This version June 2005

Socially Optimal Districting*

Abstract

This paper provides a welfare economic analysis of the problem of districting. In the context of a simple micro-founded model intended to capture the salient features of U.S. politics, it studies how a social planner should allocate citizens of different ideologies across districts to maximize aggregate utility. In the model, districting determines the *equilibrium seat-vote curve* which is the relationship between the aggregate vote share of the political parties and their share of seats in the legislature. To understand optimal districting, the paper first characterizes the *optimal seat-vote curve* which describes the ideal relationship between votes and seats. It then shows that under rather weak conditions the optimal seat-vote curve is *implementable* in the sense that there exist districtings which make the equilibrium seat-vote curve equal to the optimal seat-vote curve. The nature of these optimal districtings is described. Finally, the paper provides a full characterization of the *constrained optimal seat-vote curve* and the districtings that underlie it when the optimal seat-vote curve is not achievable.

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

*For helpful comments we thank Tim Besley, Andrew Gelman and seminar participants at UC Berkeley, Princeton and Rutgers.

1 Introduction

Districting plans, which allocate voters across districts for the purpose of electing representatives to a central legislature, are frequently perceived as unfair to voters of certain ideologies or race. These perceptions of unfairness have lead to conflicts over how district lines should be drawn. As computer technology and the information available to officials charged with districting have improved, districting plans have become more refined, and these conflicts between groups of voters have intensified. As a result of these conflicts, courts have become more involved in the process, and independent commissions have been established in some cases to oversee the districting process.

There is little consensus, however, on what types of districting plans are socially desirable. Should all districts be identical in their composition of voter types or should districts be heterogenous? Should all districts be competitive or should some districts be "safe seats"? How should the allocation of seats in the legislature respond to changes in national support for the parties? Should the system be biased in favor of certain groups of voters? In addressing these normative questions, some have advocated an axiomatic approach, which adheres as closely as possible to "traditional districting principles", such as the spatial notions of compactness and contiguity as well as the democratic ideals of respecting political subdivisions and recognizing communities of actual shared interest.[1]

As an alternative to this axiomatic method of evaluating districting plans, this paper explores an approach rooted in traditional welfare economics. This approach begins with the observation that citizens have preferences over policy outcomes, which depend upon the representation of groups of voters in the legislature, which in turn depends upon how different groups of voters are allocated across districts. This induced linkage between citizen preferences and districting plans allows for an explicit characterization of how different groups of voters should be allocated across districts in order to maximize social welfare.

The paper studies a theoretical model of a community divided into political districts each of which elects a single representative to a legislature. There are three types of voters: Democrats, Republicans, and Independents. Democrats and Republicans have fixed ideologies, while Independents' ideologies may vary across elections. There are two political parties, one representing

[1] The U.S. Supreme Court defined traditional districting principles in 1990s redistricting cases, including Shaw vs. Reno and Miller vs. Johnson.

Democrats and the other Republicans. These parties field candidates in each district and the candidates with the most votes are elected. The legislature's policy choices depend upon the average ideology of the elected legislators which in turn depends upon the share of seats each party holds in the legislature. The allocation of voters across districts determines the *equilibrium seat-vote curve* which is the relationship between the aggregate vote share of the two parties and their share of seats. This relationship determines how responsive the legislature's policy choices are to swings in the aggregate vote share created by changes in the ideological leanings of Independents.

In the context of this model, we analyze how the three types of voters should be allocated across districts to maximize social welfare. We approach the problem by first characterizing the *optimal seat-vote curve*, which relates the optimal fraction of Democrats in the legislature to the aggregate fraction of voters supporting Democrat candidates across all districts. Under our assumptions, the optimal relationship between aggregate votes and seats is linear, with a slope that depends on the degree of variation in the preferences of Independents. Interestingly, we also find that the optimal seat-vote curve is biased in favor of the party with the largest partisan base.

We then explore whether this optimal seat-vote curve is *implementable*, in the sense that there exist feasible allocations of Democrat, Republican, and Independent citizens across districts that would make the equilibrium seat-vote curve equal to the optimal seat-vote curve. If so, then such allocations clearly represent socially optimal districtings. We develop simple necessary and sufficient conditions for the optimal seat-vote curve to be implementable. These conditions are in terms of the fractions of the various groups in the community and the Independents' preference parameters. We also describe some of the districtings that generate the optimal seat-vote curve.

While the conditions under which the optimal seat-vote curve is implementable are permissive, there are interesting situations in which they are not satisfied. To characterize optimal allocations of voters across districts in these cases requires a more sophisticated approach. First, we must characterize implementable seat-vote curves - those that can be generated by some feasible districting. Then, we must choose the best of these implementable seat-vote curves. We develop an analytical approach that permits a complete characterization of the shape of the constrained optimal seat-vote curve. We also identify the districtings that generate these constrained optimal seat-vote curves.

Throughout the paper we ignore geographical constraints in the way in which districts may be formed. Thus, we assume that the planner can allocate citizens to districts in any way he likes, rather than requiring districts be connected subsets of some geographic space. While this is certainly a weakness of the analysis, 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, when the optimal seat-vote curve is implementable, we show that it can typically be implemented by a large class of districtings, some of which look quite "straightforward", and hence geographic constraints may actually be easily accommodated.

This paper fits into the growing literature applying contemporary political economy modelling and welfare economic methods to explore the optimal design of political institutions.[2] 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 (2004a) 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.

The organization of the remainder of the paper is as follows. The next section discusses the relationship of the analysis to the existing literature on districting. Section 3 outlines the model and introduces the notion of an equilibrium seat-vote curve. Section 4 introduces the idea of the optimal seat-vote curve and characterizes it. This section also shows that the optimal seat-vote curve is not necessarily implementable. Section 5 describes a general method for determining whether a seat-vote curve is implementable and this is used in Section 6 to find the conditions under which the optimal seat-vote curve is implementable. Section 7 characterizes the constrained optimal seat-vote curve and discusses the districtings that generate it. Section 8 discusses the role of some of the key assumptions of the model and Section 9 concludes with a summary of the lessons of the analysis.

[2] A parallel empirical literature explores the consequences of political institutions for observed policy choices. See Besley and Case (2003) and Persson and Tabellini (2004) for surveys.

2 Relation to the districting literature

There are two strands of political science literature on districting - one empirical and the other theoretical. The main focus of the empirical literature has been on understanding how redistricting in the U.S. States has impacted *partisan bias* and *responsiveness*. In a two-party system, partisan bias and responsiveness are conceptualized in terms of properties of the *seat-vote curve* that a districting generates. The seat-vote curve is formally represented by a function $S(V)$ where V is the aggregate fraction of votes received by (say) the Democrats and S is the fraction of seats in the legislature that they hold. A seat-vote curve exhibits *partisan symmetry* if the fraction of seats that one Party gets with any particular share of the vote is the same as the other Party would receive with the same share. Formally, the condition is that $S(V) = 1 - S(1 - V)$ for all V. A seat-vote curve exhibits *partisan bias* if it deviates from partisan symmetry in a systematic way by giving one Party more seats. The *responsiveness* of a seat-vote curve is measured by the proportionate change in seat share following an increase in vote share. If the seat-vote curve is differentiable, then its responsiveness at vote share V is measured by the derivative $S'(V)$.

A common approach in the literature has been to specify parameterized functional forms for seat-vote curves and estimate them. One popular specification is the *linear seat-vote curve*, which can be written as

$$S(V) = \frac{1}{2} + b + r(V - \frac{1}{2}). \tag{1}$$

The parameter b measures partisan bias and r measures responsiveness. In a well-known study, Tufte (1973) estimates linear seat-vote curves using historical data for the U.K., the U.S., New Zealand and three U.S. States and found that the linear form fits the data well.[3] In an influential series of papers, Gary King and co-authors worked with *bilogit seat-vote curves* of the general form

$$S(V) = \frac{1}{1 + \exp(-b - r \ln \frac{V}{1-V})}. \tag{2}$$

Again, the parameters b and r can be interpreted as measuring bias and responsiveness.[4] This family of curves admits a broad range of possible shapes (see Browning and King (1987) and King

[3] He found, for example, that for New York State in the time period 1934-66 with S measuring democratic seats, r equalled 1.28 and b equalled -0.055.

[4] To see this, consider the equivalent log-odds formulation: $\ln[S(V)/(1 - S(V))] = b + r \ln[V/(1 - V)]$. This makes clear that the responsiveness parameter r determines how changes in votes are translated into changes in seats. Further, note that when voters are equally divided between parties [$V = 0.5$], seats are given by $S = \exp(b)/[1 + \exp(b)]$, and Democrats thus secure a majority (minority) of seats if the bias parameter (b) is positive (negative). While these parameters share the same interpretation as Tufte (1973), the exact formulations are somewhat different.

(1989)). King (1989) developed techniques to estimate bias and responsiveness parameters using only data from a single redistricting period.

Gelman and King (1990), (1994) significantly advance the literature by dispensing with the assumption of a particular functional form for the seat-vote curve. Instead, they specify an underlying statistical model of the district-by-district vote generating process which implies a relationship between expected votes and expected seats. They then develop a procedure for estimating the parameters of this underlying statistical model and explore how the implied relationship between seats and votes is impacted by redistricting. In particular, they study whether cross-state variation in redistricting institutions gives rise to systematically different patterns of change in bias and responsiveness.

While this general line of inquiry strikes us as very interesting, the underlying foundations of the analysis are somewhat opaque. Rather than beginning with a functional form for the seat-vote curve or a statistical model for the vote generating process, it would seem more satisfying to begin with a description of the voters, their political preferences and what is generating the variation in these preferences. A districting, or distribution of voter types across districts, would then imply both a district-specific vote generating process and a seat-vote curve. It would also seem useful to spell out how the composition of the legislature matters for citizens' welfare, so that the normative significance of partisan bias and responsiveness could be assessed.

The theoretical literature has largely focused on understanding how political districts should be crafted with the aim of maximizing a Party's expected seat share. Its motivation has been the purely positive one of shedding light on how partisan redistricting committees might further their political objectives. Important strategies for expected seat maximization are *concentration* - the packing of an opponent's supporters into a few districts - and *dispersion* - the spreading of the remainder thinly over the remaining districts.

Owen and Grofman (1988) present a classic analysis of this problem that incorporates aggregate uncertainty in voters' behavior (see also Gilligan and Matsusaka (1999) and Sherstyuk (1998)). Their model assumes that each district j is characterized by some threshold $\alpha_j \in [-1, 1]$ and that there is some random variable Z such that district j votes for the Party controlling the districting if and only if $\alpha_j < Z$. The districting determines the α_j for each district, but subject to two constraints. First, if $y(\alpha)$ is the fraction of districts with $\alpha_j = \alpha$, it is required that $\sum_\alpha y(\alpha) = 1$. Second, the average value of the thresholds across districts must be zero, so that $\sum_\alpha \alpha y(\alpha) d\alpha = 0$.

The districting problem is to choose the function $y(\alpha)$ to maximize the controlling Party's expected seat share. The solution is very simple: there exists some $\alpha^* > 0$ such that $y(-1) = \alpha^*/(1+\alpha^*)$ and $y(\alpha^*) = 1/(1+\alpha^*)$. Thus, a fraction $\alpha^*/(1+\alpha^*)$ of districts will be overwhelmingly for the opposition, while the complementary fraction will be solidly for the controlling Party.

While the principles emerging from the theoretical literature seem natural, the mapping from the models used to the problem of districting is again somewhat opaque. For example, in Owen and Grofman's formulation it is not clear precisely what the threshold α_j is, nor why the average value of the thresholds across districts must be zero. Moreover, the interpretation of the random variable Z is unclear.

What our paper contributes to both the empirical and theoretical literatures is a micro-founded model for the study of districting questions. The model is simple, but captures important aspects of the U.S. political scene. It permits a clear understanding of the mapping between districtings and seat-vote curves. It also provides a consistent story for why the properties of the seat-vote curve matter for welfare, as is implicitly assumed in the empirical literature. As in Owen and Grofman's theoretical analysis, each district does indeed have a critical threshold and this threshold depends upon the distribution of voter types in the district.[5] The random variable in the model is the distribution of the aggregate vote between the two Parties and this randomness is generated by variation in the ideological attachments of Independent voters.

Some of the same concerns about the empirical literature on seat-vote curves motivate the independent work of Besley and Preston (2005). These authors develop an alternative micro-founded model that generates an equilibrium relationship between seats and votes. They use their model to solve for what the distribution of voter types must be across districts if the equilibrium seat-vote curve is to be of the bilogit form. Their main theoretical point is to show that this distribution, and hence the shape of the seat-vote curve, is a key determinant of Parties' electoral incentives to put in effort on the part of their constituents. They provide empirical evidence in favor of their theory by showing that local government performance in the U.K. is related to the parameters of the local seat-vote curve in the way the theory suggests. Their work therefore suggests a novel theoretical mechanism why the form of the seat-vote curve (and hence districting) matters for citizens' welfare and provides evidence for this. By contrast, our model reflects the

[5] As we will note, however, Owen and Grofman's constraint that the average threshold is zero is not implied by the model.

conventional view that districting matters because it determines which Party gets the most seats and hence the ideological composition of the legislature.

Also in the spirit of this paper is the recent work of Epstein and O'Hallaran (2004) on racial gerrymandering.[6] They seek to understand the allocation of voter types across districts that would maximize the welfare of blacks. Their model 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 Esptein and O'Hallaran's model is simpler than the one presented here 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' positions are fixed.

3 The model

We consider a community divided into n equally sized districts, indexed by $i = 1, ..., n$. Policies are chosen by a legislature consisting of a representative from each district. Each district chooses its representative in an election. The policy outcomes chosen by the legislature depend upon the average ideology of the elected representatives, where ideology is measured on a 0 to 1 scale.[7]

In terms of ideologies, citizens are divided into three groups - Democrats, Republicans, and Independents. Democrats and Republicans have ideologies 0 and 1, respectively. Independents

[6] See also the interesting work of Shotts (2002) who studies the impact of federally mandated majority-minority districts on policy outcomes under the assumptions that districting at the state level is done by partisan districters and that the median voter theorem applies in district level elections.

[7] This assumption should be distinguished from the obvious alternative that the policy outcome chosen by the legislature depends upon the median ideology of the elected representatives. While it is certainly possible to undertake the analysis under the median assumption, it implies that the properties of the seat-vote curve are irrelevant for citizens' welfare over almost of its range and hence makes the problem much less interesting. 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 range is irrelevant. All that matters for welfare is the vote share at which the Democrats become the majority party. In essence, to make sense of the concern in the districting literature over the responsiveness of seat-vote curves one needs to assume something like average legislator ideology matters and this motivates the modelling choice we have made. From a theoretical perspective, whether policy choices reflect the preferences of the median or mean legislator ultimately depends upon the nature of legislative bargaining.

have ideologies that are uniformly distributed on the interval $[m-\tau, m+\tau]$ where $\tau > 0$. Reflecting the fluid nature of these voters' attitudes, the ideology of the median Independent is ex ante uncertain. 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 the ideologies of the Independents are always between those of Democrats and Republicans, while the former guarantees that some Independents lean Democrat and some lean Republican. The fraction of voters in district i who are Democrats, Republicans, and Independents are, respectively, $\pi_D(i)$, $\pi_R(i)$ and $\pi_I(i)$.[8] Let π_D, π_R and π_I denote, respectively, the fraction of voters in the entire community who are Democrats, Republicans, and Independents.

Each district must elect a representative. Candidates are put forward by two political organizations: the Democrat and Republican Parties. Following the *citizen-candidate approach*, candidates are citizens and are characterized by their ideologies (see Besley and Coate (1997) and Osborne and Slivinski (1996)). Each Party must select from the ranks of its membership, so that the Democrat Party always selects a Democrat and the Republican Party a Republican.[9] Elections are held simultaneously in each of the n districts and the candidate with the most votes wins. If the average ideology of the elected representatives is α', a citizen with ideology α experiences a payoff given by $-(\alpha - \alpha')^2$. Thus, citizens have quadratic loss functions.[10]

In each district, every citizen votes sincerely for the representative whose ideology is closest to his own.[11] Accordingly, if the median independent has ideology m, the fraction of voters in

[8] Note for future reference that since $\pi_R(i) = 1 - \pi_I(i) - \pi_D(i)$ the allocation of voters in district i is fully described by the pair $(\pi_D(i), \pi_I(i))$.

[9] This assumption substantially simplifies the problem because it means that Parties have no strategic choices to make as regards candidates. It would be interesting to extend the model to allow Parties some flexibility in candidate choice, perhaps by assuming that Democrats and Republicans come in varying ideologies (as in Coate (2004b)). Districting would then shape the incentives for Parties to put up moderate or extreme candidates. It is important to note, however, that going all the way to a Downsian vision of political competition in which candidates adopt the ideology that makes them most likely to win would devoid the problem of much of its content. In each district both Parties' candidates would adopt the position of the expected median voter and which candidate won would have no signifcance for welfare. Thus, while the problem of optimal districting could still be posed, the seat-vote curve and the ideas of partisan bias or responsiveness would cease to have much meaning.

