon}]$. In order that (47) be satisfied we need that

$$\underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\overline{i}_o)x = \pi_I.$$

Thus, for this to be a solution, it must be that $\pi_I \in [\underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\overline{i}_o)\frac{\tau}{\tau+\varepsilon}]$.

Case 2: $\lambda = \frac{\tau-\varepsilon}{2\tau}$

In this case, the solution involves setting the fractions of Independents in the competitive seats equal to their maximal levels, so that

$$\pi_I(i) \in \begin{cases} \frac{\pi_I}{2(\pi_I+\pi_D-V_o^*(i))} & \text{if } i \in [\underline{i}_o, \widehat{i}_o) \\ \frac{\pi_I}{2(V_o^*(i)-\pi_D)} & \text{if } \ i \in [\widehat{i}_o, \overline{i}_o] \end{cases}$$

and the fraction of Independents in the safe Republican seats equal to zero so that $\overline{\pi}_I = 0$. The fraction of Independents in the safe Democrat seats does not effect the value of the Lagrangian and hence can be set equal to any level $x \in [0, \frac{\tau}{\tau+\varepsilon}]$. In order that constraint (47) be satisfied we need that

$$x\frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) = \pi_I.$$

Thus, for this to be a solution, it must be that $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})]$.

Case 3: $\lambda \in (\frac{\tau-\varepsilon}{2\tau}, \frac{\tau+\varepsilon}{2\tau})$

It can be shown that if the multiplier is in this range, it must be that $\pi_I < \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})$, which is not possible.

We conclude that: (i) If $\pi_I \in [\underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\overline{i}_o)\frac{\tau}{\tau+\varepsilon}]$,

then we are in Case 1 and the solution to the minimization problem is

$$\pi_I(i) = \begin{cases} \frac{\tau}{\tau+\varepsilon} & \text{if } i \in [0, \underline{i}_o) \\ \frac{\pi_I}{2(\pi_I+\pi_D-V_o^*(i))} & \text{if } i \in [\underline{i}_o, \widehat{i}_o) \\ \frac{\pi_I}{2(V_o^*(i)-\pi_D)} & \text{if } i \in [\widehat{i}_o, \overline{i}_o] \\ \frac{\pi_I-[\underline{i}_o\frac{\tau}{\tau+\varepsilon}+\pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau})]}{1-\overline{i}_o} & \text{if } i \in (\overline{i}_o, 1] \end{cases}.$$

(ii) If $\pi_I \in [\pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o\frac{\tau}{\tau+\varepsilon}+\pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau})]$, then we are in Case 2 and the solution to the minimization problem is

$$\pi_I(i) = \begin{cases} \frac{\pi_I-[\pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau})]}{\underline{i}_o} & \text{if } i \in [0, \underline{i}_o) \\ \frac{\pi_I}{2(\pi_I+\pi_D-V_o^*(i))} & \text{if } i \in [\underline{i}_o, \widehat{i}_o) \\ \frac{\pi_I}{2(V_o^*(i)-\pi_D)} & \text{if } i \in [\widehat{i}_o, \overline{i}_o] \\ 0 & \text{if } i \in (\overline{i}_o, 1] \end{cases}.$$

We can now prove the Lemma by deriving the corresponding allocation of Democrats across districts and computing the aggregate fraction of Democrats used. For example, in case (i), equations (49), (50), and the fact that $\pi_D(i) = f(\pi_I(i), V_o^*(i))$ for all $i \in [\underline{i}_o, \overline{i}_o]$, imply that

$$\pi_D(i) = \begin{cases} \frac{\varepsilon}{\tau+\varepsilon} & \text{if } i \in [0, \underline{i}_o) \\ \frac{\pi_I/2+\pi_D-V_o^*(i)}{\pi_I+\pi_D-V_o^*(i)} & \text{if } i \in [\underline{i}_o, \widehat{i}_o) \\ 0 & \text{if } i \in [\widehat{i}_o, \overline{i}_o] \\ 0 & \text{if } i \in (\overline{i}_o, 1] \end{cases}.$$

Thus, we have that

$$\begin{aligned} \underline{\Pi} &= \textstyle\int_{\underline{i}_o}^{\widehat{i}_o} (\frac{\pi_I/2+\pi_D-V_o^*(i)}{\pi_I+\pi_D-V_o^*(i)})di + \underline{i}_o\frac{\varepsilon}{\tau+\varepsilon} \\ &= \pi_I\varepsilon - \pi_I\tau\ln(1+\tfrac{\varepsilon}{\tau}) + \underline{i}_o\tfrac{\varepsilon}{\tau+\varepsilon}. \end{aligned}$$

This completes the proof of Lemma A.1. ■

7.1.2 The maximization problem $P_{\max}$

Turning to the maximization problem, we have that:

Lemma A.3: *(i) If* $\pi_I \in [\pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau}) + (1-\overline{i}_o)\frac{\tau}{\tau+\varepsilon}, \underline{i}_o\frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau\ln(1+\frac{\varepsilon}{\tau}) + (1-\overline{i}_o)\frac{\tau}{\tau+\varepsilon}]$, *then*

$$\overline{\Pi} = 1 - \pi_I - \pi_I\varepsilon + \pi_I\tau\ln(1+\frac{\varepsilon}{\tau}) - (1-\overline{i}_o)\frac{\varepsilon}{\tau+\varepsilon}.$$

(ii) If $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}]$*, then*

$$\overline{\Pi} = 1 - \pi_I - \pi_I \varepsilon + \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}) - (1-\bar{i}_o)[\frac{1}{2} - (\frac{\pi_I - \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})}{1-\bar{i}_o})(\frac{\tau-\varepsilon}{2\tau})].$$

Proof: Following the same steps as used in the proof of Lemma A.2, it may be shown that: (i) if $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}, \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}]$, then

$$\overline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}) + \underline{i}_o[1 - (\frac{\pi_I - [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}]}{\underline{i}_o})],$$

and (ii) if $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}]$, then

$$\overline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}) + \underline{i}_o + (1-\bar{i}_o)[\frac{1}{2} - (\frac{\pi_I - [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})]}{1-\bar{i}_o})(\frac{\tau+\varepsilon}{2\tau})].$$

Using the definitions of $\underline{i}_o$ and $\bar{i}_o$ in (39) and (40) and with a little work, these expressions can be shown to equal the claimed expressions in the statement of the Lemma. ■

7.1.3 Completing the proof

We will now show that the optimal inverse seat-vote curve $\{\underline{i}_o, \bar{i}_o, V_o^*(\cdot)\}$ satisfies the constraint that $\underline{\Pi} \leq \pi_D$ if and only if (12) holds for π_D and the constraint that $\overline{\Pi} \geq \pi_D$ if and only if (12) holds for π_R. We begin with the former.

