# Party Positions and the Seats/Votes Relationship with Ideological Voters

by

William Michael Leblanc

M.S. Social Science
California Institute of Technology, 2001

B.S. Political Science
Massachusetts Institute of Technology, 1999

Submitted to the Department of Political Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosphy in Political Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2007

Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Department of Political Science
August 6, 2007

Certified by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Stephen Ansolabehere
Elting E. Morison Professor of Political Science
Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Roger Peterson
Associate Professor of Political Science
Chair, Graduate Program Committee

# Party Positions and the Seats/Votes Relationship with Ideological Voters

by

William Michael Leblanc

Submitted to the Department of Political Science
on August 6, 2007, in partial fulfillment of the
requirements for the degree of
Doctor of Philosphy in Political Science

## Abstract

Chapter 2 is based on two axioms: party members must run under common platforms, but are made up of incumbents who seek their own individual re-election. Politicians seek to win their own seats in the legislature, but they must run under a common party label. In both single-member district and proportional representation systems, equilibrium platforms are shown to diverge substantially, with one party located near the 25th percentile of the voter distribution and the other near the 75th percentile, rather than converge to the median. The model also yields predictions concerning short-term economic shocks, incumbency advantages, and gerrymandering.

Chapter 3 is based on ideological voters. With purely ideological voters, party vote share depends on the distribution of voters in the entire country. Seat share depends on distribution of district medians. The seats/votes curve is therefore a combination of two different functions. This presents an identification problem for studying either function without accounting for the other. The incumbency advantage is also considered.

Chapter 4 measures ideology using the Cooperative Congressional Election Study using factor analysis on voter responses to policy questions. I discuss the robustness of the measure and implications for the model of chapter 3.

Chapter 5 finds little evidence of structural bias against either party under districting. However, in a hypothetical party-list system, there would be a massive structural advantage for the left party. The seats/votes curve is predicted to be approximately linear or logistic, unless one takes into account incumbency, in which case the curve becomes non-linear. Senate party platforms are predicted to be more converged than the House, with a midpoint right of center.

Thesis Supervisor: Stephen Ansolabehere
Title: Elting E. Morison Professor of Political Science

# Acknowledgments

Thanks to participants at the 2005 APSA conference for their comments and incites for an early version of chapter 2. Special thanks to co-chair Rebecca Morton for taking the time to comment on the paper.

Thanks to fellow students for support, incite, and perhaps most importantly, comiseration. These include, but not limited to, Michiko Ueda, Sarah Sled, Adam Ziegfeld, Jung Ho Roh, Francesco Giumelli, Christopher Wendt, Sarah Sled, and Anna De La O.

Thanks to the Department of Political Science as a whole for fostering and maintaining a robust intellectual environment and a positive place to work. Special thanks to Gabriel Lenz and Riccardo Puglisi for both specific comments on the dissertation and general discussions on theory and politics.

Extra special thanks to my committee, Stephen Ansolabehere, Jim Snyder, and Charles Stewart. Always available, always inciteful, and always supportive.

Thanks most of all to my parents and grandparents, for their unending support and understanding.

# Contents

# Chapter 1

# Introduction

This dissertation is made up of a collection of four closely related papers. The grand unifying theme is exploring some implications of taking the one dimensional ideological voter model seriously.

Chapter 2 describes step one of the process, a theoretical model, using ideological voters as a central axiom, which makes testable predictions and has implications for things of central importance to political science: election outcomes and representation. Of course, others have done ideological voting models before as well. Other assumptions are needed to generate a unique model with unique predictions. The other central assumption of the Chapter two model is that incumbents determine party platforms, and that each incumbent solely cares about her own re-election to her own personal district.

With these two central assumptions (combined with others) the model of chapter 2 predicts that party platforms will depend on the distribution of districts. If districts are close together, the party platforms will be close together. If a party does well in a previous election, winning more centrist districts, the party will become more centrist, increasing its future election prospects. If the distribution of districts is skewed, one party's platform may be closer to the median district, giving that party a permanent advantage.

Many of these predictions depend upon the distribution of voter ideology across districts. The model can not be applied unless one can estimate how voter ideology

is distributed. This project considers two approaches to finding the distribution of voters: the seats/votes curve (chapter 3) and direct measurement of voter preferences (chapters 4 and 5).

The distribution of district vote is sometimes summarized by the seats/votes relationship. The seats/votes curve describes how many seats a party wins, given its aggregate vote total. Seats/votes responsiveness is defined as the change in seat share per unit change in vote share. As the number of moderate districts increases, holding all else constant, responsiveness should increase as well. It is therefore tempting to use responsiveness as a proxy for district density. Unfortunately, as discussed in chapter 3, the number of moderate districts is not the only variable influencing responsiveness. The number of moderates in each district matters as well. Consider adding liberal and conservative extremists to each district in equal measure. This will make it harder to change vote share, since a larger fraction of the electorate is now firmly dedicated to each party. Therefore, if one DOES manage to change vote share by a unit, this implies a substantial change in the ideology of the pivotal district. Fewer moderates therefore induces a higher seats/votes responsiveness. However, adding extremists in equal measure to each district does not change district medians. Therefore, one can arbitrarily change seats/votes responsiveness without changing the distribution of districts. There is a fundamental identification problem.