[10] The roles of this assumption and the assumption that the ideologies of the Independents are uniformly distributed are discussed in Section 8.

[11] This is an assumption. An Independent voter who leans Democrat may be better off if his district elects a Republican if other districts disproportionately elect Democrats. For the average legislator ideology would be closer to his ideal point if his district elected a Republican. As an empirical matter, however, 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 the policy outcome depends upon a weighted average of the ideologies of the median congressman and the president (Alesina and Rosenthal (1995) and Fiorina (1992)). While it is certainly the case that some voters do "split their tickets", Degan and Merlo (2004) estimate that the vast majority (82%-93%) vote sincerely.

district i voting for the Democrat is

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

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

$$V(m) = \pi_D + \pi_I[\frac{1/2 - (m - \tau)}{2\tau}]. \tag{4}$$

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

We can now use the model to derive the equilibrium relationship between seats and the aggregate Democratic vote share. First, for all $V \in [\underline{V}, \overline{V}]$, 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 (4), we have that

$$m(V) = \frac{1}{2} + \tau[\frac{\pi_I + 2\pi_D - 2V}{\pi_I}]. \tag{5}$$

Substituting this into (3), we obtain

$$V(i;m(V)) = \pi_D(i) + \pi_I(i)[\frac{V - \pi_D}{\pi_I}]. \tag{6}$$

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

$$V \geq V^*(i) = \pi_D + \pi_I[\frac{1/2 - \pi_D(i)}{\pi_I(i)}], \tag{7}$$

where $V^*(i)$ is the critical aggregate vote threshold above which district i elects a Democrat. It is natural to say that 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 called *competitive.*

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

$$S(V) = \frac{Max\{i : V^*(i) \leq V\}}{n}. \tag{8}$$

This is the *equilibrium seat-vote curve.* It is determined by the allocation of citizens across districts $(\pi_D(i), \pi_I(i))_{i=1}^n$ which determine their critical vote thresholds $(V^*(i))_{i=1}^n$.[12] Note also that the

[12] It is worth noting that this model offers partial micro-foundations for the assumptions made in Owen and Grofman's (1988) analysis of optimal partisan districting discussed in section 2. The vote threshold in district i is $V^*(i)$ and the random variable is V - the aggregate Democrat vote share. Moreover, districting determines the vote thresholds across districts. However, the average value of the thresholds $\sum_{i=1}^n V^*(i)/n$ is not constant across districtings as Owen and Grofman's analysis assumes it must be. We have that $\sum_{i=1}^n V^*(i) = \pi_D + \pi_I \sum_{i=1}^n [\frac{1/2-\pi_D(i)}{\pi_I(i)}]$ and all we know is that $\sum_{i=1}^n \pi_I(i)/n = \pi_I$ and that $\sum_{i=1}^n \pi_D(i)/n = \pi_D$. Thus, their characterization of optimal partisan districtings cannot be applied to this model.

average ideology of the elected representatives is $1 - S(V)$.

4 The optimal seat-vote curve

We are interested in the problem of a planner who must choose how to allocate citizens across the districts to maximize aggregate utility. The districting matters for welfare because, as just demonstrated, it determines the equilibrium relationship between aggregate votes and the composition of the legislature - the equilibrium seat-vote curve. It is important to note, however, 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 (7), the same pattern of critical vote thresholds could in principle be achieved by many different districtings. Thus, the problem is not as simple as writing welfare as a function of the allocation of citizens and choosing the best such allocation.

To solve the problem, we need to think of the planner as choosing the seat-vote curve but subject to the constraint that it be an equilibrium for some districting. The optimal districtings will then be those that are associated with the constrained optimal seat-vote curve. However, this is a hard problem, because of the difficulties in formalizing the constraint that a seat-vote curve be an equilibrium for some districting. Accordingly, we will begin our analysis by characterizing the optimal relationship between seats and aggregate votes - the optimal seat-vote curve - ignoring the constraint that it be an equilibrium for some districting. We will then investigate whether there exist allocations of voters 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 the insights that it yields will make the problem more manageable.

Consider then the problem of the planner deciding on the number of seats S that should be allocated to the Democrats when their vote share is V given that the resulting policy outcome will be $1 - S$. Aggregate utility when the median Independent has ideology m and the Democrats have seat share S is given by:

$$W(S;m) = -[\pi_D(1-S)^2 + \pi_R S^2 + \pi_I \int_{m-\tau}^{m+\tau} (1-S-x)^2 \frac{dx}{2\tau}]. \tag{9}$$

If the vote share is V, the median Independent has ideology $m(V)$ and hence the optimal seat share is

$$S^o(V) = \arg\max_{S \in \{\frac{i}{n}\}} W(S; m(V)). \tag{10}$$

To avoid tedious integer concerns, assume that the number of districts is very large, so that we can interpret S as the fraction of seats held by the Democrats and treat the choice set in the optimization problem as the unit interval $[0,1]$. Then, $S^o(V)$ satisfies the following first order condition:

$$\partial W(S^o; m(V))/\partial S = 0. \tag{11}$$

Solving this first order condition allows us to establish the following result[13] :

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

$$S^o(V) = 1/2 + (\pi_D - \pi_R)(1/2 - \tau) + 2\tau(V - 1/2). \tag{12}$$

Recalling our discussion in section 2, Proposition 1 tells us that the optimal seat-vote curve is linear, with bias $(\pi_D - \pi_R)(1/2 - \tau)$ and responsiveness 2τ. This 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, under this optimal system, 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 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 $dm/dV = -2\tau/\pi_I$ and hence the increase in the optimal Democrat seat share is just

[13] The proofs of this and all subsequent propositions can be found in the Appendix.

2τ. Recall that τ measures the diversity of views among Independents, so that responsiveness is positive 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 (5) 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 is greater than 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 Independents.

Having understood the nature of the optimal seat-vote curve, we must tackle the question of implementability; that is, whether there exist districtings which generate an optimal relationship between seats and votes. 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.[14]

[14] In this case, the optimal seat-vote curve is $S^o(V) = 1/2 + 2\tau(V - 1/2)$, while the equilibrium seat-vote curve

This discussion leaves us with two general questions: first, what are the conditions under which the optimal seat-vote curve is implementable? Second, when it is not implementable, what does the "constrained" optimal seat-vote curve look like? The remainder of the paper is devoted to answering these questions.

5 Determining when a seat-vote curve is implementable

In this section, we outline a method for determining whether a particular seat-vote curve is implementable. This method will not only allow us to understand when the optimal seat-vote curve is implementable, but also how to specify the constraints for the problem of choosing the constrained optimal seat-vote curve.[15]

In developing this method, it is more convenient to work with inverse seat-vote curves rather than seat-vote curves. 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 aggregate 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 $S(V) = i$. In the event that $S(V)$ is flat over some part of its range, we let $V^*(i)$ be the smallest value of V such that $S(V) = i$ and, in the case in which $S(V)$ is discontinuous and there does not exist a V such that $S(V) = i$, we let $V^*(i)$ be the smallest value of V such that $S(V) \geq i$. The relationship between a seat-vote curve and its inverse is illustrated in Figure 2.

We will need the following definitions. A *districting* is a description of the fractions of voter types in each district $\{(\pi_D(i), \pi_I(i)) : i \in [0, 1]\}$. It must be the case that for all i, $(\pi_D(i), \pi_I(i))$ belongs to the two dimensional unit simplex Δ^2_+. This ensures that $\pi_D(i)$ and $\pi_I(i)$ are non-negative and satisfy the constraint that $\pi_D(i)+\pi_I(i) \leq 1$. The latter guarantees that the associated fraction of Republicans in the district $\pi_R(i) = 1 - \pi_D(i) - \pi_I(i)$ is non-negative. A districting $\{(\pi_D(i), \pi_I(i)) : i \in [0, 1]\}$ is *feasible* if it is the case that the average fractions of voter types equal the actual; i.e., $\int_0^1 \pi_I(i)di = \pi_I$ and $\int_0^1 \pi_D(i)di = \pi_D$. Notice that this definition of feasibility

is $S(V) = 0$ if $V < 1/2$ and $S(V) = 1$ if $V > 1/2$.

[15] The reader anxious to see the conditions under which the optimal seat-vote curve is implementable and/or what the constrained optimal seat-vote curve looks like, can jump ahead to the Propositions in Sections 6 and 7 with little loss of continuity.

neglects any geographic constraints on districting.

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 aggregate 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.

Consider then a particular seat-vote curve $S(V)$ with inverse $\{\underline{i}, \overline{i}, V^*(\cdot)\}$. We want to know if it is implementable. We assume only that $S(V)$ is piecewise continuously differentiable and non-decreasing. This allows $S(V)$ to have both jumps and flat spots.[16] These properties will also be shared by the function $V^*(\cdot)$.

We begin by describing the districtings that can generate the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$. 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 $i \in [0, \underline{i})$ and $(\pi_D(i), \pi_I(i)) = (\overline{\pi}_D, \overline{\pi}_I)$ for all $i \in (\overline{i}, 1]$ where $(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I) \in \Delta^2_+$.[17] Using the definitions of $\underline{V}$ and $\overline{V}$ (see (4)), requirements (i) and (ii) from above imply that

$$\underline{\pi}_D + \underline{\pi}_I[\frac{\tau - \varepsilon}{2\tau}] \geq \frac{1}{2} \tag{13}$$

and

$$\overline{\pi}_D + \overline{\pi}_I[\frac{\tau + \varepsilon}{2\tau}] \leq \frac{1}{2}. \tag{14}$$

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 $[\underline{i}, \overline{i}]$, 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, $\pi_D(i) = f(\pi_I(i), V^*(i))$ where

$$f(x, y) = \frac{1}{2} - \frac{x}{\pi_I}(y - \pi_D). \tag{15}$$

[16] By piecewise continuously differentiable we mean that $S(V)$ is continuously differentiable except possibly at a finite number of points. Thus, if $S(V)$ has jumps, it has only a finite number.

[17] For example, if $(\pi_D(i), \pi_I(i))$ varied over the safe Democrat seats $i \in [0, \underline{i})$, 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 $(\pi_D(i), \pi_I(i)) = (\int_0^{\underline{i}} \pi_D(i) \frac{di}{\underline{i}}, \int_0^{\underline{i}} \pi_I(i) \frac{di}{\underline{i}})$ for all $i \in [0, \underline{i})$.

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

$$\pi_I(i) \in [0, \min\{\frac{\pi_I}{2(V^*(i) - \pi_D)}; \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))}\}]. \tag{16}$$

Notice that $V^*(i) - \pi_D$ is less than $\pi_I + \pi_D - V^*(i)$ if and only if $V^*(i)$ is less than $\frac{\pi_I}{2} + \pi_D$. Thus, letting $\widehat{i}$ be such that $V^*(i) \leq \frac{\pi_I}{2} + \pi_D$ for all $i \in [\underline{i}, \widehat{i})$ and $V^*(i) \geq \frac{\pi_I}{2} + \pi_D$ for all $i \in (\widehat{i}, \overline{i}]$, we can write the constraint as[18]

$$\pi_I(i) \in \begin{cases} [0, \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))}] & \text{if } i < \widehat{i} \\ [0, \frac{\pi_I}{2(V^*(i) - \pi_D)}] & \text{otherwise} \end{cases}. \tag{17}$$

We conclude from this that the districtings that generate the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ 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 (13) and (14) and $\pi_I(i)$ satisfies (17) for all $i \in [\underline{i}, \overline{i}]$. We call this the *set of generating districtings* and denote it by $G(\underline{i}, \overline{i}, V^*(\cdot))$. The question of implementability is whether there exists a districting in this set which is feasible; i.e., which satisfies

$$\underline{i}\underline{\pi}_I + (1 - \overline{i})\overline{\pi}_I + \int_{\underline{i}}^{\overline{i}} \pi_I(i) di = \pi_I \tag{18}$$

and

$$\underline{i}\underline{\pi}_D + (1 - \overline{i})\overline{\pi}_D + \int_{\underline{i}}^{\overline{i}} f(\pi_I(i), V^*(i)) di = \pi_D. \tag{19}$$

How do we know when this is true? The following observation is key to the method that we use. Let $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ 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., (18)). Then we have:

Lemma 1: *Let* $\{(\underline{\pi}^o_D, \underline{\pi}^o_I), (\overline{\pi}^o_D, \overline{\pi}^o_I), \pi^o_I(i)\}$ *and* $\{(\underline{\pi}^1_D, \underline{\pi}^1_I), (\overline{\pi}^1_D, \overline{\pi}^1_I), \pi^1_I(i)\}$ *be two districtings in the set* $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ *such that*

$$\underline{i}\underline{\pi}^o_D + (1 - \overline{i})\overline{\pi}^o_D + \int_{\underline{i}}^{\overline{i}} f(\pi^o_I(i), V^*(i)) di \geq \pi_D \geq \underline{i}\underline{\pi}^1_D + (1 - \overline{i})\overline{\pi}^1_D + \int_{\underline{i}}^{\overline{i}} f(\pi^1_I(i), V^*(i)) di.$$

Then there exists a feasible districting in the set $G(\underline{i}, \overline{i}, V^*(\cdot))$.

[18] There will exist such an $\widehat{i}$ whenever there are safe seats for both Parties. If $V^*(i) = \frac{\pi_I}{2} + \pi_D$ for a set of districts, then $\widehat{i}$ can be any element of this set. If $\underline{i} = 0$ and $V^*(0) > \frac{\pi_I}{2} + \pi_D$, let $\widehat{i} = 0$, while if $\overline{i} = 1$ and $V^*(1) < \frac{\pi_I}{2} + \pi_D$, let $\widehat{i} = 1$.

Thus, if there exists two districtings in the set $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ one of which involves a higher average fraction of Democrats than there are in the population and one of which involves a lower fraction, then there must exist a feasible districting in $G^*(\underline{i}, \overline{i}, V^*(\cdot))$.

Consider now the following pair of optimization problems:

$$\min \underline{i}\underline{\pi}_D + (1 - \overline{i})\overline{\pi}_D + \int_{\underline{i}}^{\overline{i}} f(\pi_I(i), V^*(i))di \qquad P_{\min}$$
$$s.t.\ \{(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I), \pi_I(i)\} \in G^*(\underline{i}, \overline{i}, V^*(\cdot))$$

and

$$\max \underline{i}\underline{\pi}_D + (1 - \overline{i})\overline{\pi}_D + \int_{\underline{i}}^{\overline{i}} f(\pi_I(i), V^*(i))di \qquad P_{\max}$$
$$s.t.\ \{(\underline{\pi}_D, \underline{\pi}_I), (\overline{\pi}_D, \overline{\pi}_I), \pi_I(i)\} \in G^*(\underline{i}, \overline{i}, V^*(\cdot))$$

The minimization problem selects the districting in $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ that has the minimal fraction of Democrats, while the maximization problem selects the districting that has the maximal fraction of Democrats or, equivalently, the minimal fraction of Republicans. Letting the values of these problems be $\underline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$ and $\overline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$ respectively, it follows from Lemma 1 that there exists a feasible districting generating $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ if and only if $\underline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot)) \leq \pi_D \leq \overline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$. Thus, the seat-vote curve $S(V)$ is implementable if and only if π_D lies between $\underline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$ and $\overline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$.

6 When is the optimal seat-vote curve implementable?

Given the method just outlined, to see whether the optimal seat-vote curve $S^o(V)$ is implementable we proceed as follows. First, we find the associated inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(\cdot)\}$. Next, we find the values of the associated minimization and maximization problems $\underline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot))$ and $\overline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot))$. Then, we compare these values with the actual fraction of Democrats π_D. In this way, we establish the following result:

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

$$\pi_I(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})) \leq \pi_D \tag{20}$$

and

$$\pi_I(\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})) \leq 1 - \pi_D - \pi_I = \pi_R. \tag{21}$$

Thus, we need that there be "enough" Republicans and Democrats relative to Independents. This makes good intuitive sense given the discussion of implementability in Section 4. There are several points to note about the coefficient multiplying the fraction of Independents (i.e., $\frac{\varepsilon}{2\tau} + \varepsilon - (\tau + \varepsilon)\ln(1 + \frac{\varepsilon}{\tau})$). First, for all τ, its value of the coefficient converges to zero as ε converges to zero. This means that the optimal seat-vote curve is necessarily implementable when the degree of uncertainty in the identity of the median Independent is sufficiently small. 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 diversity in the ideologies of Independents. Third, and most importantly, for any values of ε and τ satisfying our assumptions, the coefficient is less than 1/2 and hence we have the following useful sufficient condition for the optimal seat-vote curve to be implementable.

Corollary: *The optimal seat-vote curve is implementable if* $\pi_I \leq 2\min\{\pi_D, \pi_R\}$.