From Lemma A.1 we know that: (i) if $\pi_I \in [\underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)\frac{\tau}{\tau+\varepsilon}]$, then

$$\underline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}) + \underline{i}_o \frac{\varepsilon}{\tau+\varepsilon},$$

and (ii) if $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})]$, then

$$\underline{\Pi} = \pi_I \varepsilon - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}) + \underline{i}_o[\frac{1}{2} - (\frac{\pi_I - \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})}{\underline{i}_o})(\frac{\tau-\varepsilon}{2\tau})].$$

In addition, observe that after substituting in for $\underline{i}_o$, we have that $\pi_I \geq \underline{i}_o \frac{\tau}{\tau+\varepsilon} + \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})$ if and only if

$$\pi_I \geq \frac{\pi_D}{(1+\frac{\varepsilon}{\tau})[1-2\tau \ln(1+\frac{\varepsilon}{\tau})] + \varepsilon - \frac{1}{2}}, \tag{52}$$

so that case (i) arises if (52) holds and case (ii) otherwise.

Suppose that (52) holds so that case (i) arises. Then, after substituting in for $\underline{i}_o$, we have that

$$\underline{\Pi} = (\pi_D + \frac{\pi_I}{2})\frac{\varepsilon}{\tau+\varepsilon} + \pi_I \varepsilon \frac{\tau}{\tau+\varepsilon} - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}).$$

Thus, in this case, the constraint that $\underline{\Pi} \leq \pi_D$ is satisfied if and only if

$$\pi_D \geq \pi_I(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})),$$

which is just (12).

Next suppose that (52) does not hold and case (ii) arises. Then, after substituting in for $\underline{i}_o$, we have

$$\underline{\Pi} = \frac{\pi_I}{2}(\varepsilon + \frac{\varepsilon}{\tau} - \frac{1}{2}) + \frac{\pi_D}{2} - \pi_I \varepsilon \ln(1 + \frac{\varepsilon}{\tau})$$

and thus the constraint that $\underline{\Pi} \leq \pi_D$ is satisfied if and only if

$$\frac{\pi_I}{2}(\varepsilon + \frac{\varepsilon}{\tau} - \frac{1}{2}) - \pi_I \varepsilon \ln(1 + \frac{\varepsilon}{\tau}) \leq \frac{\pi_D}{2}. \tag{53}$$

To summarize, if (52) holds the constraint $\underline{\Pi} \leq \pi_D$ will be satisfied if and only if (12) is satisfied. If (52) does not hold the constraint that $\underline{\Pi} \leq \pi_D$ will be satisfied if and only if (53) is satisfied.

We can now prove that $\underline{\Pi} \leq \pi_D$ if and only if (12) holds for π_D. Suppose first that (12) is not satisfied. This implies that (52) holds since

$$(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})) < (1 + \frac{\varepsilon}{\tau})(1 - 2\tau \ln(1 + \frac{\varepsilon}{\tau})) + \varepsilon - \frac{1}{2}.$$

It follows that the constraint $\underline{\Pi} \leq \pi_D$ will be violated. Next suppose that (12) is satisfied. Then we claim that (53) must also be satisfied. We need to show that

$$\pi_D \geq \pi_I(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})),$$

implies that

$$\pi_D \geq \pi_I\{(\varepsilon + \frac{\varepsilon}{\tau} - \frac{1}{2}) - 2\varepsilon \ln(1 + \frac{\varepsilon}{\tau})\}.$$

This amounts to $1 \geq 2\varepsilon \ln(1 + \frac{\varepsilon}{\tau})$, which holds under our assumptions on ε and τ. It follows that, irrespective of whether (52) holds, the constraint $\underline{\Pi} \leq \pi_D$ will be satisfied.

It only remains to show that $\overline{\Pi} \geq \pi_D$ if and only if (12) holds for π_R. From Lemma A.2, we know that (i) if $\pi_I \in [\pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}) + \frac{1 - \bar{i}_o}{1 + \frac{\varepsilon}{\tau}}, \frac{i_o}{1 + \frac{\varepsilon}{\tau}} + \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}) + \frac{1 - \bar{i}_o}{1 + \frac{\varepsilon}{\tau}}]$ then $\overline{\Pi} \geq \pi_D$ if and only if

$$1 - \pi_I - \pi_I \varepsilon + \pi_I \tau \ln(1 + \frac{\varepsilon}{\tau}) - (1 - \bar{i}_o)\frac{\varepsilon}{\tau + \varepsilon} \geq \pi_D$$

which is equivalent to

$$\pi_R \geq \pi_I \varepsilon - \pi_I \tau \ln(1 + \frac{\varepsilon}{\tau}) + (1 - \bar{i}_o)\frac{\varepsilon}{\tau + \varepsilon}.$$

Similarly, (ii) if $\pi_I \in [\pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}), \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau}) + \frac{(1-\bar{i}_o)}{1+\frac{\varepsilon}{\tau}}]$, then $\overline{\Pi} \geq \pi_D$ if and only if

$$1 - \pi_I - \pi_I\varepsilon + \pi_I\tau \ln(1+\frac{\varepsilon}{\tau}) - (1-\bar{i}_o)[\frac{1}{2} - (\frac{\pi_I - \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})}{1-\bar{i}_o})(\frac{\tau-\varepsilon}{2\tau})] \geq \pi_D$$

which is equivalent to

$$\pi_R \geq \pi_I\varepsilon - \pi_I\tau \ln(1+\frac{\varepsilon}{\tau}) + (1-\bar{i}_o)[\frac{1}{2} - (\frac{\pi_I - \pi_I 2\tau \ln(1+\frac{\varepsilon}{\tau})}{1-\bar{i}_o})(\frac{\tau-\varepsilon}{2\tau})].$$

Observing that $1-\bar{i}_o = \pi_R + \frac{\pi_I}{2} - \pi_I\varepsilon$, we can simply apply the argument from the first part of the proof with π_R replacing π_D to reach the desired conclusion. ■

7.2 Proof of Proposition 3

Using the definitions from the proof of Proposition 2, the optimal seat-vote curve is implementable with a districting of the form in (16) if and only if (a) the proposed districting is a feasible districting and (b) $\pi_D + 1/2 - \underline{\pi}_D \leq \underline{V}$ and $\pi_D + 1/2 - \overline{\pi}_D \geq \overline{V}$.

The proposed districting is a feasible districting if and only if the following conditions are satisfied: (a.i) $\underline{\pi}_D \in [0, 1-\pi_I]$, (a.ii) $\overline{\pi}_D \in [0, 1-\pi_I]$, (a.iii) for all $i \in [\underline{i}_o, \bar{i}_o]$, $\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau} \in [0, 1-\pi_I]$, and, (a.iv)

$$\underline{i}_o\underline{\pi}_D + \int_{\underline{i}_o}^{\bar{i}_o} [\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau}]di + (1-\bar{i}_o)\overline{\pi}_D = \pi_D. \tag{54}$$

It is straightforward to show that condition (a.iii) is satisfied if and only if $\pi_I \leq \frac{\tau}{\tau+\varepsilon}$. Condition (a.iv) can be simplified by noting that

$$\int_{\underline{i}_o}^{\bar{i}_o} [\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau}]di = \pi_I\varepsilon(1-\pi_I),$$

so that (54) can be rewritten as

$$\underline{i}_o\underline{\pi}_D + \pi_I\varepsilon(1-\pi_I) + (1-\bar{i}_o)\overline{\pi}_D = \pi_D. \tag{55}$$

Using the definitions of $\underline{V}$ and $\overline{V}$, the inequality requirements in (b) can be rewritten as $\underline{\pi}_D \geq \frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})$ and $\overline{\pi}_D \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$.

Combining all this, the optimal seat-vote curve is implementable with a districting of the form in (16) if and only if there exist $\underline{\pi}_D \in [\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau}), 1-\pi_I]$ and $\overline{\pi}_D \in [0, \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})]$ that satisfy (55). Solving (55), we have that

$$\overline{\pi}_D = \frac{\pi_D - \pi_I\varepsilon(1-\pi_I) - \underline{i}_o\underline{\pi}_D}{1-\bar{i}_o}.$$

So defining the function:

$$g(\underline{\pi}_D) = \frac{\pi_D - \pi_I\varepsilon(1-\pi_I) - \underline{i}_o\underline{\pi}_D}{1-\bar{i}_o},$$

the optimal seat-vote curve is implementable with a districting of the form in (16) if and only if there exists $\underline{\pi}_D \in [\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau}), 1-\pi_I]$ such that $g(\underline{\pi}_D) \in [0, \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})]$.

Since g is decreasing, it follows that if $g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$ the condition is met if and only if $g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) \geq 0$, while if $g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) > \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$ the condition is met if and only if $g(1-\pi_I) \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$. Observe that

$$g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) = \frac{\pi_D - \pi_I\varepsilon(1-\pi_I) - \underline{i}_o[\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})]}{1-\bar{i}_o}$$

so that $g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$ if and only if $\pi_D \leq \pi_R$. Thus, if $\pi_D \leq \pi_R$ the condition is met if and only if $g(\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})) \geq 0$ and if $\pi_D > \pi_R$ it is met if and only if $g(1-\pi_I) \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$.

So suppose that $\pi_D \leq \pi_R$. Then, the condition is

$$\frac{\pi_D - \pi_I\varepsilon(1-\pi_I) - \underline{i}_o[\frac{1}{2} - \pi_I(\frac{\tau-\varepsilon}{2\tau})]}{1-\bar{i}_o} \geq 0,$$

which is equivalent to (17) holding for π_D. On the other hand, if $\pi_D > \pi_R$, then the condition is

$$\frac{\pi_D - \pi_I\varepsilon(1-\pi_I) - \underline{i}_o(1-\pi_I)}{1-\bar{i}_o} \leq \frac{1}{2} - \pi_I(\frac{\tau+\varepsilon}{2\tau})$$