According to data from Erikson, Wright and McGuiver (1993), this sufficient condition is satisfied in all but four U.S. States.[19]

When the conditions of Proposition 2 are satisfied, we can use arguments developed in the proof of Proposition 2 to show that the optimal seat-vote curve can always be implemented by a districting of the form

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\underline{\pi}_D, \underline{\pi}_I) & \text{if } i \in [0, \underline{i}_o) \\ \left(\frac{\pi_D + \frac{\pi_I}{2} - i}{\pi_D + \frac{\pi_I}{2} - i + \pi_I\tau}, \frac{\pi_I\tau}{\pi_D + \frac{\pi_I}{2} - i + \pi_I\tau}\right) & \text{if } i \in [\underline{i}_o, \pi_D + \frac{\pi_I}{2}) \\ \left(0, \frac{\pi_I\tau}{i - (\pi_D + \frac{\pi_I}{2}) + \pi_I\tau}\right) & \text{if } i \in [\pi_D + \frac{\pi_I}{2}, \overline{i}_o] \\ (\overline{\pi}_D, \overline{\pi}_I) & \text{if } i \in (\overline{i}_o, 1] \end{cases} \tag{22}$$

The voter allocations in the safe seats $(\underline{\pi}_D, \underline{\pi}_I)$ and $(\overline{\pi}_D, \overline{\pi}_I)$ must satisfy inequalities (13) and (14) and the aggregate feasibility conditions

$$\underline{i}_o\underline{\pi}_I + 2\pi_I\tau\ln(1 + \frac{\varepsilon}{\tau}) + (1 - \overline{i}_o)\overline{\pi}_I = \pi_I \tag{23}$$

and

$$\underline{i}_o\underline{\pi}_D + 2[\pi_I\varepsilon - \pi_I\tau\ln(1 + \frac{\varepsilon}{\tau})] + (1 - \overline{i}_o)\overline{\pi}_D = \pi_D. \tag{24}$$

[19] Of course, just because a person reports that they are a Democrat does not mean that they always vote for the Democrat candidate as the model assumes. Nonetheless, the requirement that the fraction of swing voters is less than twice the fraction of voters who either always vote Democrat or always vote Republican seems permissive.

Under the conditions of Proposition 2, there will exist some $(\underline{\pi}_D, \underline{\pi}_I)$ and $(\overline{\pi}_D, \overline{\pi}_I)$ that satisfy all these requirements.

The allocations of voters in the competitive districts in districtings of this form are of particular interest. They are divided into *Democrat-leaning districts* $(i \in [\underline{i}_o, \pi_D + \frac{\pi_I}{2}))$ and *Republican-leaning districts* $(i \in [\pi_D + \frac{\pi_I}{2}, \bar{i}_o])$. 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 $\pi_D + \frac{\pi_I}{2}$. 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 $\overline{V}$.

In general, not much of interest can be said about the allocation of voters in the safe seats. However, when one of the two conditions in Proposition 2 holds with equality, there is a unique districting (in the class of districtings with homogeneous safe seats) that generates the optimal seat-vote curve. Accordingly, the allocation of voters in the safe seats is tied down uniquely. It will be helpful in understanding constrained optimal seat-vote curves to see what this looks like. Consider the case in which condition (20) holds with equality, so that there are just enough Democrats. Then, $(\underline{\pi}_D, \underline{\pi}_I) = (\varepsilon/(\tau+\varepsilon), \tau/(\tau+\varepsilon))$ and $(\overline{\pi}_D, \overline{\pi}_I) = (0, (\frac{\pi_I}{2} - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau}))/(1-\frac{\pi_I}{2} - \pi_I \varepsilon - \pi_D))$. Thus, the safe Democrat districts are just populated by Democrats and Independents and the safe Republican districts by Republicans and Independents. The fraction of Democrats in the safe Democrat districts is no more than necessary to ensure that the fraction of Democrats and Democrat-favoring Independents always exceeds the fraction of Republican-favoring Independents. Assuming that condition (21) holds as an inequality, the fraction of Republicans in the safe Republican districts is greater than the minimal sufficient level.[20] Thus, there are surplus Republicans in the safe Republican seats.

The districtings of the form described in (22) are extreme in the sense that the competitive

[20] If condition (21) holds as an inequality, then $\frac{\frac{\pi_I}{2} - \pi_I \tau \ln(1+\frac{\varepsilon}{\tau})}{1-\frac{\pi_I}{2} - \pi_I \varepsilon - \pi_D} < \frac{\tau}{\tau+\varepsilon}$.

districts have no voters of one type. It is reasonable to object that such districts are unlikely to be practically feasible when account is taken of geographic constraints. However, 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*

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\underline{\pi}_D, \pi_I) & \text{if } i \in [0, \underline{i}_o) \\ (\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau}, \pi_I) & \text{if } i \in [\underline{i}_o, \bar{i}_o] \\ (\overline{\pi}_D, \pi_I) & \text{if } i \in (\bar{i}_o, 1] \end{cases} \tag{25}$$

if and only if

$$\frac{\pi_I \varepsilon(1-\pi_I) + (\frac{1}{2} - \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau}))\pi_I(\frac{1}{2} - \varepsilon)}{\frac{1}{2} + \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau})} \leq \pi_D \tag{26}$$

and

$$\frac{\pi_I \varepsilon(1-\pi_I) + (\frac{1}{2} - \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau}))\pi_I(\frac{1}{2} - \varepsilon)}{\frac{1}{2} + \pi_I(\frac{1}{2} - \frac{\varepsilon}{2\tau})} \leq 1 - \pi_D - \pi_I = \pi_R. \tag{27}$$

The encouraging point to note is that the conditions of Proposition 3 are not that much more restrictive than those of Proposition 2. Figure 3 illustrates the sets of (π_D, π_I) that satisfy the conditions of Propositions 2 and 3 under the assumption that $\varepsilon = 0.1$ and $\tau = 0.2$. The horizontal axis measures π_I and the vertical axis measures π_D. The two dimensional unit simplex Δ_+^2 is the area below the line connecting the points $(0, 1)$ and $(1, 0)$. The set of (π_D, π_I) that satisfy the conditions of Proposition 3 is the smaller triangular area between the two lines that are closest to each other and the set satisfying the conditions of Proposition 2 is the larger triangular area. The two sets are almost the same.

The competitive districts in districtings of the form described in Proposition 3 can still be divided into Democrat-leaning districts ($i \in [\underline{i}_o, \pi_D + \frac{\pi_I}{2}]$) and Republican-leaning districts ($i \in [\pi_D + \frac{\pi_I}{2}, \bar{i}_o]$). 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 $[1 - \pi_I(\frac{\tau+\varepsilon}{\tau})]/[1 - \pi_I(\frac{\tau-\varepsilon}{\tau})]$.

7 The constrained optimal seat-vote curve

While the optimal seat-vote curve is implementable in a broad class of circumstances, there are interesting situations in which it might not be. For example, according to Erikson, Wright and McGuiver (1993), the sufficient conditions of the Corollary are not satisfied in four New England States, where many voters are not affiliated with either political party. In two states (MA and RI) there are enough Democrats ($\pi_I \leq 2\pi_D$) but too few Republicans ($\pi_I > 2\pi_R$), while these conditions are reversed in two others (NH and VT). In such cases, what does the constrained optimal seat-vote curve look like?

We find the constrained optimal seat-vote curve by solving for the implementable inverse seat-vote curve that maximizes aggregate welfare.[21] Let F^{-1} denote the set of all inverse seat-vote curves $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ that have the property that $V^*(\cdot)$ is piecewise continuously differentiable. In addition, let $EW(\{\underline{i}, \overline{i}, V^*(\cdot)\})$ denote expected aggregate utility under the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$. Then, the problem we solve is

$$\max_{\{\underline{i}, \overline{i}, V^*(\cdot)\} \in F^{-1}} EW(\{\underline{i}, \overline{i}, V^*(\cdot)\}) \qquad P_{con}$$
$$s.t.\ \overline{\Omega}(\{\underline{i}, \overline{i}, V^*(\cdot)\}) \geq \pi_D \ \geq \ \underline{\Omega}(\{\underline{i}, \overline{i}, V^*(\cdot)\}).$$

The constrained optimal seat-vote curve is the seat-vote curve corresponding to the solution of this problem.

When (20) and (21) are satisfied, Proposition 2 implies that the optimal inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(\cdot)\}$ solves this problem and the constraints are not binding. When this is not the case, there are three possibilities. First, only (21) is satisfied and there are *not enough Democrats.* Second, only (20) is satisfied and there are *not enough Republicans.* Finally, neither inequality is satisfied and there are *not enough Democrats or Republicans.* We discuss each of these cases in turn.

7.1 Not enough Democrats

In this case, we are able to establish the following result:

Proposition 4: *Suppose that there are not enough Democrats and let $S^*(V)$ denote the constrained optimal seat-vote curve. (a) If $\pi_D \leq \frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, then $S^*(V) = \pi_D \frac{\tau+\varepsilon}{\varepsilon}$ on the interval*

[21] By an "implementable" inverse seat-vote curve we mean one for which there exists a feasible districting that generates it.

$[\underline{V}, \pi_D + \frac{\pi_I}{2})$ *and* $S^*(V) = S^o(V)$ *on the interval* $[\pi_D + \frac{\pi_I}{2}, \overline{V}]$. *(b) If* $\pi_D > \frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$ *there exists* $\widetilde{V} \in (\underline{V}, \pi_D + \frac{\pi_I}{2})$ *such that: (i)* $S^*(V)$ *is positive, increasing, and strictly convex on the interval* $[\underline{V}, \widetilde{V})$*; (ii)* $S^*(V)$ *is constant on the interval* $[\widetilde{V}, \pi_D + \frac{\pi_I}{2})$*; and (iii)* $S^*(V) = S^o(V)$ *on the interval* $[\pi_D + \frac{\pi_I}{2}, \overline{V}]$.

The result is illustrated in Figure 4. 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 is greater than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$. In the former case, the constrained optimal seat-vote curve is constant on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2})$ and then jumps up discontinuously to equal the optimal seat-vote curve on the interval $[\pi_D + \frac{\pi_I}{2}, \overline{V}]$. The logic of the constrained optimum is to allocate the available Democrats to make as many safe Democrat districts as possible. In the case illustrated in panel (b) the seat-vote curve is first increasing and at an increasing rate. However, at some aggregate vote level between $\underline{V}$ and $\pi_D + \frac{\pi_I}{2}$ the curve becomes flat. It then jumps up discontinuously to equal the optimal seat-vote curve on the interval $[\pi_D + \frac{\pi_I}{2}, \overline{V}]$. It can be shown that as π_D gets larger (holding constant π_I) the point at which the curve flattens ($\widetilde{V}$) moves to the right and, for sufficiently large π_D, equals $\pi_D + \frac{\pi_I}{2}$ and the flat spot disappears.

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 by Independents. In case (a) the shortage of Democrats is dealt with by creating as many safe Democrat seats as possible. This means that the seat-vote curve is non-responsive on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2})$, implying that the divergence between average population and legislator ideology is increasing. In case (b) the seat-vote curve is first increasingly responsive, and then becomes unresponsive. This implies that the divergence between the average population and legislator ideology displays a more complex pattern, first increasing and then decreasing. This counter-intuitive pattern stems from an inherent non-convexity in Problem P_{con} that is discussed in the proof of Proposition 4.

What can be said about the districting underlying the constrained optimal seat-vote curve? In contrast to the situation when the optimal seat-vote curve can be implemented, there is a unique districting (in the class with homogeneous safe seats) generating the constrained optimal seat-vote curve. When π_D is less than $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, this optimal districting is

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\frac{\varepsilon}{\tau+\varepsilon}, \frac{\tau}{\tau+\varepsilon}) & \text{if } i \in [0, \pi_D \frac{\tau+\varepsilon}{\varepsilon}) \\ (0,1) & \text{if } i \in [\pi_D \frac{\tau+\varepsilon}{\varepsilon}, \pi_D + \frac{\pi_I}{2}) \\ (0, \frac{\pi_I \tau}{i-(\pi_D+\frac{\pi_I}{2})+\pi_I \tau}) & \text{if } i \in [\pi_D + \frac{\pi_I}{2}, \bar{i}_o] \\ (0, \frac{\frac{\pi_I}{2}-\pi_I \tau \ln(1+\frac{\varepsilon}{\tau})}{1-\pi_D-\frac{\pi_I}{2}-\pi_I \varepsilon}) & \text{if } i \in (\bar{i}_o, 1] \end{cases}. \tag{28}$$

It is instructive to compare this with the optimal districting when there are "just enough Democrats" discussed in Section 6. The safe Democrat districts look exactly the same, but there are less of them since $\pi_D(\tau+\varepsilon)/\varepsilon$ is smaller than $\underline{i}_o$. However, the Democrat-leaning competitive districts from (22) have been replaced by a group of districts ($i \in [\pi_D \frac{\tau+\varepsilon}{\varepsilon}, \pi_D + \frac{\pi_I}{2}]$) that are populated solely by Independents. These districts all vote in the same way and elect a Democrat candidate if and only if the median Independent votes Democrat or, equivalently, if the aggregate vote share for the Democrats exceeds $\pi_D + \frac{\pi_I}{2}$. This is what generates the discontinuity in the seat-vote curve illustrated in Figure 4(a). The Republican-leaning competitive districts and the safe Republican districts look the same as in the districting described by (22).

In the case in which π_D exceeds $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, the optimal districting is more complicated. There exists scalars $\underline{i}$, i^* and a function φ defined on $[\underline{i}, i^*]$ such that

$$(\pi_D(i), \pi_I(i)) = \begin{cases} (\frac{\varepsilon}{\tau+\varepsilon}, \frac{\tau}{\tau+\varepsilon}) & \text{if } i \in [0, \underline{i}) \\ (\frac{\pi_I/2+\pi_D-\varphi(i)}{\pi_I+\pi_D-\varphi(i)}, \frac{\pi_I/2}{\pi_I+\pi_D-\varphi(i)}) & \text{if } i \in [\underline{i}, i^*) \\ (0,1) & \text{if } i \in [i^*, \pi_D + \frac{\pi_I}{2}) \\ (0, \frac{\pi_I \tau}{i-(\pi_D+\frac{\pi_I}{2})+\pi_I \tau}) & \text{if } i \in [\pi_D + \frac{\pi_I}{2}, \bar{i}_o) \\ (0, \frac{\frac{\pi_I}{2}-\pi_I \tau \ln(1+\frac{\varepsilon}{\tau})}{1-\pi_D-\frac{\pi_I}{2}-\pi_I \varepsilon}) & \text{if } i \in [\bar{i}_o, 1] \end{cases} \tag{29}$$

where $\varphi(i)$ is increasing, strictly concave and satisfies $\varphi(\underline{i}) = \underline{V}$. Again, the safe Democrat districts look exactly the same as when there are just enough Democrats, but there are less of them since $\underline{i}$ is smaller than $\underline{i}_o$. The Democrat-leaning competitive districts from (22) are now replaced by two groups of districts. One group ($i \in [\underline{i}, i^*)$) contains both Democrats and Independents. This group have differing critical vote thresholds, with the fraction of Independents increasing from $\tau/(\tau+\varepsilon)$ to $\pi_I/2(\pi_I+\pi_D-\varphi(i^*))$. Accordingly, the fraction of these districts electing a Democrat candidate varies smoothly with the aggregate Democrat vote. However, in contrast to the case where the optimal seat-vote curve is implementable, the critical vote threshold (which is $\varphi(i)$) increases at a decreasing rate in i as opposed to a linear rate. This generates a strictly convex seat-vote curve. The other group of districts ($i \in [i^*, \pi_D + \frac{\pi_I}{2})$) are populated solely by Independents as in the earlier case. Notice that the aggregate vote level $\widetilde{V}$ described in Proposition 4 part (b) is $\varphi(i^*)$. When $i^* = \pi_D + \frac{\pi_I}{2}$, the group of districts populated solely by Independents disappears and $\varphi(i^*) = \pi_D + \frac{\pi_I}{2}$.

7.2 Not enough Republicans

The properties of the constrained optimal seat-vote curve when there are not enough Republicans, can be deduced from Proposition 4. As noted in the proof of Proposition 2, one can redefine the seat-vote curve as representing the relationship between the fraction of seats held by the Republican Party and its share of the aggregate vote. Such a *Republican seat-vote curve* is denoted by $S_R(V_R)$, where S_R is the fraction of seats held by Republicans and V_R is the fraction of votes they received. One can then apply Proposition 4 to deduce the properties of the constrained optimal Republican seat-vote curve $S_R^*(V_R)$ when there are not enough Republicans. Finally, one can use the fact that $S^*(V) = 1 - S_R^*(1 - V)$ to find the properties of the constrained optimal Democrat seat-vote curve. In this way, the following result can be established:

Proposition 5: *Suppose that there are not enough Republicans and let $S^*(V)$ denote the constrained optimal seat-vote curve. (a) If $\pi_R \leq \frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$, then $S^*(V) = S^o(V)$ on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$ and $S^*(V) = 1 - \pi_R \frac{\tau+\varepsilon}{\varepsilon}$ on the interval $(\pi_D + \frac{\pi_I}{2}, \overline{V}]$. (b) If $\pi_R > \frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$ there exists $\widehat{V} \in (\pi_D + \frac{\pi_I}{2}, \overline{V})$ such that: (i) $S^*(V) = S^o(V)$ on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$; (ii) $S^*(V)$ is constant on the interval $(\pi_D + \frac{\pi_I}{2}, \widehat{V}]$; and (iii) $S^*(V)$ is increasing and strictly concave on the interval $(\widehat{V}, \overline{V}]$.*

This result is illustrated in Figure 5. Panel (a) illustrates the case in which π_R is less than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$ and panel (b) the case in which π_R is greater than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$. In the former case, the constrained optimal seat-vote curve equals the optimal one on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$ and then jumps up discontinuously and flattens out on the interval $(\pi_D + \frac{\pi_I}{2}, \overline{V}]$. Again, the logic of the constrained optimum is to allocate the available Republicans to make as many safe Republican districts as possible. In the latter case, the seat-vote curve equals the optimal one on the interval $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$, jumps up at $\pi_D + \frac{\pi_I}{2}$ and stays constant until $\widehat{V}$. It then starts increasing at a decreasing rate on the interval $[\widehat{V}, \overline{V}]$. As π_R gets larger, $\widehat{V}$ moves to the left and eventually equals $\pi_D + \frac{\pi_I}{2}$.