which with a little work can be shown equivalent to (17) holding for π_R. ■

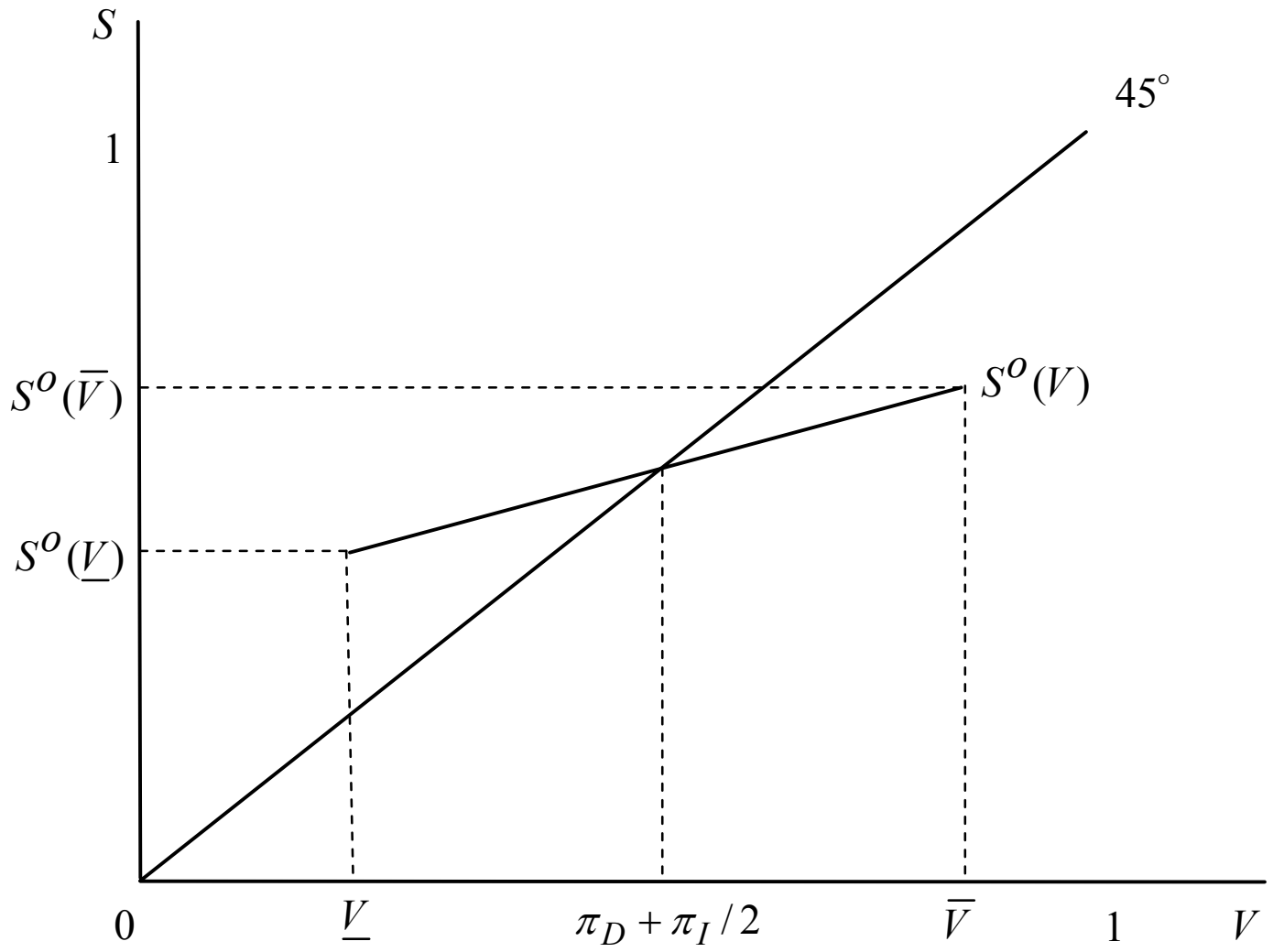
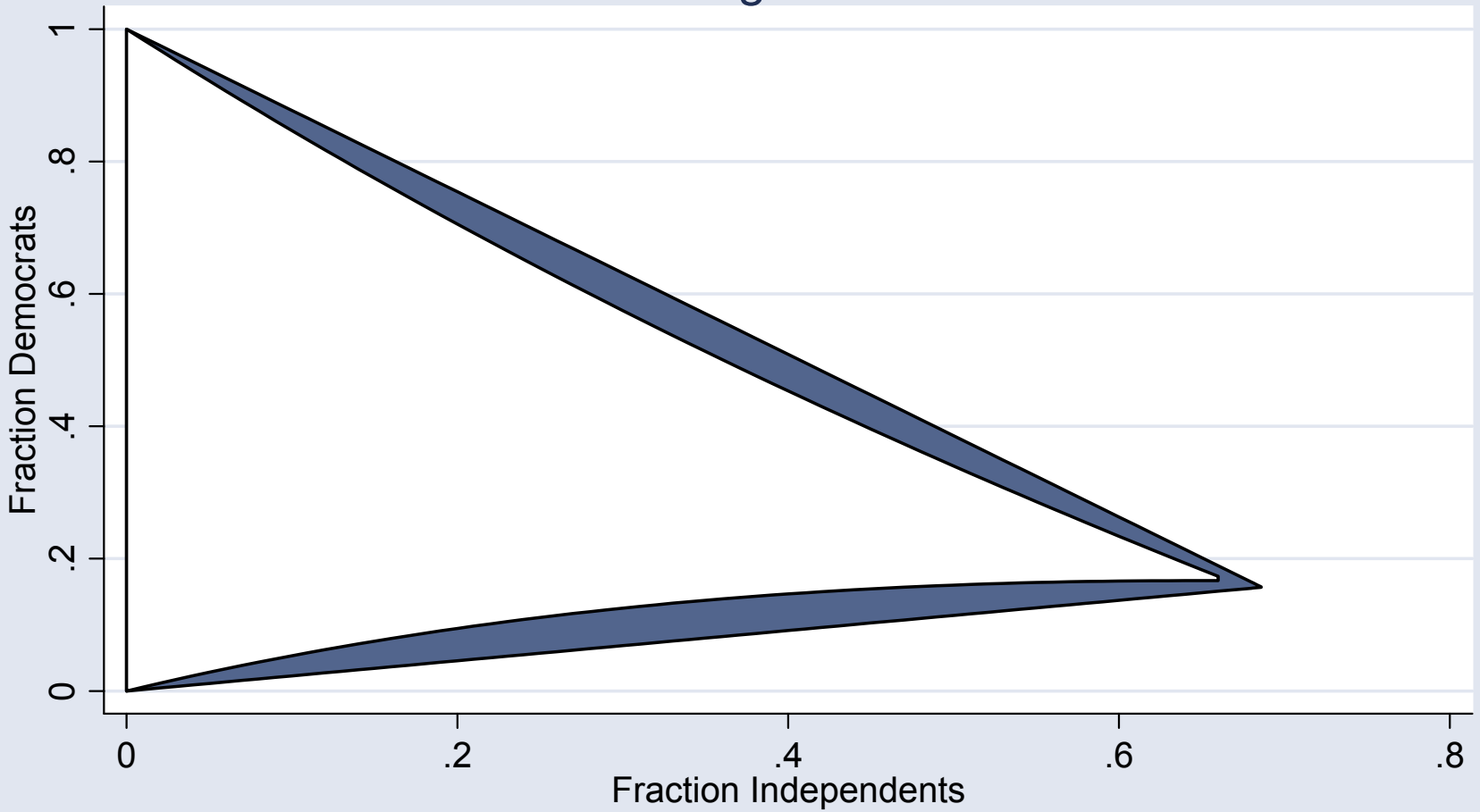
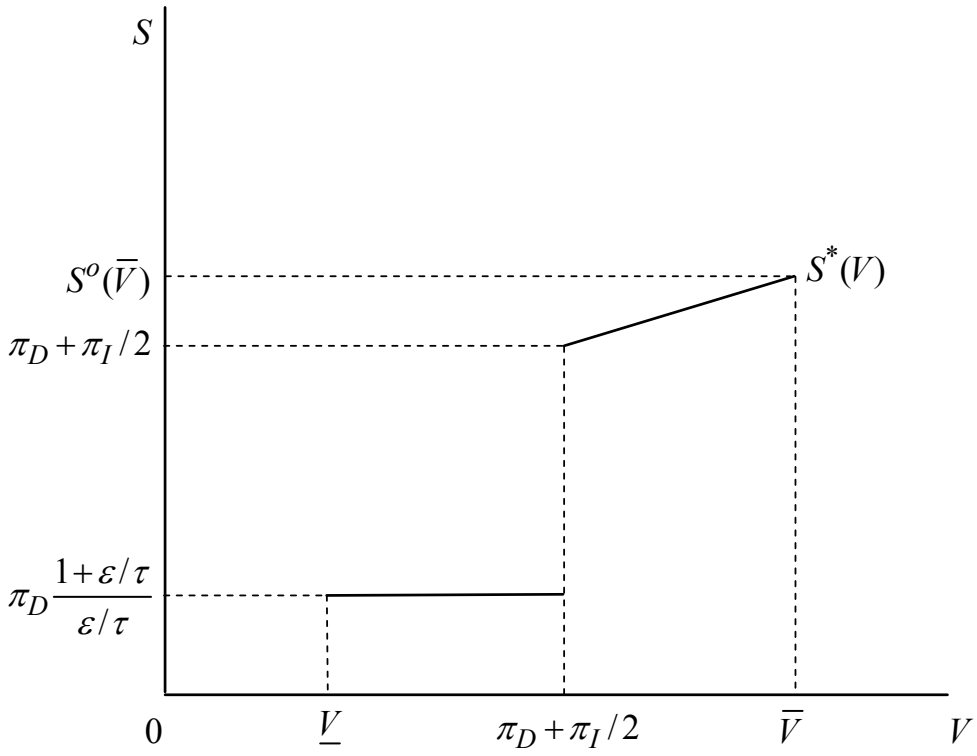

S
1
45°
$S^o(\overline{V})$
$S^o(V)$
$S^o(\underline{V})$
0
$\underline{V}$
$\pi_D + \pi_I / 2$
$\overline{V}$
1
V

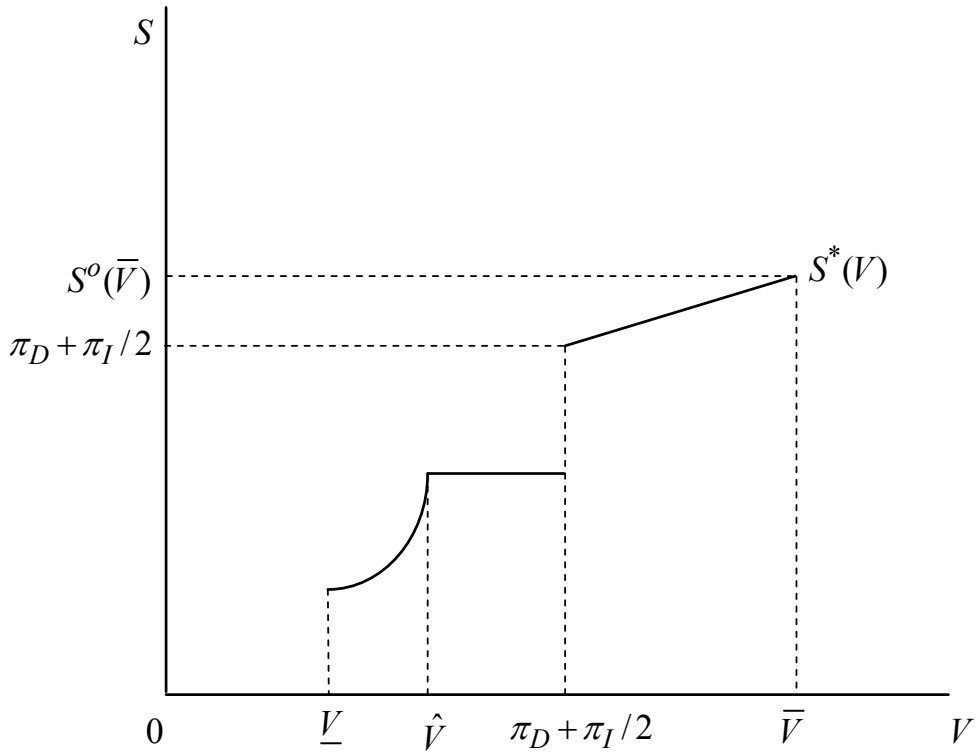
Figure 1

Figure 2


Fraction Democrats
0
.2
.4
.6
.8
1
0
.2
.4
.6
.8
Fraction Independents
Entire area represents pairs satisfying Proposition 2
Shaded area represents pairs satisfying Proposition 2 but not Proposition 3