7.3 Not enough Democrats or Republicans

The optimal districting when there are not enough Democrats allocates all the available Democrats in the districts $[0, \pi_D + \frac{\pi_I}{2})$. Similarly, when there are not enough Republicans, the available Republicans are allocated to the districts $(\pi_D + \frac{\pi_I}{2}, 1]$. Accordingly, when there are neither enough Democrats or Republicans the optimal districting is just an amalgam of the two cases: the Democrats are allocated optimally over the districts $[0, \pi_D + \frac{\pi_I}{2})$ and the Republicans over the districts $(\pi_D + \frac{\pi_I}{2}, 1]$. The corresponding constrained optimal seat-vote curve therefore just pieces together the two distorted ends of the seat-vote curves.

This is illustrated in Figure 6. Panel (a) depicts the case in which both π_D and π_R are smaller than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$ and in panel (b) the case in which both π_D and π_R are greater than $\frac{\pi_I \varepsilon}{2\tau}(1 - \tau - 2\varepsilon)$. In the former case, the constrained optimal seat-vote curve is flat on $[\underline{V}, \pi_D + \frac{\pi_I}{2})$, jumps up discontinuously at $\pi_D + \frac{\pi_I}{2}$ and then is constant on the interval $(\pi_D + \frac{\pi_I}{2}, \overline{V}]$. In the latter case, the seat-vote curve is first increasing and at an increasing rate. However, at aggregate vote level $\widetilde{V}$ the curve becomes constant, jumps up discontinuously at $\pi_D + \frac{\pi_I}{2}$ and stays constant until $\widehat{V}$. It then starts increasing at a decreasing rate on the interval $[\widehat{V}, \overline{V}]$. In the case in which both $\widetilde{V}$ and $\widehat{V}$ equal $\pi_D + \frac{\pi_I}{2}$, the seat-vote curve is S-shaped.

7.4 General lessons

There are three general lessons we can draw concerning the properties of constrained optimal seat-vote curves. The first is that they always have safe seats. When either Democrats or Republicans are in short supply, at least some fraction of them are optimally concentrated together to make safe seats for their party.

The second lesson is that when there is a shortage of one group of partisans, the constrained optimal seat-vote curve is biased toward the party with the largest partisan base, but when there is a shortage of both groups this is not uniformly the case. Consider first the case in which there is a shortage of one group - say, Republicans. The constrained optimal seat-vote curve is biased in favor of the Democrats if for all V we have that $S^*(V) > 1 - S^*(1 - V)$.[22] The optimal seat-vote curve is biased in favor of the Democrats in this case, so that $S^o(V) > 1 - S^o(1 - V)$.

[22] To be more precise, the inequality must hold for all $V \in [\underline{V}, \overline{V}]$ such that $1 - V \in [\underline{V}, \overline{V}]$.

Moreover, from Figure 5, the constrained optimal seat-vote curve lies on or above the optimal seat-vote curve in this case, so that $S^*(V) \geq S^o(V)$ and $S^*(1-V) \geq S^o(1-V)$. Combining these inequalities yields the result.

By contrast, consider a case in which there are too few Republicans and Democrats. Assume that both π_D and π_R are smaller than $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$ so that we have the situation illustrated in Figure 6(a) and suppose that π_D is larger than π_R. Consider the situation in which exactly one half the population vote for the Democrats so that $V = 1/2$. Then, since π_D is larger than π_R, it must be the case that $1/2 < \pi_D + \pi_I/2$, implying that the Democrats' seat share is $S^*(1/2) = \pi_D(\tau+\varepsilon)/\varepsilon$ (see Figure 6(a)). In order for $S^*(1/2) > 1 - S^*(1/2)$, it must be the case that $S^*(1/2)$ exceeds $1/2$. But this will not be the case whenever $\pi_D < 2\varepsilon/(\tau+\varepsilon)$. The difficulty that arises here is that because π_D is larger than π_R, it must be the case that the majority of Independents favor the Republicans when $V = 1/2$. Thus all the Independent districts elect a Republican giving the Republicans an advantage in this case. It should be stressed that this anomaly does not arise for all V because whenever V is sufficiently small so that $1-V > \pi_D+\pi_I/2$ the condition that $S^*(V) > 1 - S^*(1-V)$ will be satisfied. But the existence of the anomaly means that the constrained optimal seat-vote curve is not necessarily uniformly biased toward the larger party.

The final lesson is that the responsiveness of the constrained optimal seat-vote curve can be anywhere from zero to infinity. Moreover, as is clear from Figures 4-6, responsiveness can vary discontinuously as one moves along the seat-vote curve. Accordingly, while the notion of *the* optimal degree of responsiveness makes sense for the optimal seat-vote curve, it does not for the constrained optimal seat-vote curve.

8 The role of the assumptions

In analyzing socially optimal districting in our model, we have worked with quite specific assumptions on citizens' political preferences. Specifically, we have assumed that citizens have quadratic loss functions and that the distribution of Independents' ideologies is uniform across its support. To highlight the role these play in the analysis, this section briefly discusses the implications of working with more general assumptions.

With respect to the distribution of Independents' ideologies, the basic model can be generalized by assuming that the fraction of Independents with ideologies less than $x \in [m-\tau, m+\tau]$ is

$H(\frac{x-(m-\tau)}{2\tau})$ where $H : [0,1] \to [0,1]$ is a continuously differentiable, increasing distribution function with a density h that is symmetric around 1/2. This allows us to capture the possibility, say, that there are more Independents with ideologies closer to the median than in the tails of the support. With respect to citizens' loss functions, the model can be generalized by assuming that if the average ideology of the elected representatives is α', a citizen with ideology α experiences a payoff given by $-v(|\alpha - \alpha'|)$ where $v : \Re_+ \to \Re_+$ is increasing, strictly convex, twice continuously differentiable and satisfies $v'(0) = 0$. This allows us to vary the degree of convexity in citizens' loss functions.

It is important to note that these generalizations make little difference to the positive aspects of the analysis. The derivation of the equilibrium seat-vote curve is basically the same and, in particular, the critical aggregate vote threshold for district i is still given by (7).[23] Consequently, the method for determining the implementability of a seat-vote curve outlined in Section 5 generalizes straightforwardly under these assumptions.[24]

Where the generalizations have implications is for the normative analysis; in particular, the form of the optimal seat-vote curve. Consider first the implications of generalizing the distribution of Independents' ideologies. Maintaining the assumption of quadratic loss functions, the optimal seat-vote curve can now be written as:

$$S^o(V) = 1/2 + (\pi_D - \pi_R)(1/2 - \tau) + 2\tau(\pi_I H^{-1}(\frac{V - \pi_D}{\pi_I}) - (1/2 - \pi_D)). \tag{30}$$

This seat-vote curve remains biased in favor of the party with the largest partisan base and its responsiveness continues to depend upon the degree of preference variation among the Independents, as measured by 2τ. However, the responsiveness is $2\tau/h(\frac{1/2-(m(V)-\tau)}{2\tau})$ and thus depends upon the density of Independents' ideologies. Intuitively, this is because the size of this density determines the change in the mean Independent's ideology that is signalled by a marginal increase in the Democrats' vote share. The implication of this is that the optimal seat-vote curve is no longer linear. In particular, under the assumption that the density h is increasing on $[0, 1/2]$, the optimal seat-vote curve will be strictly concave on $[\underline{V}, \pi_D + \frac{\pi_I}{2})$ and strictly convex on $(\pi_D + \frac{\pi_I}{2}, \overline{V}]$.

[23] If the median independent has ideology m, the fraction of voters in district i voting for the Democrat is $V(i;m) = \pi_D(i) + \pi_I(i)H(\frac{1/2-(m-\tau)}{2\tau})$ and the average fraction of voters voting for the Democrat Party is $V(m) = \pi_D + \pi_I H(\frac{1/2-(m-\tau)}{2\tau})$. Accordingly, $m(V) = 1/2 + \tau - 2\tau H^{-1}(\frac{V-\pi_D}{\pi_I})$ and $V(i;m(V)) = \pi_D(i) + \pi_I(i)[\frac{V-\pi_D}{\pi_I}]$ which is just (6). Hence (7) still holds.

[24] The only difference is that equations (13) and (14) become $\underline{\pi}_D + \underline{\pi}_I H(\frac{\tau-\varepsilon}{2\tau}) \geq 1/2$ and $\overline{\pi}_D + \overline{\pi}_I H(\frac{\tau+\varepsilon}{2\tau}) \leq 1/2$.

With the more general loss function, it is no longer possible to obtain a closed form solution for the optimal seat-vote curve. Rather, it is defined implicitly by the first order condition that the social marginal benefit from having more Democrat seats just equals the social marginal cost. Maintaining the uniform distribution assumption, this condition is given by[25]

$$\pi_D v'(1-S^o)+\pi_I\int_{m(V)-\tau}^{1-S^o} v'(1-S^o-\alpha)\frac{d\alpha}{2\tau}=\pi_R v'(S^o)+\pi_I\int_{1-S^o}^{m(V)+\tau} v'(\alpha-(1-S^o))\frac{d\alpha}{2\tau}. \quad (31)$$

Despite the lack of a closed form solution, it is possible to explore the bias and responsiveness of the optimal seat-vote curve. It is straightforward to show that when $V = 1/2$, the optimal Democratic seat share $S^o(V)$ is greater or smaller than $1/2$ as π_D is greater or smaller than π_R, so that the optimal seat-vote curve remains biased towards the party with the largest partisan base. Moreover, it can be shown that there exists $\alpha_D \in (0, 1-S^o(V)-(m(V)-\tau))$ and $\alpha_R \in (0, m(V)+\tau-(1-S^o(V)))$ such that

$$\frac{dS^o(V)}{dV}=2\tau[\frac{(v''(\alpha_D)+v''(\alpha_R))/2}{Ev''}]+[\frac{v''(\alpha_D)-v''(\alpha_R)}{Ev''}](1-S^o(V)-m(v)), \quad (32)$$

where Ev'' is the population average second derivative of the loss function. Thus, the responsiveness of the optimal seat-vote curve still depends upon the degree of preference variation among the Independents, but now also on the behavior of the second derivative of the loss function.

It should now be clear that the role of our assumptions concerning citizens' preferences and the distribution of Independents' ideologies is to ensure that the optimal seat-vote curve has a simple tractable form. This allows us to easily compute the values of the minimization and maximization problems associated with the optimal inverse seat-vote curve $\underline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot))$ and $\overline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot))$ and hence derive the simple conditions for implementability presented in Proposition 2. Moreover, these assumptions permit the characterization of the optimal constrained seat-vote curve by ensuring that the social welfare function $EW(\{\underline{i}, \overline{i}, V^*(\cdot)\})$ has a relatively tractable form. In this sense, our assumptions are key. Nonetheless, they are not misleading because, as just demonstrated, the determinants of partisan bias and responsiveness in the basic model remain in more general models. Thus, the nature of socially optimal districting in more general models is not going to be fundamentally different - it is just that additional considerations will come into play. Obviously, understanding precisely how such considerations impact both the conditions for implementability and the nature of the constrained optimal seat-vote curve is an interesting (and

[25] This assumes that the optimal Democrat seat-share S^o lies between $m(V)-\tau$ and $m(V)+\tau$.

challenging) subject for further research.

9 Conclusion

This paper has developed a welfare economic analysis of the problem of districting. In the context of a simple micro-founded model intended to capture salient features of U.S. politics, it has studied how a social planner should allocate citizens of different ideologies across districts to maximize aggregate utility. Ideally, the social planner would like the Democratic seat share in the legislature to be such that the social gains from marginally increasing it just equal the social costs. Since changes in the parties' aggregate vote share reflect changes in voters' ideologies, the optimal composition of the legislature will depend on the aggregate vote share. This yields the key conceptual innovation of the paper - the optimal seat-vote curve. Under the assumptions of the model, the optimal seat-vote curve is of the same simple linear form estimated in the early empirical literature. Its "responsiveness" depends on the magnitude of the change in average voter ideology signalled by a change in vote share, which in turn depends on the degree of preference variation among Independents. Its "bias" depends on the difference in the fractions of Democrats and Republicans in the population; specifically, it is biased in favor of the party with the largest partisan base.

If there exists a way of districting voters that makes the equilibrium seat-vote curve equal to the optimal seat-vote curve, then the social planner can do no better than to choose such a districting. The first analytical achievement of the paper is to show that there exist such districtings if (and only if) the fraction of Independents in the population is not "too large" relative to either the fraction of Democrats or Republicans. These conditions appear permissive and would be satisfied in the vast majority of U.S. States. Moreover, while the analysis does not take into account the geographical constraints faced by officials charged with redistricting in the real world, the optimal seat-vote curve can typically be generated by districtings that look straightforward to achieve. This nurtures the hope that the optimal seat-vote curve may be an attainable benchmark for districters.

When the fraction of Independents in the population is large, the optimal seat-vote curve will not be implementable even if the planner has the flexibility in allocating voter types that we have assumed. The second analytical achievement of the paper is to fully characterize the constrained optimal seat-vote curve. In contrast to the situation when the first best is implementable, the constrained optimal seat-vote curve is generated by a unique districting. These optimal districtings involve a complex pattern of voter types, with some districts being all Independent and the remainder containing only Independents and Democrats or Independents and Republicans.

While the shape of the constrained optimal seat-vote curve differs from that of the optimal seat-vote curve, they do share several general features, which can be interpreted as lessons for districting practices. While many commentators consider uncompetitive districts to be undesirable from the perspective of democracy, our welfare economic perspective provides general support for a mix of competitive and safe seats. In addition, the analysis provides support for partisan bias as both the optimal and constrained optimal seat-vote curves are typically biased in favor of the party with the largest voter base. Regarding the districtings underlying these seat-vote curves, our analysis provides support for districts that are heterogeneous, rather than identical, in their compositions of voter ideology. While the optimal and constrained optimal systems concur on these issues of the number of safe seats, partisan bias, and cross-district heterogeneity, they differ on the appropriate degree of responsiveness. In particular, while the first-best system has a constant responsiveness, the constrained optimal seat-vote curve exhibits responsiveness that varies from zero to infinity.

The model and techniques developed in this paper can be used to address other districting questions. One could study the classic question of optimal partisan gerrymandering by characterizing the implementable seat-vote curve that maximizes the expected utility of (say) the Democrats. This requires solving a problem similar to that studied in Section 7, except the objective function would be the expected welfare of the Democrats rather than the population at large. This exercise might be useful for developing predictions concerning the districtings that a partisan redistricting committee might choose. The model would also facilitate a precise understanding of the determinants of the welfare loss associated with partisan districting.[26]

The model can also be used as a basis to empirically estimate and evaluate seat-vote curves. Coate and Knight (2005) use the model to develop an empirical methodology for estimating seat-vote curves for the U.S. States and measuring citizen welfare. This allows the comparison of actual and optimal seat-vote curves and the estimation of the welfare loss associated with observed districtings. Given our argument that it may be reasonably easy to achieve the optimal relationship between seats and votes, we might hope this welfare loss to be small. Following

[26] It would also be interesting to explore the determinants of the level of partisan bias under the optimal partisan gerrymander as in Gilligan and Matsusaka [1999].

King (1989) and Gelman and King (1994), one could also investigate the correlation between redistricting institutions and welfare loss.

Finally, it will be clear to the reader that this paper is very much a first cut at the problem and there are numerous ways the model could usefully be extended. First, it would be interesting to see how strategic voting of the sort discussed in the split-ticket voting literature (Alesina and Rosenthal (1995) and Fiorina (1992)) would impact the analysis. Second, it would be highly desirable to be able to incorporate geographic constraints in a meaningful way. Perhaps the most fruitful approach would be to devise a way of studying the welfare consequences of local changes in districting. Third, it would be useful to incorporate a governor or president into the model. Fourth, it would be interesting to make the model dynamic and incorporate incumbency. In reality, incumbents have a significant advantage (perhaps due to greater experience) and, it is often argued that redistricting is done with an eye to preserving the seats of incumbents. Fifth, it would be interesting to give parties a strategic role in terms of candidate selection, perhaps by assuming that they can choose between moderate and extremist candidates.

References

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

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.

Coate, Stephen, [2004a], "Pareto-Improving Campaign Finance Policy," *American Economic Review*, 94, 628-655.

Coate, Stephen, [2004b], "Political Competition with Campaign Contributions and Informative Advertising," *Journal of the European Economic Association*, 2(5), 772-804.

Coate, Stephen and Brian Knight, [2005], "Socially Optimal Districting: An Empirical Analysis," mimeo, Cornell University.

Degan, Arianna and Antonio Merlo, [2004], "Do Citizens Vote Sincerely (If They Vote at All)? Theory and Evidence from U.S. National Elections," 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.

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.

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.

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.

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, [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.

10 Appendix[27]

Proof of Proposition 1: Differentiating (9) 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 (5), we obtain

$$S^o(V) = 1/2 + (\pi_D - \pi_R)(1/2 - \tau) + 2\tau(V - 1/2)$$

as required. *QED*

Proof of Lemma 1: Let

$$\Omega^o = \underline{i}\underline{\pi}_D^o + (1-\overline{i})\overline{\pi}_D^o + \int_{\underline{i}}^{\overline{i}} f(\pi_I^o(i), V^*(i))di$$

and

$$\Omega^1 = \underline{i}\underline{\pi}_D^1 + (1-\overline{i})\overline{\pi}_D^1 + \int_{\underline{i}}^{\overline{i}} f(\pi_I^1(i), V^*(i))di.$$

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

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

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 $G(\underline{i}, \overline{i}, V^*(\cdot))$ and is feasible. *QED*

Proof of Proposition 2: The proof has four parts. In Part I, we develop expressions for the values of the minimization and maximization problems $\underline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$ and $\overline{\Omega}(\underline{i}, \overline{i}, V^*(\cdot))$ associated with an arbitrary inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$. This is more general than we need, but we will use these expressions later in the paper. In Part II, we compute the inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(\cdot)\}$ associated with the optimal seat-vote curve $S^o(V)$. In Part III, we show that the optimal inverse seat-vote curve $\{\underline{i}_o, \overline{i}_o, V_o^*(\cdot)\}$ satisfies the constraint that $\underline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot)) \leq \pi_D$ if

[27] In the interests of brevity, some of the details of the proofs are omitted. Detailed proofs can be found in the version available at http://www.econ.brown.edu/fac/Brian_Knight/optlong.pdf.

and only if (20) holds and in Part IV we show that it satisfies the constraint that $\overline{\Omega}(\underline{i}_o, \overline{i}_o, V_o^*(\cdot)) \geq \pi_D$ if and only if (21) holds.

Part I

Let $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ be an arbitrary inverse seat-vote curve satisfying the requirement that $V^*(\cdot)$ is piecewise continuously differentiable and consider 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 (13), we have that

$$\underline{\pi}_D = \frac{1}{2} - \underline{\pi}_I(\frac{\tau - \varepsilon}{2\tau}). \tag{33}$$

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

$$\overline{\pi}_D = 0. \tag{34}$$

It follows from (34) that we can rewrite (14) as $\overline{\pi}_I \leq \frac{\tau}{\tau+\varepsilon}$. Similarly, (33) 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:

$$\min_{\{\pi_I(i), \overline{\pi}_I, \underline{\pi}_I\}} \int_{\underline{i}}^{\overline{i}} f(\pi_I(i), V^*(i))di + \underline{i}[\tfrac{1}{2} - \underline{\pi}_I(\tfrac{\tau-\varepsilon}{2\tau})] \qquad P_{\min}$$
$$s.t. \;\; \underline{\pi}_I \in [0, \tfrac{\tau}{\tau+\varepsilon}]; \;\; \overline{\pi}_I \in [0, \tfrac{\tau}{\tau+\varepsilon}]; \text{ (17) and (18)}$$

In order for this problem to have a solution, it must be the case that the set $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ 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

$$\pi_I \leq \underline{i}\frac{\tau}{\tau+\varepsilon} + \int_{\underline{i}}^{\widehat{i}} \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))}di + \int_{\widehat{i}}^{\overline{i}} \frac{\pi_I}{2(V^*(i) - \pi_D)}di + (1 - \overline{i})\frac{\tau}{\tau+\varepsilon}. \tag{35}$$

The expression on the right hand side is the fraction of Independents associated with the generating districting that maximizes the use of Independents. We will assume that $\{\underline{i}, \overline{i}, V^*(\cdot)\}$ satisfies this inequality.

To state the value of the minimization problem, it is convenient to introduce some additional notation. Let $\underline{\beta}(\underline{i}, \overline{i}, V^*(\cdot))$ denote the fraction of Independents that would be used up if in each competitive district $i \in [\underline{i}, \widehat{i}]$ $\pi_I(i)$ were set equal to its maximal level; that is,

$$\underline{\beta}(\underline{i}, \overline{i}, V^*(\cdot)) = \int_{\underline{i}}^{\widehat{i}} \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))}di. \tag{36}$$

Similarly, let $\overline{\beta}(\underline{i}, \overline{i}, V^*(\cdot))$ denote the fraction of Independents that would be used up if in each competitive district $i \in [\widehat{i}, \overline{i}]$ $\pi_I(i)$ were set equal to its maximal level; that is,

$$\overline{\beta}(\underline{i}, \overline{i}, V^*(\cdot)) = \int_{\widehat{i}}^{\overline{i}} \frac{\pi_I}{2(V^*(i) - \pi_D)} di. \tag{37}$$

Then we have:[28]

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

$$\underline{\Omega} = \int_{\underline{i}}^{\widehat{i}} (\frac{\pi_I/2 + \pi_D - V^*(i)}{\pi_I + \pi_D - V^*(i)}) di + \underline{i}\frac{\varepsilon}{\tau + \varepsilon}.$$

(ii) If $\pi_I \in [\underline{\beta} + \overline{\beta}, \underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta}]$, *then*

$$\underline{\Omega} = \int_{\underline{i}}^{\widehat{i}} (\frac{\pi_I/2 + \pi_D - V^*(i)}{\pi_I + \pi_D - V^*(i)}) di + \underline{i}[\frac{1}{2} - (\frac{\pi_I - \underline{\beta} - \overline{\beta}}{\underline{i}})(\frac{\tau - \varepsilon}{2\tau})].$$

(iii) If $\pi_I \in [\overline{\beta}, \underline{\beta} + \overline{\beta}]$, *then*

$$\underline{\Omega} = \int_{\underline{i}}^{i^*} \frac{1}{2} di + \int_{i^*}^{\widehat{i}} (\frac{\pi_I/2 + \pi_D - V^*(i)}{\pi_I + \pi_D - V^*(i)}) di + \underline{i}\frac{1}{2}$$

where i^* is defined by

$$\int_{i^*}^{\widehat{i}} \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))} di + \overline{\beta} = \pi_I.$$

(iv) If $\pi_I \in [0, \overline{\beta}]$, *we have that*

$$\underline{\Omega} = \int_{\underline{i}}^{i^{**}} \frac{1}{2} di + \underline{i}\frac{1}{2}$$

where i^{**} is defined by

$$\int_{i^{**}}^{\overline{i}} \frac{\pi_I}{2(V^*(i) - \pi_D)} di = \pi_I.$$

Proof of Lemma A.1: Ignoring the inequality constraints on the choice variables, the Lagrangian for the problem is

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

where λ is the Lagrange multiplier on the aggregate constraint (18). Using the definition of the function $f(\cdot)$ we can write this as

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

[28] To economize on notation and where it will not cause confusion, we will not recognize the dependence of $\underline{\beta}$, $\overline{\beta}$, $\underline{\Omega}$ and $\overline{\Omega}$ on the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$.

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 (18) is satisfied.

We know that

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

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 (18) 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

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

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 (18) be satisfied we need that

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

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

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

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

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 (18) be satisfied we need that

$$x\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta} = \pi_I.$$

Thus, for this to be a solution, it must be that $\pi_I \in [\underline{\beta} + \overline{\beta}, \underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta}]$.

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

Let $i(\lambda)$ denote the value of i at which λ is at least as large as $\frac{V^*(i)-\pi_D}{\pi_I}$ for all $i \in [\underline{i}, i(\lambda)]$ and smaller than $\frac{V^*(i)-\pi_D}{\pi_I}$ for all $i \in (i(\lambda), \overline{i}]$. There are two subcases depending on whether $i(\lambda)$ is greater or less than $\widehat{i}$.

Case 3a: $i(\lambda) \in [\underline{i}, \widehat{i}]$

In this case, the fraction of Independents in the safe Democrat and Republican seats equals zero, so that $\underline{\pi}_I = 0$ and $\overline{\pi}_I = 0$. In the competitive seats,

$$\pi_I(i) = \begin{cases} 0 \text{ if } & i \leq i(\lambda) \\ \frac{\pi_I}{2(\pi_I+\pi_D-V^*(i))} & \text{if } i \in (i(\lambda), \widehat{i}) \\ \frac{\pi_I}{2(V^*(i)-\pi_D)} & \text{if } i \geq \widehat{i} \end{cases}.$$

The value of the multiplier must be such that $i(\lambda)$ satisfies the constraint that

$$\int_{i(\lambda)}^{\widehat{i}} \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))} di + \overline{\beta} = \pi_I,$$

and lies in the interval $[\underline{i}, \widehat{i}]$. Thus, it must be that $\pi_I \in [\overline{\beta}, \underline{\beta} + \overline{\beta}]$.

Case 3b: $i(\lambda) \in [\widehat{i}, \overline{i}]$

In this case, we still have that the fraction of Independents in the safe Democrat and Republican seats equals zero, so that $\underline{\pi}_I = 0$ and $\overline{\pi}_I = 0$, but in the competitive seats,

$$\pi_I(i) = \begin{cases} 0 \text{ if } & i \leq i(\lambda) \\ \frac{\pi_I}{2(V^*(i)-\pi_D)} & \text{if } i > i(\lambda) \end{cases}.$$

The value of the multiplier must be such that $i(\lambda)$ satisfies the constraint that

$$\int_{i(\lambda)}^{\overline{i}} \frac{\pi_I}{2(V^*(i) - \pi_D)} di = \pi_I,$$

and lies in the interval $[\widehat{i}, \overline{i}]$. Thus, it must be that $\pi_I < \overline{\beta}$.

We conclude that: (i) If $\pi_I \in [\underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta}, \underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta} + (1-\overline{i})\frac{\tau}{\tau+\varepsilon}]$, then we are in Case 1 and the solution to the minimization problem is

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

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

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

(iii) If $\pi_I \in [\overline{\beta}, \underline{\beta}+\overline{\beta}]$, then we are in Case 3a and the solution to the minimization problem is

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

where i^* is defined by

$$\int_{i^*}^{\widehat{i}} \frac{\pi_I}{2(\pi_I+\pi_D-V^*(i))} di + \overline{\beta} = \pi_I.$$

(iv) If $\pi_I \in [0, \overline{\beta}]$, then we are in Case 3b and the solution to the minimization problem is

$$\pi_I(i) = \begin{cases} 0 & \text{if } i \in [0, i^{**}) \\ \frac{\pi_I}{2(V^*(i)-\pi_D)} & \text{if } \ i \in [i^{**}, \overline{i}] \\ 0 & \text{if } \ i \in (\overline{i}, 1] \end{cases}.$$

where i^{**} is defined by

$$\int_{i^{**}}^{\overline{i}} \frac{\pi_I}{2(V^*(i)-\pi_D)} di = \pi_I.$$

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 (33), (34), and the fact that $\pi_D(i) = f(\pi_I(i), V^*(i))$ for all $i \in [\underline{i}, \overline{i}]$, imply that

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

Thus, we have that

$$\underline{\Omega} = \int_{\underline{i}}^{\widehat{i}} (\frac{\pi_I/2 + \pi_D - V^*(i)}{\pi_I + \pi_D - V^*(i)})di + \underline{i}\frac{\varepsilon}{\tau + \varepsilon}.$$

This completes the proof of Lemma A.1. ■

To understand the result, recall that the problem is to choose the districting that uses as few Democrats as possible from the set of districtings that both generate the inverse seat-vote curve and satisfy the constraint that the fraction of Independents used equals π_I. Precisely what that districting looks like will depend upon the actual fraction of Independents available. In case (i) of the Lemma, there are a large fraction of Independents available, and it is optimal to set the fractions of Independents in both the safe Democrat and competitive districts ($\underline{\pi}_I$ and $\pi_I(i)$ for all $i \in [\underline{i}, \overline{i}]$) equal to their maximal level, with the remaining Independents allocated to the safe Republican districts. The opposite extreme is case (iv), in which there are only a small fraction of Independents available and it is only in Republican-leaning competitive districts ($i \in [i^{**}, \overline{i}]$) that the fractions of Independents are set equal to their maximal level. In all other districts, the fraction of Independents equals its minimal level - which is 0. Cases (ii) and (iii) lie in between these extremes.

The nature of the solution to the maximization problem can be deduced from the observation that selecting the districting in $G^*(\underline{i}, \overline{i}, V^*(\cdot))$ that has the maximal fraction of Democrats is equivalent to choosing the districting that has the minimal fraction of Republicans. One can alternatively define the seat-vote curve as representing the relationship between the fraction of seats held by the Republican Party and its share of the aggregate vote. Let such a *Republican seat-vote curve* be denoted by $S_R(V_R)$, where S_R is the fraction of seats held by Republicans and V_R is the fraction of votes they received. We can analogously define $\underline{V}_R$ and $\overline{V}_R$ to be the minimal and maximal vote shares received by the Republican Party. Associated with this Republican seat-vote curve, we can define an inverse Republican seat-vote curve $\{\underline{i}_R, \overline{i}_R, V_R^*(\cdot)\}$ and deduce the minimal fraction of Republicans - call it $\underline{\Omega}_R$ - directly from Lemma A.1. The value of the maximization problem will then be given by $\overline{\Omega} = 1 - \pi_I - \underline{\Omega}_R$. The only drawback with this procedure is that the expressions for the value $\overline{\Omega}$ will be in terms of the inverse Republican seat-vote curve $\{\underline{i}_R, \overline{i}_R, V_R^*(\cdot)\}$. However, these expressions are readily converted into ones in terms of the inverse (Democrat) seat-vote curve $\{\underline{i}, \overline{i}, V^*(\cdot)\}$, by noting that $\underline{i}_R = 1 - \overline{i}$, $\overline{i}_R = 1 - \underline{i}$, and $V_R^*(i) = 1 - V^*(1-i)$. In this way, we can establish:

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

$$\overline{\Omega} = 1 - \pi_I - \int_{\widehat{i}}^{\overline{i}} (\frac{V^*(i) - \pi_D - \pi_I/2}{V^*(i) - \pi_D}) di - (1-\overline{i})\frac{\varepsilon}{\tau+\varepsilon}.$$

(ii) If $\pi_I \in [\underline{\beta} + \overline{\beta}, \underline{\beta} + \overline{\beta} + (1-\overline{i})\frac{\tau}{\tau+\varepsilon}]$*, then*

$$\overline{\Omega} = 1 - \pi_I - \int_{\widehat{i}}^{\overline{i}} (\frac{V^*(i) - \pi_D - \pi_I/2}{V^*(i) - \pi_D}) di - (1-\overline{i})[\frac{1}{2} - (\frac{\pi_I - \underline{\beta} - \overline{\beta}}{1-\overline{i}})(\frac{\tau-\varepsilon}{2\tau})]$$

(iii) If $\pi_I \in [\underline{\beta}, \underline{\beta} + \overline{\beta}]$*, then*

$$\overline{\Omega} = 1 - \pi_I - \int_{\widehat{i}}^{i^*} (\frac{V^*(i) - \pi_D - \pi_I/2}{V^*(i) - \pi_D}) di - \int_{i^*}^{\overline{i}} \frac{1}{2} di - (1-\overline{i})\frac{1}{2}$$

where i^* *is defined by*

$$\int_{\widehat{i}}^{i^*} \frac{\pi_I}{2(V^*(i) - \pi_D)} di + \underline{\beta} = \pi_I.$$

(iv) If $\pi_I \in [0, \underline{\beta}]$*, we have that*

$$\overline{\Omega} = 1 - \pi_I - \int_{i^{**}}^{\overline{i}} \frac{1}{2} di - (1-\overline{i})\frac{1}{2}$$

where i^{**} *is defined by*

$$\int_{\underline{i}}^{i^{**}} \frac{\pi_I}{2(\pi_I + \pi_D - V^*(i))} di = \pi_I.$$

Part II

Using the definition of an inverse seat-vote curve and the expression for the optimal seat-vote curve in Proposition 1, we find that the optimal inverse seat-vote curve is given by

$$\underline{i}_o = \pi_D + \pi_I(1/2 - \varepsilon), \tag{38}$$

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

and

$$V_o^*(i) = \frac{[i - (\pi_D + \frac{\pi_I}{2})(1 - 2\tau)]}{2\tau}. \tag{40}$$

Part III

For the optimal inverse seat-vote curve, it is straightforward to show that $\underline{\beta} = \overline{\beta} = \pi_I \tau \ln(1+\frac{\varepsilon}{\tau})$. This means that

$$\underline{\beta} + \overline{\beta} = \pi_I 2\tau \ln(1 + \frac{\varepsilon}{\tau}) \leq \pi_I \ln(2) < \pi_I.$$

Accordingly, only cases (i) and (ii) of Lemma A.1 are possible which simplifies matters. Furthermore, it can be shown that

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

Thus, we can deduce from Lemma A.1 that (a) 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

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

and (b) 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

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

In addition, observe that after substituting in for $\underline{i}_o$ from (38), 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

$$\pi_I \geq \frac{\pi_D}{(1 + \frac{\varepsilon}{\tau})[1 - 2\tau \ln(1 + \frac{\varepsilon}{\tau})] + \varepsilon - \frac{1}{2}} \tag{41}$$

so that case (a) arises if (41) holds and case (b) otherwise.