S
$S^o(\overline{V})$
$S^*(V)$
$\pi_D + \pi_I/2$
$\pi_D \frac{1+\varepsilon/\tau}{\varepsilon/\tau}$
0
$\underline{V}$
$\pi_D + \pi_I/2$
$\overline{V}$
V

(a)


S
$S^o(\overline{V})$
$S^*(V)$
$\pi_D + \pi_I/2$
0
$\underline{V}$
$\hat{V}$
$\pi_D + \pi_I/2$
$\overline{V}$
V

(b)

Figure 3

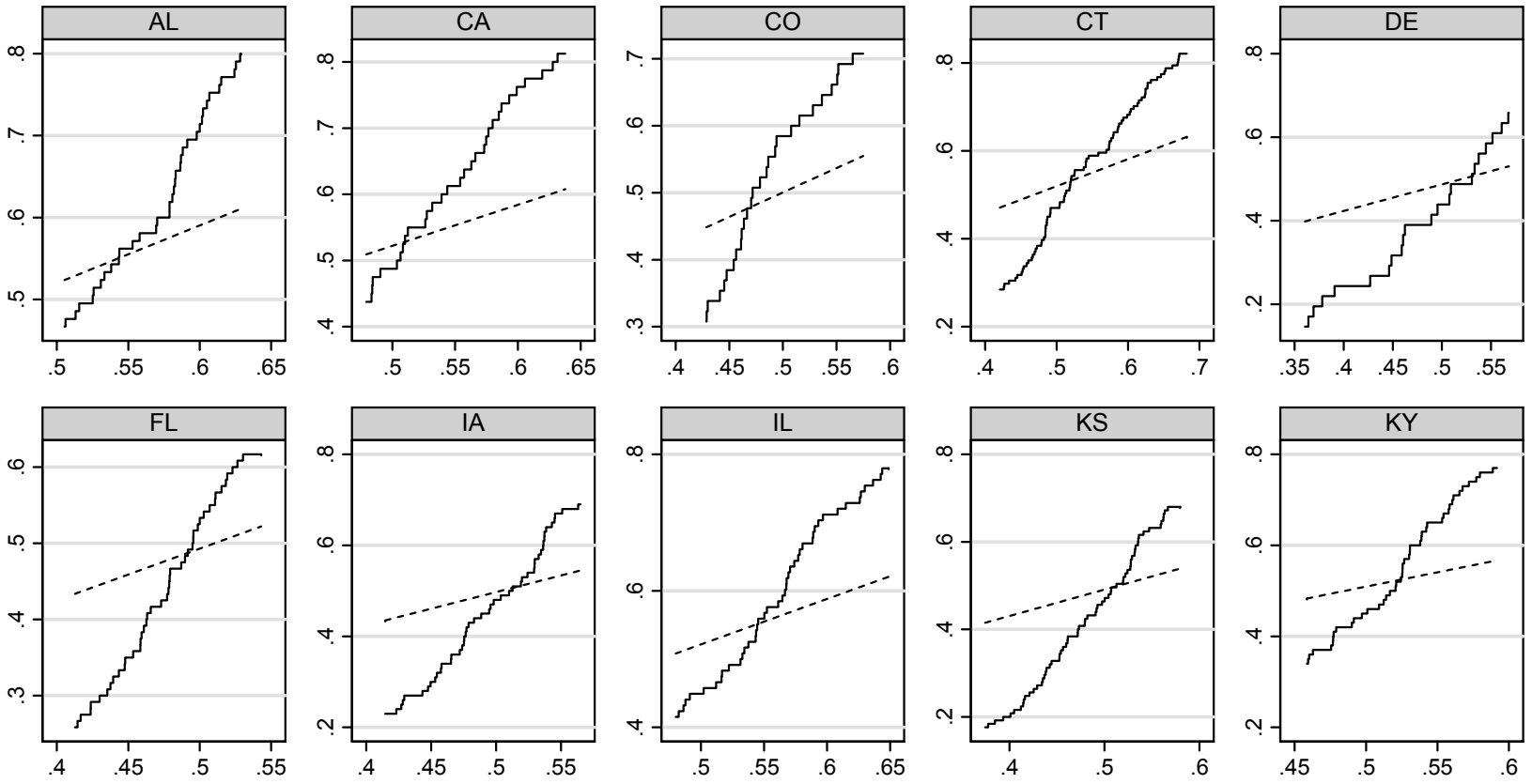


Figure 4a: Seat-Vote Curves

Graphs by state

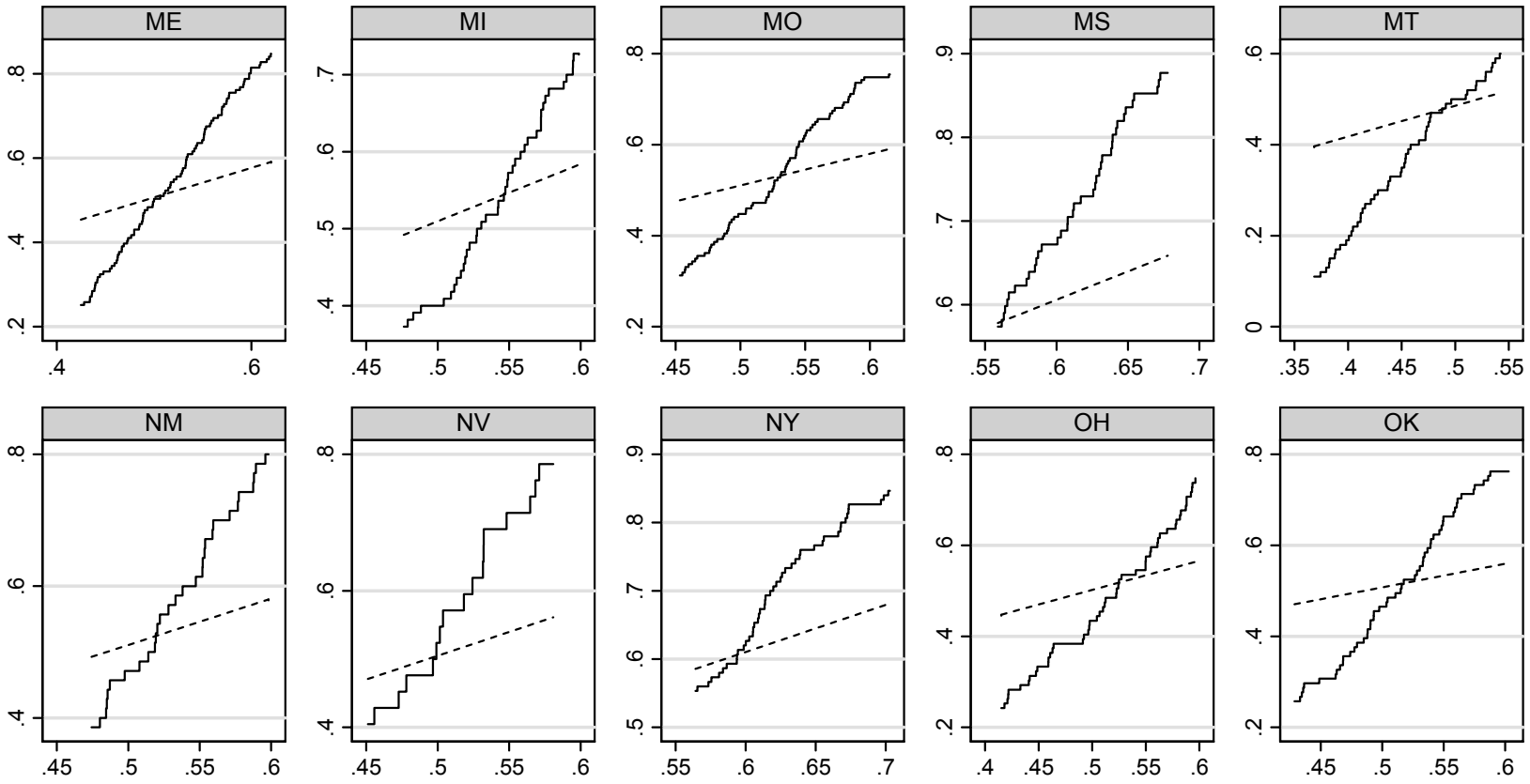


Figure 4b: Seat-Vote Curves

Graphs by state

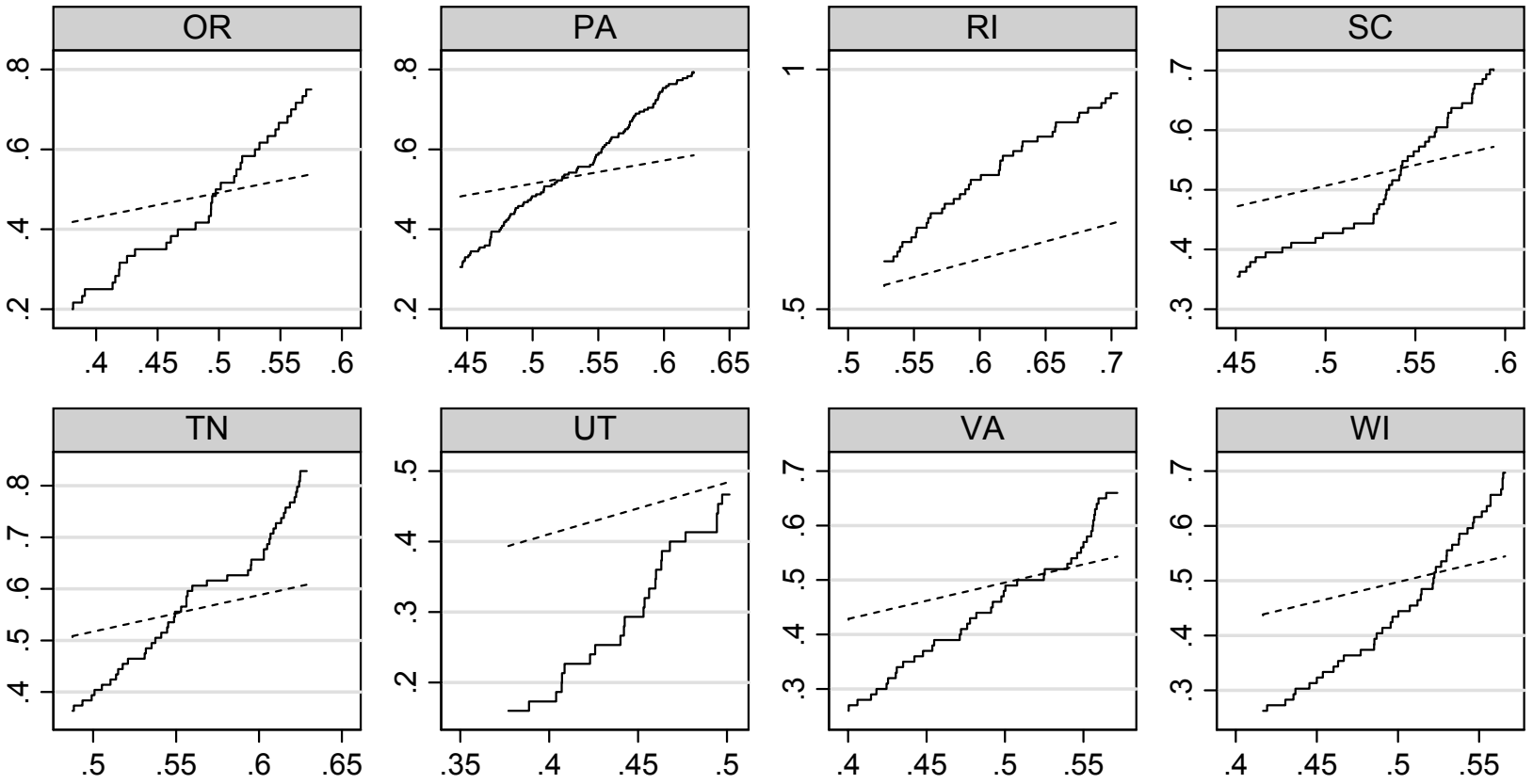


Figure 4c: Seat-Vote Curves

S(V) Optimal S(V)

Graphs by state

Table 1: States and Years Included in Analysis

state	*first redistricting*	*subsequent redistricting*	*number districts*
AL	1994	2002	105
CA	1992	2002	80
CO	1992	1998	65
CT	1992	2002	151
DE	1992	2002	41
FL	1994	2002	120
IA	1992	2002	100
IL	1992	2002	118
KS	1992	2002	125
KY	1996	2002	100
ME	1994	2002	151
MI	1992	2002	110
MO	1992	2002	163
MS	1995	2002	122
MT	1994	2002	100
NM	1992	2002	70
NV	1992	2002	42
NY	1992	2002	150
OH	1992	2002	99
OK	1992	2002	101
OR	1992	2002	60
PA	1992	2002	203
RI	1992	2002	100
SC	1992	1998	124
TN	1994	2002	99
UT	1992	2002	75
VA	1991	2001	100
WI	1992	2002	99

Table 2: Random Effects Regression Results

(all regressions include state-specific constant terms, bootstrap standard errors in parentheses)