Suppose that (41) holds so that case (a) arises. Then, after substituting in for $\underline{i}_o$, we have that

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

Thus, in this case, the constraint that $\underline{\Omega} \leq \pi_D$ 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 (20).

Next suppose that (41) does not hold and case (b) arises. Then, after substituting in for $\underline{i}_o$, we have

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

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

$$\frac{\pi_I}{2}(\varepsilon + \frac{\varepsilon}{\tau} - \frac{1}{2}) - \pi_I \varepsilon \ln(1 + \frac{\varepsilon}{\tau}) \leq \frac{\pi_D}{2} \tag{42}$$

To summarize, if (41) holds the constraint $\underline{\Omega} \leq \pi_D$ will be satisfied if and only if (20) is satisfied. If (41) does not hold the constraint that $\underline{\Omega} \leq \pi_D$ will be satisfied if and only if (42) is satisfied.

We can now prove Part III. Suppose first that (20) is not satisfied. This implies that (41) 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 $\underline{\Omega} \leq \pi_D$ will be violated. Next suppose that (20) is satisfied. Then we claim that (42) 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 (41) holds, the constraint $\underline{\Omega} \leq \pi_D$ will be satisfied. This completes Part III.

Part IV

Note first that since $\overline{\Omega} = 1 - \pi_I - \underline{\Omega}_R$, the constraint that $\overline{\Omega} \geq \pi_D$ is equivalent to the constraint that $\pi_R \geq \underline{\Omega}_R$ where $\underline{\Omega}_R$ is the minimized fraction of Republicans defined above. But by applying the argument just presented to the optimal Republican inverse seat-vote curve $\{\underline{i}_{Ro}, \overline{i}_{Ro}, V_{Ro}^*(\cdot)\}$, we can show that $\pi_R \geq \underline{\Omega}_R$ if and only if

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

This completes the proof of Proposition 2. *QED*

Proof of Corollary: We need to show that

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

As already noted, for a given value of ε, the coefficient is decreasing in τ. Thus, since $0 < \varepsilon < \tau$ by assumption, it suffices to show that

$$1/2 + \varepsilon - 2\varepsilon \ln(2) \leq 1/2.$$

This follows from the fact that $2\ln(2) > 1$. *QED*

Proof of Proposition 3: Using the definitions in section 5, the optimal seat-vote curve is implementable with a districting of the form in (25) 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 $i \in [\underline{i}_o, \overline{i}_o]$, $\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau} \in [0, 1-\pi_I]$; and (a.iv)

$$\underline{i}_o \underline{\pi}_D + \int_{\underline{i}_o}^{\overline{i}_o} [\frac{1}{2} - \frac{\pi_I}{2} + \frac{\pi_D + \frac{\pi_I}{2} - i}{2\tau}] di + (1 - \overline{i}_o)\overline{\pi}_D = \pi_D. \tag{43}$$

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

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

so that (43) can be rewritten as

$$\underline{i}_o \underline{\pi}_D + \pi_I \varepsilon (1 - \pi_I) + (1 - \overline{i}_o)\overline{\pi}_D = \pi_D. \tag{44}$$

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 (25) 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 (44). Solving (44), we have that

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

So defining the function:

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

the optimal seat-vote curve is implementable with a districting of the form in (25) 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

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

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

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

which is equivalent to (26). On the other hand, if $\pi_D>\pi_R$, then the condition is

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

which with a little work can be shown equivalent to (27). *QED*

Proof of Proposition 4: The problem we need to solve is

$$\max_{\{\underline{i},\overline{i},V^*(i)\}\in F^{-1}} EW(\{\underline{i},\overline{i},V^*(i)\}) \qquad\qquad P_{con}$$
$$s.t.\ \ \overline{\Omega}(\{\underline{i},\overline{i},V^*(i)\})\geq\pi_D\ \geq\ \underline{\Omega}(\{\underline{i},\overline{i},V^*(i)\}),$$

under the assumption that condition (21) is satisfied but that condition (20) is not. The idea of the proof is to first hope that the constraint that $\overline{\Omega}(\{\underline{i},\overline{i},V^*(i)\})\geq\pi_D$ will not be binding and second substitute in for the expression $\underline{\Omega}(\{\underline{i},\overline{i},V^*(i)\})$ the formula from part (i) of Lemma A.1. The logic for the second step is that when condition (20) is not satisfied, this is the range in which the constraint is violated (see the proof of Proposition 2). Thus, we consider the problem

$$\max_{\{\underline{i},\overline{i},V^*(i)\}\in F^{-1}} EW(\{\underline{i},\overline{i},V^*(i)\}) \qquad P_{conD}$$
$$s.t.\quad \int_{\underline{i}}^{\widehat{i}}(\frac{\frac{\pi_I}{2}+\pi_D-V^*(i)}{\pi_I+\pi_D-V^*(i)})di+\underline{i}\frac{\varepsilon}{\tau+\varepsilon}\leq\pi_D. \tag{45}$$

We will first characterize the solution to this problem and then show that it indeed solves Problem P_{con}.

Before we can do this, however, we must develop an expression for the objective function $EW(\{\underline{i},\overline{i},V^*(i)\})$.

Lemma A.3: *Let $\{\underline{i},\overline{i},V^*(i)\}\in F^{-1}$. Then,*

$$\begin{aligned}EW(\{\underline{i},\overline{i},V^*(i)\}) \;=\;& [\int_{\underline{i}}^{\overline{i}}\{2[i-(\pi_D+\frac{\pi_I}{2})(1-2\tau)]V^*(i)-2\tau V^*(i)^2\}di\\ &+[2\tau\overline{i}\overline{V}^2+2(\pi_D+\frac{\pi_I}{2})(1-2\tau)\overline{i}\overline{V}-\overline{i}^2\overline{V}]\\ &-[2\tau\underline{i}\underline{V}^2+2(\pi_D+\frac{\pi_I}{2})(1-2\tau)\underline{i}\underline{V}-\underline{i}^2\underline{V}]+constant]/[\overline{V}-\underline{V}].\end{aligned}$$

Proof of Lemma A.3: Let $S(V) \in F$ be the seat-vote curve associated with the inverse seat-vote curve $\{\underline{i}, \overline{i}, V^*(i)\}$. Welfare with aggregate votes V is

$$W(S(V); m(V)) = -[\pi_D(1 - S(V))^2 + \pi_R S(V)^2 + \pi_I \int_{m(V)-\tau}^{m(V)+\tau} (1 - S(V) - x)^2 \frac{dx}{2\tau}],$$

which can be rewritten as:

$$W(S(V); m(V)) = -[c(V) + S(V)^2 - 2\pi_D S(V) - 2\pi_I S(V)(1 - m(V))],$$

where

$$c(V) = \pi_D + \frac{\pi_I \tau^2}{3} + \pi_I(1 - m(V))^2.$$

Note that $c(V)$ is independent of the number of seats and hence the seat-vote curve. Using the equation for $m(V)$ given in (5), we can re-write welfare as follows:

$$W(S(V); m(V)) = -[c(V) + S(V)^2 + 2(\pi_D + \frac{\pi_I}{2})(2\tau - 1)S(V) - 4\tau V S(V)].$$

The expected welfare associated with the seat-vote curve $S(V)$ is accordingly given by

$$EW(S(V)) = \int_{\underline{V}}^{\overline{V}} [4\tau V S(V) + 2(\pi_D + \frac{\pi_I}{2})(1 - 2\tau)S(V) - S(V)^2 - c(V)][\frac{dV}{\overline{V} - \underline{V}}]. \tag{46}$$

Now note that

$$\int_{\underline{V}}^{\overline{V}} S(V)dV = \overline{i}\overline{V} - \underline{i}\underline{V} - \int_{\underline{i}}^{\overline{i}} V^*(i)di,$$

$$\int_{\underline{V}}^{\overline{V}} S(V)^2 dV = \overline{i}^2\overline{V} - \underline{i}^2\underline{V} - \int_{\underline{i}}^{\overline{i}} 2iV^*(i)di,$$

and

$$\int_{\underline{V}}^{\overline{V}} S(V)VdV = \frac{\overline{i}}{2}\overline{V}^2 - \frac{\underline{i}}{2}\underline{V}^2 - \int_{\underline{i}}^{\overline{i}} \frac{1}{2}V^*(i)^2 di.$$

Substituting these formulas into (46) yields the result. ■

The first point to note about the solution to Problem P_{conD}, is that it can be shown straightforwardly that $\overline{i}$ and $V^*(i)$ on the range $[\pi_D + \frac{\pi_I}{2}, \overline{i}]$ are exactly as in the unconstrained problem. Thus, we have:

Fact A.1: Let $\{\underline{i}, \overline{i}, V^*(i)\}$ solve Problem P_{conD}. Then, $\overline{i} = \overline{i}_o$ and on the interval $[\pi_D + \frac{\pi_I}{2}, \overline{i}]$, $V^*(i) = V_o^*(i)$.

To understand this intuitively, observe that constraint (45) is independent of $\bar{i}$ and the behavior of the function $V^*(i)$ for $i \geq \widehat{i}$.

Since $V_o^*(\pi_D + \frac{\pi_I}{2}) = \pi_D + \frac{\pi_I}{2}$, it follows from Fact A.1 that we can assume that if $\{\underline{i}, \bar{i}, V^*(i)\}$ solves Problem P_{conD}, then $\widehat{i} = \pi_D + \frac{\pi_I}{2}$. It remains to solve for $\underline{i}$ and the behavior of the function $V^*(i)$ on the range $[\underline{i}, \pi_D + \frac{\pi_I}{2})$. Using Lemma A.3, these must solve the problem:

$$\begin{aligned}
&\max_{\{\underline{i}, V^*(i)\}} \int_{\underline{i}}^{\pi_D + \frac{\pi_I}{2}} \{2[i - (\pi_D + \tfrac{\pi_I}{2})(1-2\tau)]V^*(i) - 2\tau V^*(i)^2\} di \\
&\quad -[2\tau \underline{i}\underline{V}^2 + 2(\pi_D + \tfrac{\pi_I}{2})(1-2\tau)\underline{i}\underline{V} - \underline{i}^2 \underline{V}] \\
&\quad s.t. \;\; \pi_D \geq \int_{\underline{i}}^{\pi_D + \frac{\pi_I}{2}} (\tfrac{\frac{\pi_I}{2} + \pi_D - V^*(i)}{\pi_I + \pi_D - V^*(i)}) di + \underline{i}\tfrac{\varepsilon}{\tau + \varepsilon} \\
&V^*(i) \in [\underline{V}, \pi_D + \tfrac{\pi_I}{2}] \text{ for all } i \in [\underline{i}, \pi_D + \tfrac{\pi_I}{2}) \text{ and } \underline{i} \geq 0.
\end{aligned} \tag{47}$$

The constraint that $V^*(i) \in [\underline{V}, \pi_D + \frac{\pi_I}{2}]$ for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$ is implied by the requirement that $V^*(\cdot)$ must be a non-decreasing function defined on $[\underline{i}, \bar{i}]$ with range $[\underline{V}, \overline{V}]$ given that we know that $V^*(\pi_D + \frac{\pi_I}{2}) = \pi_D + \frac{\pi_I}{2}$. It is not necessary to impose the constraint that $V^*(i)$ be non-decreasing on $[\underline{i}, \pi_D + \frac{\pi_I}{2})$ since it will not bind.

The Lagrangian for the problem is

$$\pounds = \int_{\underline{i}}^{\pi_D + \frac{\pi_I}{2}} h(V^*(i), i, \lambda) di - [2\tau \underline{i}\,\underline{V}^2 + 2(\pi_D + \frac{\pi_I}{2})(1 - 2\tau)\underline{i}\,\underline{V} - \underline{i}^2\,\underline{V}] + \lambda[\pi_D \; - \underline{i}\frac{\varepsilon}{\tau + \varepsilon}]$$

where λ is the Lagrange multiplier and

$$h(V, i, \lambda) = 2[i - (\pi_D + \frac{\pi_I}{2})(1 - 2\tau)]V - 2\tau V^2 - \lambda(\frac{\frac{\pi_I}{2} + \pi_D - V}{\pi_I + \pi_D - V}).$$

Differentiating the Lagrangian with respect to $\underline{i}$, we have that

$$\frac{\partial \pounds}{\partial \underline{i}} = h(\underline{V}, \underline{i}, \lambda) - h(V^*(\underline{i}), \underline{i}, \lambda).$$

Thus, the Kuhn-Tucker condition for $\underline{i}$ is that

$$h(\underline{V}, \underline{i}, \lambda) \leq h(V^*(\underline{i}), \underline{i}, \lambda) \;\; (= \text{if } \underline{i} > 0). \tag{48}$$

In addition, it must be the case that for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$

$$V^*(i) \in \arg\max\{h(V, i, \lambda) : V \in [\underline{V}, \pi_D + \frac{\pi_I}{2}]\}. \tag{49}$$

Before we develop the implications of these conditions, it is useful to note two properties of the function $h(V, i, \lambda)$. The first property is its shape. Differentiating, we have that

$$\frac{\partial h(V; i, \lambda)}{\partial V} = 2[i - (\pi_D + \frac{\pi_I}{2})(1 - 2\tau)] - 4\tau V + \frac{\lambda \pi_I}{2(\pi_I + \pi_D - V)^2},$$

so that

$$\frac{\partial^2 h(V;i,\lambda)}{\partial V^2} = -4\tau + \frac{\lambda\pi_I}{(\pi_I+\pi_D-V)^3}.$$

Notice that for all $V \in [\underline{V}, \pi_D + \frac{\pi_I}{2}]$

$$\frac{\lambda 8}{\pi_I^2}(\frac{\tau}{\tau+\varepsilon})^3 \leq \frac{\lambda\pi_I}{(\pi_I+\pi_D-V)^3} \leq \frac{\lambda 8}{\pi_I^2}.$$

Thus, if $\lambda \in (0, \tau\pi_I^2/2]$ then $h(\cdot, i, \lambda)$ is a strictly concave function on $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$ for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$. On the other hand, if $\lambda \geq (\tau+\varepsilon)^3\pi_I^2/2\tau^2$ then $h(\cdot, i, \lambda)$ is strictly convex. In the intermediate case in which $\lambda \in (\tau\pi_I^2/2, (\tau+\varepsilon)^3\pi_I^2/2\tau^2)$ then, for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$, $h(\cdot, i, \lambda)$ is strictly concave on $[\underline{V}, \pi_D + \pi_I - (\lambda\pi_I/4\tau)^{1/3})$ and strictly convex on $(\pi_D + \pi_I - (\lambda\pi_I/4\tau)^{1/3}, \pi_D + \frac{\pi_I}{2}]$.

The second important property of the function $h(V, i, \lambda)$ is monotonicity. In particular, it is straightforward to show that if $V > V'$, $i > i'$, and $h(V, i', \lambda) \geq h(V', i', \lambda)$, then it must be the case that $h(V, i, \lambda) > h(V', i, \lambda)$.

We can now develop the implications of conditions (48) and (49). Matters are simplified by noting that it can be shown that there is no loss of generality in assuming that $\underline{i} > 0$ and hence (48) can be assumed to hold with equality. We begin by describing what the solution to conditions (48) and (49) must look like for given λ, denoting this by $\{\underline{i}(\lambda), V^*(i;\lambda)\}$.

Suppose first that $\lambda \in (0, \tau\pi_I^2/2]$ so that $h(\cdot, i, \lambda)$ is a strictly concave function on $[\underline{V}, \pi_D + \frac{\pi_I}{2}]$ for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$. Let $\underline{i}(\lambda)$ be such that

$$\frac{\partial h(\underline{V};\underline{i},\lambda)}{\partial V} = 2[\underline{i} - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau\underline{V} + \frac{\lambda\pi_I}{2(\pi_I+\pi_D-V)^2} = 0$$

and let $i^*(\lambda)$ be such that

$$\frac{\partial h(\pi_D + \frac{\pi_I}{2}; i^*, \lambda)}{\partial V} = 2[i^* - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau(\pi_D + \frac{\pi_I}{2}) + \frac{2\lambda}{\pi_I} = 0.$$

It is straightforward to show that $0 < \underline{i}(\lambda) < i^*(\lambda) < \pi_D + \frac{\pi_I}{2}$. For all $i \in [\underline{i}(\lambda), i^*(\lambda)]$, the concavity of $h(\cdot, i, \lambda)$ means that $V^*(i;\lambda)$ is implicitly defined by the first order condition

$$\frac{\partial h(V^*;i,\lambda)}{\partial V} = 2[i - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau V^* + \frac{\lambda\pi_I}{2(\pi_I+\pi_D-V^*)^2} = 0,$$

while for all $i \in (i^*(\lambda), \pi_D + \frac{\pi_I}{2})$ the monotonicity property implies that $V^*(i;\lambda) = \pi_D + \frac{\pi_I}{2}$.

Suppose next that $\lambda \geq (\tau+\varepsilon)^3\pi_I^2/2\tau^2$ so that $h(\cdot, i, \lambda)$ is a strictly convex function for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$. Then, let $\underline{i}(\lambda)$ be such that $h(\underline{V}; \underline{i}, \lambda) = h(\pi_D + \frac{\pi_I}{2}; \underline{i}, \lambda)$, which implies that

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

This will be greater than 0 provided that $\lambda < \pi_I(\frac{\tau+\varepsilon}{\tau})(\pi_D + \frac{\pi_I(1-\varepsilon)}{2})$. Since $h(\cdot, \underline{i}(\lambda), \lambda)$ is a convex function, it is clear that

$$\pi_D + \frac{\pi_I}{2} \in \arg\max\{h(V, \underline{i}(\lambda), \lambda) : V \in [\underline{V}, \pi_D + \frac{\pi_I}{2}]\}$$

and hence we can choose $V^*(\underline{i}(\lambda); \lambda) = \pi_D + \frac{\pi_I}{2}$. Condition (48) is then satisfied by construction. For all $i \in (\underline{i}(\lambda), \pi_D + \frac{\pi_I}{2})$ the monotonicity property implies that $V^*(i; \lambda) = \pi_D + \frac{\pi_I}{2}$.

Finally, consider the intermediate case in which $\lambda \in (\tau\pi_I^2/2, (\tau+\varepsilon)^3\pi_I^2/2\tau^2)$. In this case, there are two possibilities depending on the value of λ. The first possibility is that the solution is exactly as in the convex case; that is, $\underline{i}(\lambda)$ is such that $h(\underline{V}; \underline{i}, \lambda) = h(\pi_D + \frac{\pi_I}{2}; \underline{i}, \lambda)$ and, for all $i \in (\underline{i}(\lambda), \pi_D + \frac{\pi_I}{2})$, $V^*(i; \lambda) = \pi_D + \frac{\pi_I}{2}$. This is the solution if and only if

$$\frac{\partial h(\underline{V}; \underline{i}, \lambda)}{\partial V} = 2[\underline{i}(\lambda) - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau\underline{V} + \frac{\lambda\pi_I}{2(\pi_I + \pi_D - \underline{V})^2} \leq 0.$$

Since

$$2[\underline{i}(\lambda) - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau\underline{V} + \frac{\lambda\pi_I}{2(\pi_I + \pi_D - \underline{V})^2} = \pi_I\varepsilon - \frac{2\lambda\varepsilon\tau}{\pi_I(\tau+\varepsilon)^2},$$

this requires that $\lambda \geq \frac{\pi_I^2(\tau+\varepsilon)^2}{2\tau}$.

For $\lambda < \frac{\pi_I^2(\tau+\varepsilon)^2}{2\tau}$, let $\widetilde{V}(\lambda) \in (\underline{V}, \pi_D + \frac{\pi_I}{2})$ and $i^*(\lambda)$ satisfy the following equations:

$$h(\widetilde{V}; i^*, \lambda) = h(\pi_D + \frac{\pi_I}{2}; i^*, \lambda)$$

and

$$\frac{\partial h(\widetilde{V}; i^*, \lambda)}{\partial V} = 2[i^* - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau\widetilde{V} + \frac{\lambda\pi_I}{2(\pi_I + \pi_D - \widetilde{V})^2} = 0.$$

It should be clear that $\widetilde{V}(\lambda)$ must belong to the region in which $h(\cdot, i^*, \lambda)$ is concave. It follows that

$$\arg\max\{h(V, i^*(\lambda), \lambda) : V \in [\underline{V}, \pi_D + \frac{\pi_I}{2}]\} = \{\widetilde{V}(\lambda), \pi_D + \frac{\pi_I}{2}\}$$

and hence we can choose $V^*(i^*(\lambda); \lambda) = \pi_D + \frac{\pi_I}{2}$. For all $i \in (i^*(\lambda), \pi_D + \frac{\pi_I}{2})$ the monotonicity property implies that $V^*(i; \lambda) = \pi_D + \frac{\pi_I}{2}$. Then, let $\underline{i}(\lambda)$ be such that

$$\frac{\partial h(\underline{V}; \underline{i}, \lambda)}{\partial V} = 2[\underline{i} - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau\underline{V} + \frac{\lambda\pi_I}{2(\pi_I + \pi_D - V)^2} = 0.$$

It is straightforward to show that $0 < \underline{i}(\lambda) < i^*(\lambda)$. For all $i \in [\underline{i}(\lambda), i^*(\lambda))$, the concavity of $h(\cdot, i, \lambda)$ means that $V^*(i; \lambda)$ is implicitly defined by the first order condition

$$\frac{\partial h(V^*; i, \lambda)}{\partial V} = 2[i - (\pi_D + \frac{\pi_I}{2})(1-2\tau)] - 4\tau V^* + \frac{\lambda\pi_I}{2(\pi_I + \pi_D - V^*)^2} = 0.$$

We have now described the solution $\{\underline{i}(\lambda), V^*(i;\lambda)\}$ for any given λ. The value of the multiplier must be such that constraint (45) holds with equality, implying that λ equals $\widehat{\lambda}$ where

$$\underline{i}(\widehat{\lambda})\frac{\varepsilon}{\tau+\varepsilon} + \int_{\underline{i}(\widehat{\lambda})}^{\pi_D+\frac{\pi_I}{2}} (\frac{\frac{\pi_I}{2} + \pi_D - V^*(i;\widehat{\lambda})}{\pi_I + \pi_D - V^*(i;\widehat{\lambda})})di = \pi_D.$$

The solution to the problem described in (47) is then given by $\underline{i}(\widehat{\lambda})$ and $V^*(i;\widehat{\lambda})$.

The next step is to provide conditions that inform us as to the type of solution that will arise.

Fact A.2: Let $\{\underline{i}, \bar{i}, V^*(i)\}$ solve Problem P_{conD}. Then, if $\pi_D \leq \frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$ we have that $\underline{i} = \pi_D \frac{\tau+\varepsilon}{\varepsilon}$ and for all $i \in [\underline{i}, \pi_D + \frac{\pi_I}{2})$, $V^*(i) = \pi_D + \frac{\pi_I}{2}$.

Proof of Fact A.2: To prove this, all we need to show is that under the stated condition, the value of the multiplier $\widehat{\lambda}$ is such that $\widehat{\lambda} \geq \frac{\pi_I^2(\tau+\varepsilon)^2}{2\tau}$. Notice that the proposed solution $\{\underline{i}, V^*(i)\}$ necessarily satisfies the constraint (45). However, we can obtain the value of the multiplier $\widehat{\lambda}$ from the requirement that

$$h(\underline{V}; \pi_D\frac{\tau+\varepsilon}{\varepsilon}, \widehat{\lambda}) = h(\pi_D + \frac{\pi_I}{2}; \pi_D\frac{\tau+\varepsilon}{\varepsilon}, \widehat{\lambda})$$

which implies that

$$\widehat{\lambda} = \frac{\pi_I^2(1-\varepsilon)(\varepsilon+\tau)}{2\tau} - \frac{\tau\pi_I\pi_D(\varepsilon+\tau)}{\varepsilon\tau}.$$

Thus, we need that

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

which is equivalent to

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

■

Fact A.3: Let $\{\underline{i}, \bar{i}, V^*(i)\}$ solve Problem P_{conD}. Then, if $\pi_D > \frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, there exists some $i^* \in (\underline{V}, \pi_D + \frac{\pi_I}{2}]$ such that $V^*(i)$ is increasing and strictly concave on $[\underline{i}, i^*]$ and equal to $\pi_D + \frac{\pi_I}{2}$ thereafter. For π_D sufficiently close to $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, $V^*(i^*)$ will be strictly less than $\pi_D + \frac{\pi_I}{2}$ and hence $V^*(i)$ will be discontinuous at i^*.

Proof of Fact A.3: This follows almost immediately from the above discussion of the properties of the solution. When π_D exceeds $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$ but is close to it, the value of the multiplier $\widehat{\lambda}$ will only be slightly less than $\frac{\pi_I^2(\tau+\varepsilon)^2}{2\tau}$ and $V^*(i^*) = \widetilde{V}(\widehat{\lambda}) < \pi_D + \frac{\pi_I}{2}$. When π_D is much larger than $\frac{\pi_I \varepsilon}{2\tau}(1-\tau-2\varepsilon)$, the value of the multiplier will be less than $\tau\pi_I^2/2$ and $V^*(i^*) = \pi_D + \frac{\pi_I}{2}$. In either case, on the interval $[\underline{i}, i^*]$, $V^*(i)$ is defined by the first order condition $\partial h(V^*; i, \widehat{\lambda})/\partial V = 0$ implying that

$$\frac{dV^*}{di} = \frac{-\frac{\partial^2 h(V^*;i,\widehat{\lambda})}{\partial V \partial i}}{\frac{\partial^2 h(V^*;i,\widehat{\lambda})}{\partial V^2}} = \frac{2}{\frac{\widehat{\lambda}\pi_I}{(\pi_I+\pi_D-V^*)^3} - 4\tau} > 0$$

It is also apparent that $\frac{d^2V^*}{di^2} < 0$. Thus, $V^*(i)$ is increasing and strictly concave as claimed.■

We have now characterized the solution to Problem P_{conD}. It remains to show that it solves Problem P_{con}.

Fact A.4: Suppose that condition (20) does not hold and and that condition (21) holds and let $\{\underline{i}, \overline{i}, V^*(i)\}$ solve Problem P_{conD}. Then it solves Problem P_{con}.

Proof of Fact A.4: To prove this, we first need to show that $\{\underline{i}, \overline{i}, V^*(i)\}$ is feasible for Problem P_{con}. This requires demonstrating that it satisfies the constraints $\overline{\Omega}(\{\underline{i}, \overline{i}, V^*(i)\}) \geq \pi_D$ and $\underline{\Omega}(\{\underline{i}, \overline{i}, V^*(i)\}) \leq \pi_D$.

To show that $\underline{\Omega}(\{\underline{i}, \overline{i}, V^*(i)\}) \leq \pi_D$, we can use what we know about $\{\underline{i}, \overline{i}, V^*(i)\}$ to demonstrate that $\pi_I \in [\underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta}, \underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta} + (1-\overline{i})\frac{\tau}{\tau+\varepsilon}]$. It then follows from Lemma A.1 that

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

which by construction is equal to π_D.

To see that $\overline{\Omega}(\{\underline{i}, \overline{i}, V^*(i)\}) \geq \pi_D$ suppose first that $\pi_I \in [\underline{\beta} + \overline{\beta} + (1-\overline{i})\frac{\tau}{\tau+\varepsilon}, \underline{i}\frac{\tau}{\tau+\varepsilon} + \underline{\beta} + \overline{\beta} + (1-\overline{i})\frac{\tau}{\tau+\varepsilon}]$. Then, by Lemma A.2

$$\overline{\Omega} = 1 - \pi_I - \int_{\widehat{i}}^{\overline{i}} (\frac{V^*(i) - \pi_D - \pi_I/2}{V^*(i) - \pi_D})di - (1-\overline{i})\frac{\varepsilon}{\tau + \varepsilon},$$

which is bigger than π_D when (21) holds. If $\pi_I \in [\underline{\beta} + \overline{\beta}, \frac{1-\overline{i}}{1+\frac{\varepsilon}{\tau}} + \underline{\beta} + \overline{\beta}]$ then by Lemma A.2

$$\overline{\Omega} = 1 - \pi_I - \int_{\widehat{i}}^{\overline{i}} (\frac{V^*(i) - \pi_D - \pi_I/2}{V^*(i) - \pi_D})di - (1-\overline{i})[\frac{1}{2} - (\frac{\pi_I - \underline{\beta} - \overline{\beta}}{1 - \overline{i}})(\frac{\tau - \varepsilon}{2\tau})],$$

which is again bigger than π_D when (21) holds.

Given that $\{\underline{i}, \overline{i}, V^*(i)\}$ is feasible for Problem P_{con}, if it were not a solution there would exist some alternative inverse seat-vote curve $\{\underline{i}_a, \overline{i}_a, V_a^*(i)\}$ which was also feasible but yielded a higher level of welfare. Now clearly it must be the case that

$$\int_{\underline{i}_a}^{\widehat{i}_a} (\frac{\frac{\pi_I}{2} + \pi_D - V_a^*(i)}{\pi_I + \pi_D - V_a^*(i)})di + \underline{i}_a\frac{\varepsilon}{\tau + \varepsilon} > \pi_D,$$

because otherwise $\{\underline{i}, \bar{i}, V^*(i)\}$ could not solve Problem P_{conD}. However, we can show using Lemma A.1 that it must be that

$$\underline{\Omega}(\{\underline{i}_a, \bar{i}_a, V_a^*(i)\}) \geq \int_{\underline{i}_a}^{\widehat{i}_a} (\frac{\frac{\pi_I}{2} + \pi_D - V_a^*(i)}{\pi_I + \pi_D - V_a^*(i)}) di + \underline{i}_a \frac{\varepsilon}{\tau + \varepsilon}$$

which contradicts the assumption that $\{\underline{i}_a, \bar{i}_a, V_a^*(i)\}$ is feasible for Problem P_{con}. ■