Moment	mean	variance
percent urban	0.0694**	-0.0570
	(0.0078)	(0.0947)
percent suburban	0.0357**	-0.1332
	(0.0083)	(0.1011)
household income (thousands)	-0.0035**	-0.0120**
	(0.0004)	(0.0040)
percent with college degree	0.1438**	-0.2667
	(0.0388)	(0.4534)
percent over age 65	0.3649**	0.0648
	(0.0364)	(0.5930)
percent African American	0.4779**	-2.7374**
	(0.0144)	(0.2649)
percent Hispanic	0.2926**	-0.1256
	(0.0320)	(0.3273)
R-squared	0.4386	0.0494
Number of districts	2707	2707
Number of elections	8504	8504

Table 3: Properties of Estimated and Optimal Seat-Vote Curves

(for replication associated with median welfare loss by state)

state	*r*	*optimal r*	*difference*	*b*	*optimal b*	*difference*	*expected seats difference*
AL	2.7483	0.7044	2.0439				0.0480
CA	2.4170	0.6212	1.7958	-0.0125	0.0162	-0.0287	0.0826
CO	2.6867	0.7277	1.9589	0.0846	-0.0038	0.0884	0.0480
CT	2.1870	0.6155	1.5716	-0.0298	0.0146	-0.0444	0.0183
DE	2.2597	0.6354	1.6243	-0.0610	-0.0158	-0.0452	-0.0951
FL	3.0914	0.6791	2.4123	0.0250	-0.0080	0.0330	-0.0332
IA	3.3310	0.7288	2.6022	-0.0200	-0.0041	-0.0159	-0.0457
IL	2.3693	0.6658	1.7035	-0.0508	0.0157	-0.0665	0.0364
KS	2.7312	0.6033	2.1279	-0.0360	-0.0106	-0.0254	-0.0627
KY	3.5330	0.6289	2.9041	-0.0500	0.0075	-0.0575	0.0314
ME	3.1589	0.7003	2.4587	0.0033	0.0067	-0.0034	0.0361
MI	3.1192	0.7416	2.3776	-0.1000	0.0010	-0.1010	-0.0086
MO	3.0136	0.6960	2.3176	-0.0521	0.0093	-0.0615	0.0160
MS	2.6200	0.6771	1.9429				0.1156
MT	2.8717	0.6763	2.1954	0.0000	-0.0170	0.0170	-0.0922
NM	3.3887	0.7038	2.6849	-0.0286	0.0090	-0.0376	0.0525
NV	3.1484	0.6965	2.4519	0.0238	0.0042	0.0196	0.0697
NY	2.3173	0.6900	1.6272				0.0812
OH	2.5471	0.6413	1.9058	-0.0657	0.0005	-0.0662	-0.0424
OK	3.2740	0.5186	2.7554	-0.0347	0.0063	-0.0410	0.0000
OR	2.7487	0.6145	2.1342	0.0000	-0.0101	0.0101	-0.0313
PA	2.6887	0.5803	2.1084	-0.0172	0.0139	-0.0311	0.0242
RI	1.9708	0.7432	1.2276				0.1757
SC	2.3915	0.6985	1.6930	-0.0726	0.0057	-0.0783	-0.0277
TN	2.9736	0.7059	2.2677	-0.1061	0.0165	-0.1225	0.0078
UT	2.6488	0.7316	1.9172	-0.0333	-0.0182	-0.0151	-0.1501
VA	2.1373	0.6678	1.4695	-0.0200	-0.0052	-0.0148	-0.0442
WI	2.8097	0.7094	2.1003	-0.0657	-0.0049	-0.0608	-0.0587
Average	2.7565	0.6715	2.0850	-0.0300	0.0012	-0.0312	0.0054

Table 4: Properties of Estimated and Optimal Seat-Vote Curves
(properties of distribution across all replications)

	difference in responsiveness				*difference in bias*				*difference in expected seats*			
	mean	*median*	*confidence interval*		*mean*	*median*	*confidence interval*		*mean*	*median*	*confidence interval*	
AL	1.8239	1.8652	1.0131 ,	2.4326	-0.0804	-0.0812	-0.1421 ,	-0.0379	0.0541	0.0487	0.0108 ,	0.1202
CA	1.8327	1.8481	1.0972 ,	2.4980	-0.0283	-0.0271	-0.0798 ,	0.0329	0.0792	0.0838	-0.0001 ,	0.1356
CO	1.8700	1.8641	1.1077 ,	2.7489	-0.0059	-0.0047	-0.0724 ,	0.0542	-0.0560	-0.0549	-0.1323 ,	0.0170
CT	1.7339	1.7479	1.4181 ,	2.0376	-0.0577	-0.0555	-0.1023 ,	-0.0244	0.0265	0.0258	-0.0064 ,	0.0551
DE	2.0212	2.0231	1.2378 ,	2.8136	-0.0427	-0.0416	-0.1273 ,	0.0361	-0.0584	-0.0543	-0.1365 ,	0.0248
FL	2.0488	2.0093	1.5851 ,	2.6673	-0.0306	-0.0283	-0.0840 ,	0.0206	-0.0239	-0.0263	-0.0767 ,	0.0361
IA	2.3761	2.3970	1.8240 ,	2.9774	-0.0014	-0.0048	-0.0550 ,	0.0487	-0.0134	-0.0145	-0.0628 ,	0.0418
IL	1.7695	1.8173	1.2300 ,	2.2815	-0.0878	-0.0865	-0.1295 ,	-0.0507	0.0105	0.0104	-0.0297 ,	0.0455
KS	2.0040	1.9937	1.5830 ,	2.5635	-0.0014	-0.0013	-0.0494 ,	0.0427	-0.0857	-0.0833	-0.1269 ,	-0.0445
KY	2.5209	2.5256	1.7970 ,	3.2608	-0.0490	-0.0508	-0.0999 ,	0.0033	0.0496	0.0487	-0.0038 ,	0.1140
ME	2.3666	2.3653	2.0194 ,	2.6793	-0.0061	-0.0035	-0.0442 ,	0.0429	0.0435	0.0473	-0.0008 ,	0.0830
MI	2.0319	2.0469	1.3785 ,	2.6249	-0.0951	-0.0973	-0.1413 ,	-0.0515	-0.0233	-0.0265	-0.0771 ,	0.0313
MO	2.1379	2.1538	1.7466 ,	2.6283	-0.0767	-0.0783	-0.1163 ,	-0.0378	0.0138	0.0158	-0.0215 ,	0.0500
MS	1.6717	1.7072	1.0177 ,	2.3327	-0.0241	-0.0241	-0.0440	-0.0043	0.1135	0.1118	0.0630 ,	0.1655
MT	2.2381	2.2022	1.6500 ,	3.1036	0.0124	0.0115	-0.0414 ,	0.0649	-0.1023	-0.1017	-0.1568 ,	-0.0490
NM	1.9778	1.9869	1.1543 ,	2.7721	-0.0035	-0.0028	-0.0704 ,	0.0580	0.0743	0.0721	0.0020 ,	0.1387
NV	2.1884	2.1779	1.3243 ,	3.0263	-0.0291	-0.0303	-0.1053 ,	0.0475	0.0434	0.0419	-0.0399 ,	0.1231
NY	1.5752	1.5660	1.1122 ,	2.0576					0.0851	0.0849	0.0529 ,	0.1128
OH	2.0223	1.9576	1.5078 ,	2.5498	-0.0551	-0.0557	-0.1058 ,	-0.0125	-0.0353	-0.0330	-0.0839 ,	0.0135
OK	2.4721	2.4915	1.9608 ,	3.0119	-0.0375	-0.0386	-0.0878 ,	0.0111	0.0388	0.0412	-0.0164 ,	0.0923
OR	2.3696	2.2990	1.6972 ,	3.0857	-0.0080	-0.0132	-0.0700 ,	0.0628	-0.0008	-0.0017	-0.0804 ,	0.0821
PA	2.1454	2.1410	1.7926 ,	2.5518	-0.0649	-0.0655	-0.1012 ,	-0.0226	-0.0143	-0.0145	-0.0566 ,	0.0283
RI	1.2585	1.2690	0.6227 ,	1.8859					0.1712	0.1735	0.1336 ,	0.2144
SC	1.8004	1.7805	1.1822 ,	2.3830	-0.0310	-0.0283	-0.0763 ,	0.0135	-0.0060	-0.0035	-0.0583 ,	0.0416
TN	2.0311	2.0160	1.5091 ,	2.6308	-0.0775	-0.0800	-0.1260 ,	-0.0171	-0.0014	-0.0020	-0.0591 ,	0.0550
UT	1.8290	1.7967	1.0440 ,	2.6165	0.0202	0.0191	-0.0300 ,	0.0758	-0.1496	-0.1517	-0.2136 ,	-0.0717
VA	1.7549	1.7406	1.2524 ,	2.2849	-0.0345	-0.0339	-0.0769 ,	0.0117	-0.0095	-0.0116	-0.0534 ,	0.0466
WI	2.2543	2.2377	1.7681 ,	2.7728	-0.0385	-0.0415	-0.0811 ,	0.0075	-0.0220	-0.0264	-0.0812 ,	0.0464
Average	2.0045	2.0010			-0.0359	-0.0363			0.0072	0.0071		