The Proposition now follows from combining Facts A.1 - A.4. *QED*

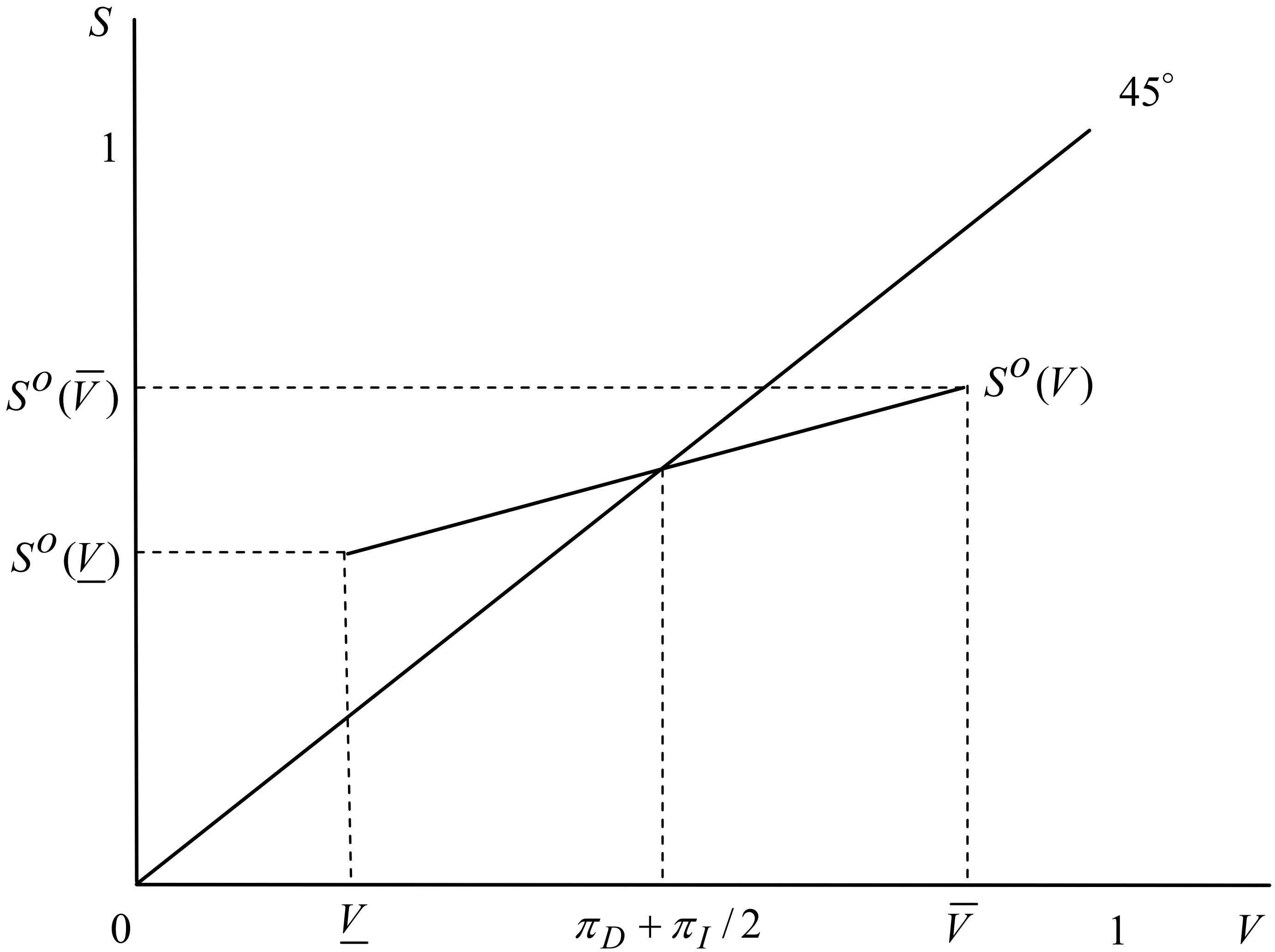
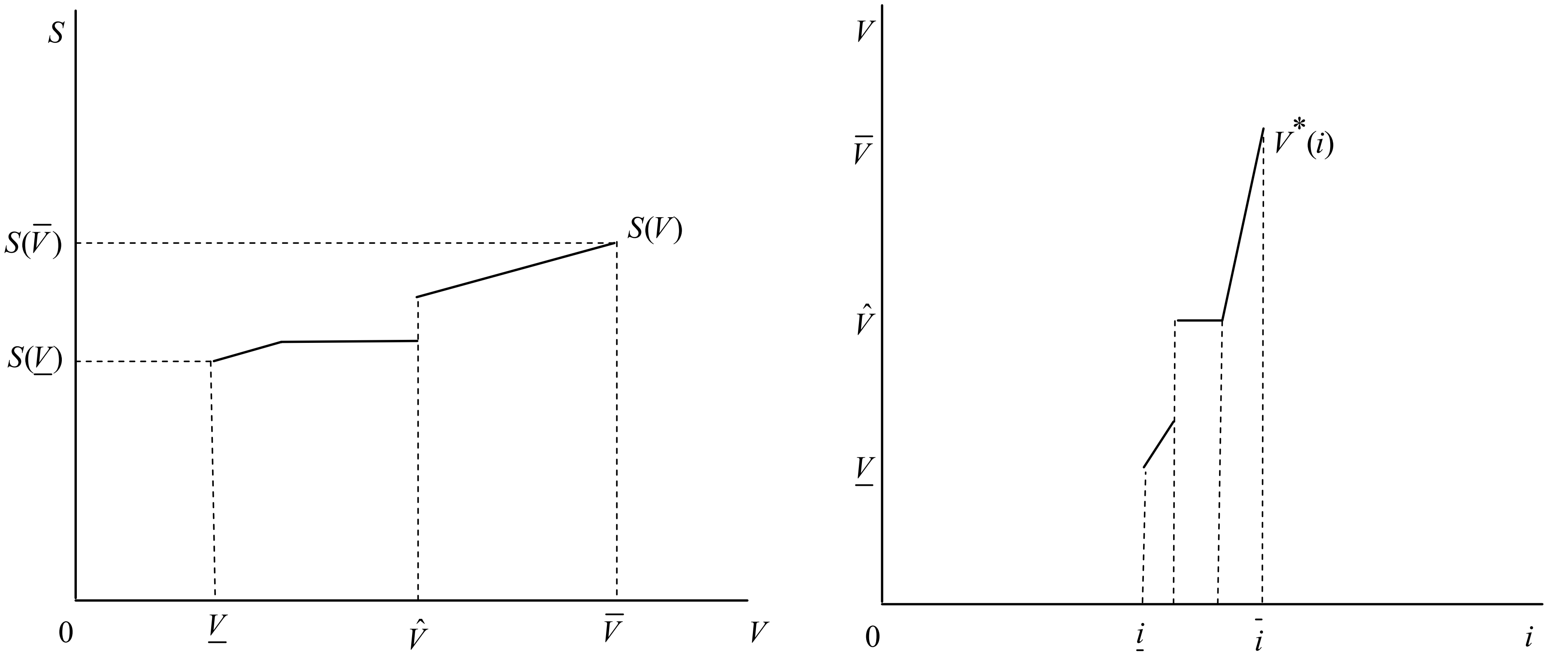


Figure 1



S
$S(\overline{V})$
$S(V)$
$S(\underline{V})$
0
$\underline{V}$
$\hat{V}$
$\overline{V}$
V
V
$\overline{V}$
$V^*(i)$
$\hat{V}$
$\underline{V}$
0
$\underline{i}$
$\bar{i}$
i

Figure 2

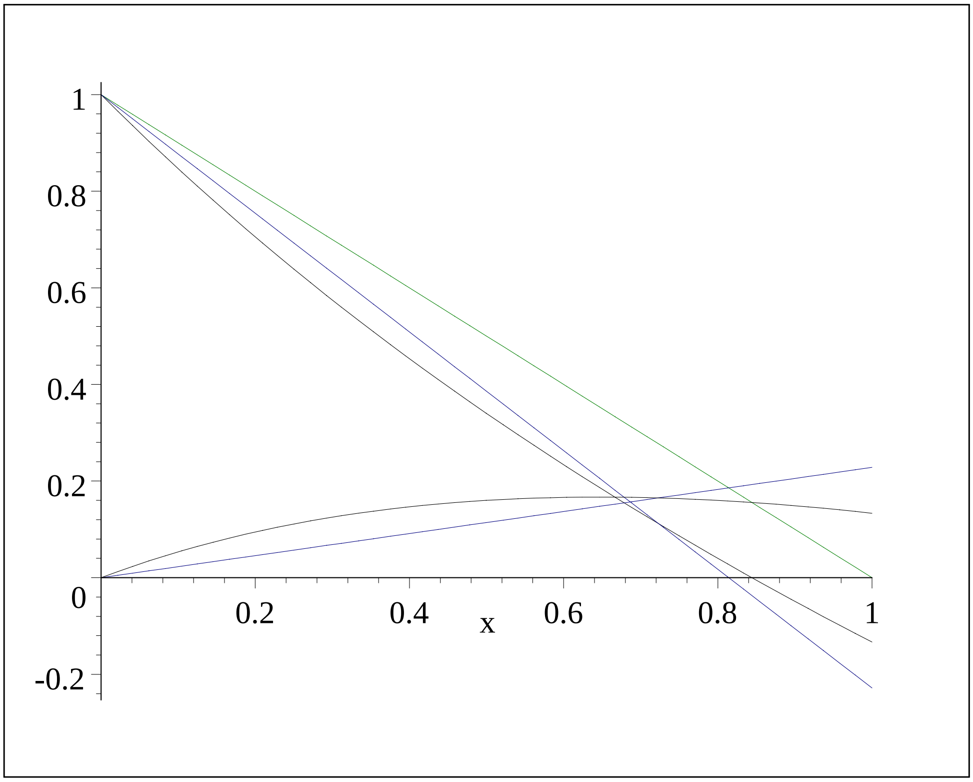
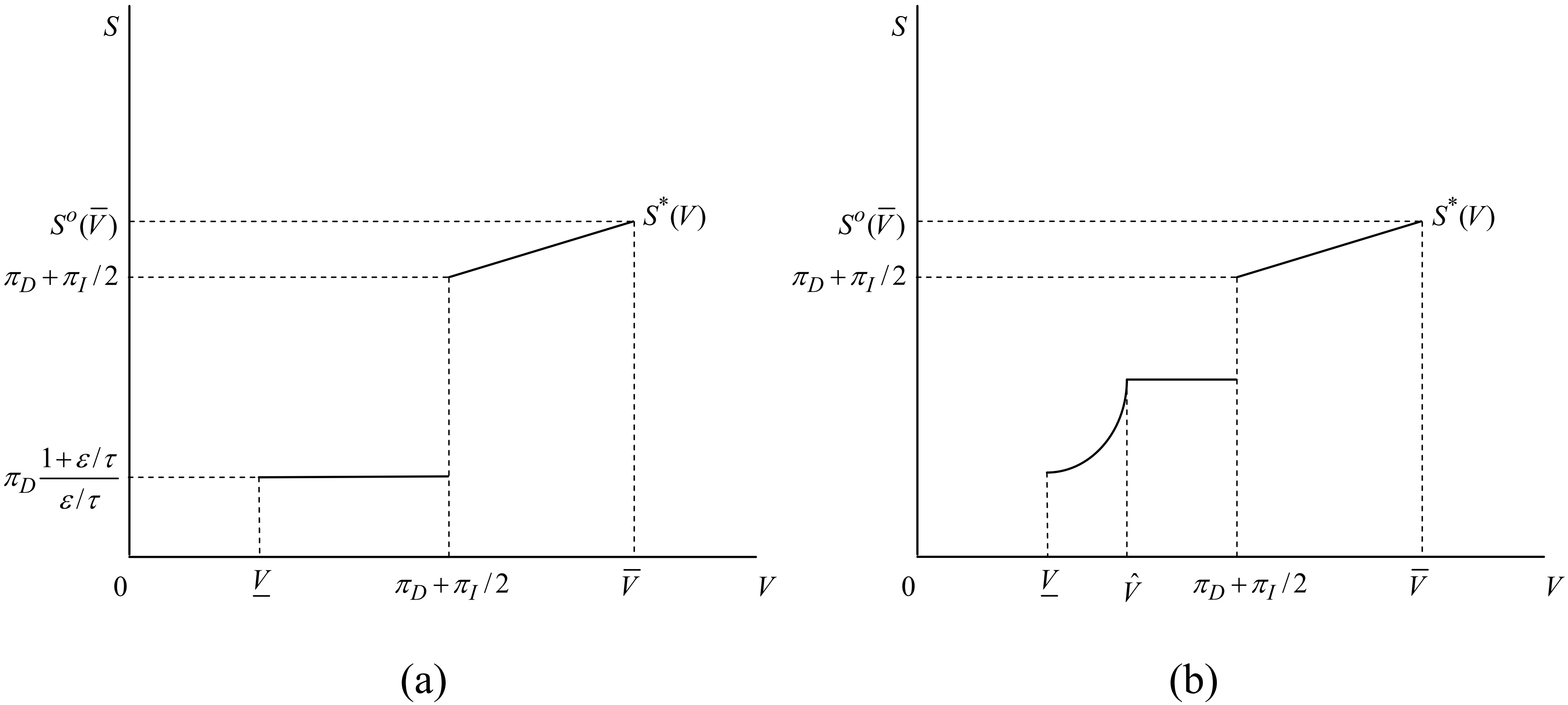
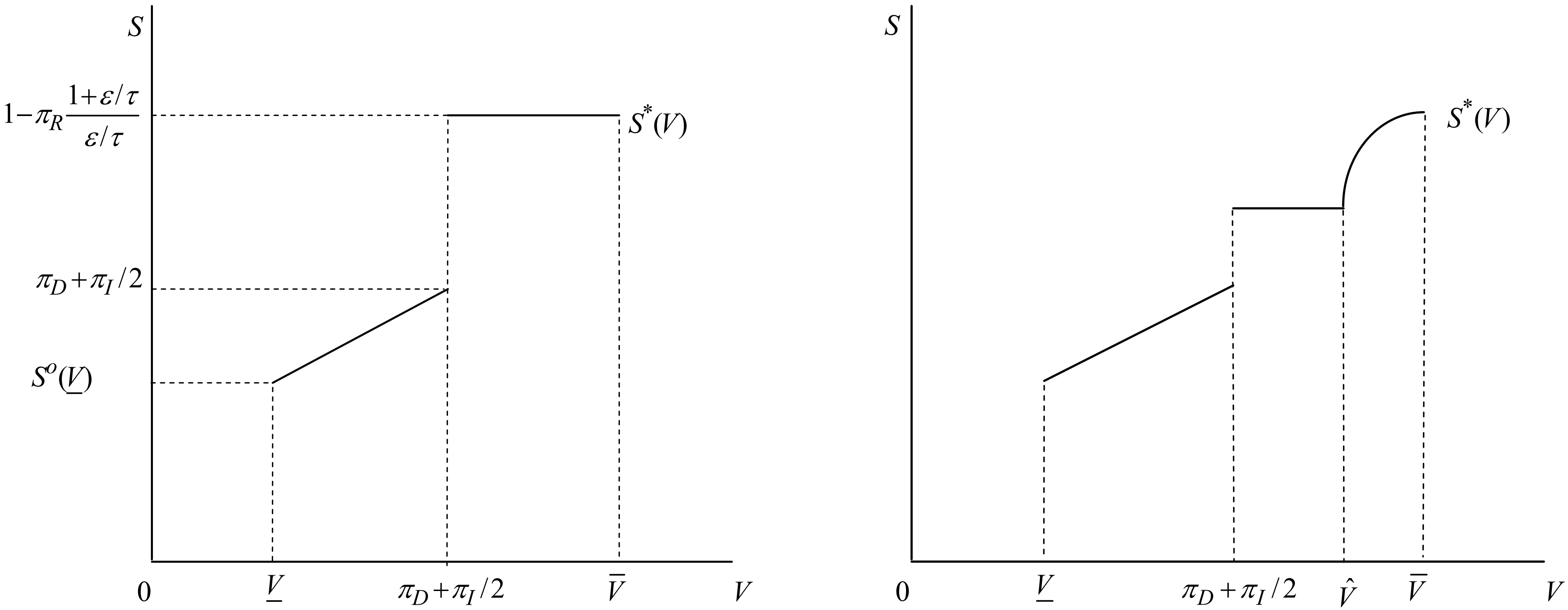


Figure 3



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)

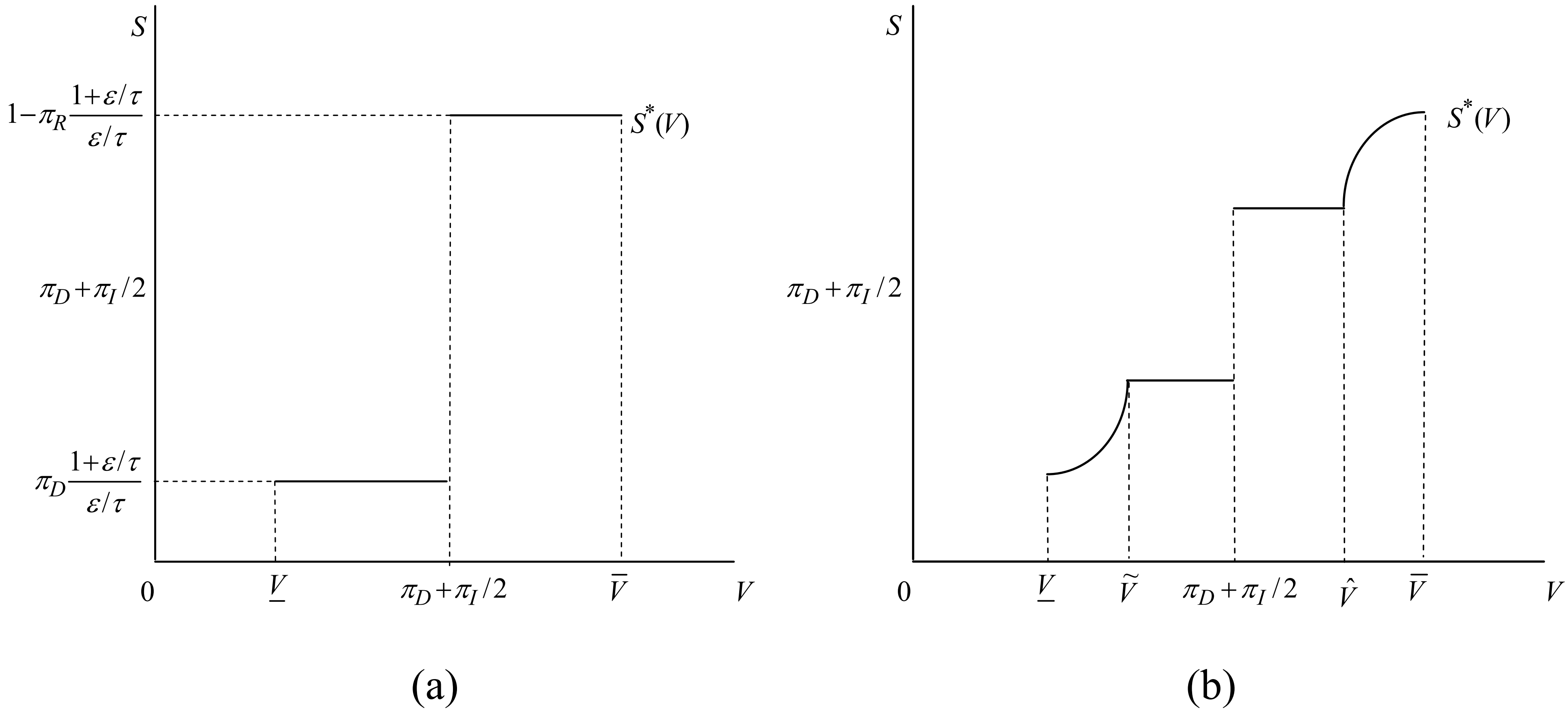
Figure 4


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

(a)

(b)

Figure 5



S
$1-\pi_R \frac{1+\varepsilon/\tau}{\varepsilon/\tau}$
$S^*(V)$
$\pi_D+\pi_I/2$
$\pi_D \frac{1+\varepsilon/\tau}{\varepsilon/\tau}$
0
$\underline{V}$
$\pi_D+\pi_I/2$
$\bar{V}$
V
(a)
S
$S^*(V)$
$\pi_D+\pi_I/2$
0
$\underline{V}$
$\tilde{V}$
$\pi_D+\pi_I/2$
$\hat{V}$
$\bar{V}$
V
(b)

Figure 6