Table 5: Conditions for Implementability

state	*% independents*	*average maximum % independents*	*Pr(implementable)*
AL	29.44%	70.77%	100%
CA	26.01%	73.13%	100%
CO	39.12%	80.16%	100%
CT	41.89%	67.34%	100%
DE	36.20%	76.83%	100%
FL	27.59%	82.25%	100%
IA	40.46%	80.60%	100%
IL	33.64%	72.11%	100%
KS	31.18%	73.56%	100%
KY	22.55%	79.14%	100%
ME	45.71%	75.60%	100%
MI	35.16%	80.19%	100%
MO	37.15%	74.65%	100%
MS	24.94%	63.52%	100%
MT	36.36%	75.27%	100%
NM	29.43%	78.03%	100%
NV	29.78%	77.84%	100%
NY	30.92%	60.60%	100%
OH	32.41%	80.25%	100%
OK	18.79%	75.86%	100%
OR	30.99%	78.82%	100%
PA	24.59%	78.58%	100%
RI	51.19%	58.14%	100%
SC	33.02%	83.00%	100%
TN	33.94%	77.89%	100%
UT	33.91%	70.49%	100%
VA	34.42%	80.60%	100%
WI	36.43%	81.68%	100%

Table 6: Welfare Gains to Optimal Districting

state	*median % welfare gains*	*average % welfare gains*	*confidence interval*	
AL	0.99%	1.11%	0.36% ,	2.27%
CA	1.74%	1.86%	0.72% ,	3.54%
CO	1.21%	1.45%	0.45% ,	3.08%
CT	1.80%	1.82%	1.19% ,	2.53%
DE	2.38%	2.43%	0.77% ,	4.14%
FL	1.20%	1.24%	0.68% ,	2.02%
IA	1.87%	1.90%	1.04% ,	2.99%
IL	1.04%	1.06%	0.47% ,	1.74%
KS	2.52%	2.50%	1.53% ,	3.60%
KY	1.76%	1.76%	0.82% ,	3.08%
ME	2.52%	2.62%	1.84% ,	3.72%
MI	0.91%	0.94%	0.50% ,	1.52%
MO	1.51%	1.50%	0.92% ,	2.11%
MS	2.26%	2.34%	1.08% ,	4.06%
MT	2.62%	2.75%	1.32% ,	4.47%
NM	1.53%	1.73%	0.57% ,	3.23%
NV	1.74%	1.82%	0.70% ,	3.45%
NY	1.37%	1.41%	0.76% ,	2.15%
OH	1.49%	1.60%	0.90% ,	2.57%
OK	2.55%	2.54%	1.53% ,	3.64%
OR	2.00%	2.22%	1.10% ,	3.96%
PA	1.57%	1.61%	1.04% ,	2.31%
RI	4.21%	4.15%	2.67% ,	6.02%
SC	0.78%	0.84%	0.40% ,	1.47%
TN	1.10%	1.15%	0.55% ,	1.89%
UT	3.56%	3.78%	1.98% ,	6.20%
VA	0.93%	1.01%	0.45% ,	1.82%
WI	1.50%	1.55%	0.83% ,	2.54%
Average	1.81%	1.88%		

Table 7: Average Welfare Gains Under Alternative Parameter Values

	$\eta=1$	$\eta=0.75$	$\eta=0.50$	$\eta=0.25$
$\gamma/\beta=1$	1.88%	2.06%	2.25%	2.46%
$\gamma/\beta=0.75$	1.33%	1.46%	1.59%	1.74%
$\gamma/\beta=0.50$	0.84%	0.92%	1.01%	1.10%
$\gamma/\beta=0.25$	0.40%	0.44%	0.48%	0.52%

Table 8: Average Welfare Gains Associated With Moving from Identical to Optimal Districting

state	*% welfare gains (relative to identical districting)*
AL	27.49%
CA	32.30%
CO	35.23%
CT	31.91%
DE	35.32%
FL	38.44%
IA	35.86%
IL	32.33%
KS	34.71%
KY	36.73%
ME	33.82%
MI	35.37%
MO	33.83%
MS	22.39%
MT	33.04%
NM	34.66%
NV	35.67%
NY	19.86%
OH	37.06%
OK	38.20%
OR	36.86%
PA	38.15%
RI	19.29%
SC	37.59%
TN	34.92%
UT	27.80%
VA	36.65%
WI	36.69%
Average	33.29%

Table 9: Average Welfare Gains from PR

state	*% welfare gains to optimal districting from PR*	*% welfare gains to implementing PR*
AL	0.02%	1.09%
CA	0.04%	1.83%
CO	0.02%	1.43%
CT	0.07%	1.75%
DE	0.06%	2.39%
FL	0.03%	1.22%
IA	0.03%	1.88%
IL	0.03%	1.03%
KS	0.05%	2.45%
KY	0.03%	1.73%
ME	0.04%	2.59%
MI	0.02%	0.93%
MO	0.03%	1.47%
MS	0.03%	2.32%
MT	0.02%	2.73%
NM	0.03%	1.71%
NV	0.05%	1.79%
NY	0.02%	1.39%
OH	0.04%	1.56%
OK	0.08%	2.45%
OR	0.05%	2.18%
PA	0.05%	1.56%
RI	0.02%	4.13%
SC	0.02%	0.82%
TN	0.02%	1.13%
UT	0.02%	3.77%
VA	0.03%	0.99%
WI	0.02%	1.53%
Average	0.03%	1.85%