Chapter 3 discusses how one might disentangle the distribution of districts from the seats votes curve. The solution is to use outside information about the distribution of ideology to supplement the seats/votes relationship and thereby generate the distribution of districts. The chapter also discusses how one might run the process in reverse, generating seats/votes curves using information on the distribution of ideologies. Implementing ALL of these methods is beyond the scope of this project.

However, chapters 4 and 5 implement one of the techniques: the direct measurement of voter ideology via a survey. The Cooperative Congressional Election Study provided a unique opportunity to measure the ideology of over 36501 subjects. With so many observations, it is possible to get fairly good approximations of the median voter in every state and congressional district. Chapter 4 discusses the survey and

the technique used to measure voter ideology. Chapter 5 brings everything together. Using the ideology measure and survey discussed in chapter 4, the distribution of district ideologies can be interpreted through the lens of the model discussed in chapter 2. For example, the distribution of states is less dispersed than the distribution of House districts. Therefore, according to the party position model, Senate parties should be closer together, and Senate incumbents should be more vulnerable to electoral shocks. The model of Chapter 2 gives context to the data described in Chapter 4.

# Chapter 2

# Party Platforms

## 2.1 Introduction

This chapter considers a model of ideological voters choosing between two competing parties.[1] These parties are made up of election-oriented, self-interested incumbents running for re-election in districts. Furthermore, the model assumes that voters, possibly due to informational constraints, can only observe one universal party position for each party. That is, members of each party are forced to run under one common "platform." However, the candidates of the party do not have identical goals - each candidate represents a particular district, and wishes to win her particular district. The model also assumes the current incumbents in the party decides the party platform. This assumption can either be interpreted literally, as incumbents voting based on majority rule on a platform, or as an approximation of voters perceiving the aggregate behavior of partisans during the last legislative session and thereby observing the implicit party positions. Finally, the model assumes that incumbents are not certain about voter choices. This is represented by a valence term in voter utilities that makes all voters more likely to support one party over the other, independent of ideology. These and other supporting axioms are formalized in section three.

The basic argument is as follows. In the setup here, there exists a cut-point such

---

[1] This chapter draws substantially from a working paper, "When Parties are Not Teams: Party Positions in Single Member District and Proportional Representation Systems" by Stephen Ansolabere, William Leblanc, and James M. Snyder.

that all voters left of the cut-point support the left party, and all voters right of the cut-point support the right party. We will show that each incumbent has induced ordinal preferences that correspond to the preferences of the median voter of the district she represents. As long as each party won at least some districts in the last election, the median incumbent in the left (right) party will desire a platform left (right) of the preferences of the median voter in the median district. Since one platform is left of center, and the other platform is right of center, the platforms are diverged.

While the basic argument above for divergence is fairly simple, working through the model in detail yields other, more subtle, testable predictions. These predictions are discussed in sections four and five. The degree of divergence remains large even as the uncertainty about valence becomes small (unlike Calvert (1985) and similar models). If the distribution of districts is symmetrical, there is a sense in which platforms at the first and third quartiles of district medians is stable. The ordering of parties (left and right) endures over time. Parties that do well in the past tend to moderate and therefore do better than they would have otherwise in the future. Perhaps most surprisingly of all, if the distribution of districts is asymmetrical (as might be caused by majority-minority districting), a party can be permanently electorally disadvantaged, even with ideological voters and flexible party platforms.

Section six contrasts the districted case above with a party list, proportional representation system. Surprisingly, most of the conclusions remain the same, with the distribution of voters replacing the role of the distribution of district medians. Of course, if the distribution of districts does not equal the distribution of voters, there may be significant policy implications. For example, party positions may be further apart, and election results may be less sensitive to economic shocks.

## 2.2 Past Literature

The seminal analytical model of electoral competition among parties, due to Hotelling (1929) and Downs (1957), begins with the assumption that parties are teams. All

16

candidates and legislators within a party unite around a common goal of winning control of government and present a common ideological position. Voters choose among parties on the basis of their relative distance from the national party platforms. These assumptions lead to a powerful prediction – the Median Voter Theorem. When electoral competition in a two-party system is waged along one dimension ideological spectrum, the parties will converge to the ideal point of the median voter. Many variants on this model have been developed, allowing, for example, open entry, multiple numbers of parties, varieties of electoral laws, and strategic voting.[2] This important line of inquiry, however, retains the assumption that political parties act as teams. Every politician within a party works for the common good in her or his party, even at the expense of the individual's own electoral fortunes.

The assumption that parties are teams runs counter to a second line of analytical inquiry in Political Science. Politicians, David Mayhew famously argued, single-mindedly seek election. They are self-interested and act in ways to protect and improve their own positions. As a result, in a system like the U.S., legislators are very responsive to the preferences of the voters in their own districts, and parties resemble somewhat disorganized groups more than teams. (See also, Fiorina, 1980, and Krehbiel, 2000). Robertson (1976) criticizes the application of the Downsian model to the United Kingdom because, even in a parliament with strong parties, the parties are internally divided into factions and individual politicians' interests; they are not teams. Surprisingly little theoretical research has approached the tension between the rational choice models of parties and the rational choice models of candidates.

This model is not, of course, the first model to extend the Downsian framework by examining internal party conflicts and organization. There are two important

---

[2]There is an extensive theoretical literature predicting platform divergence based factors such as entry deterrence, politicians' policy preferences, voter abstention, primary elections, valence issues, party activists, and special interest groups. Examples include Aranson and Ordeshook (1972), Wittman (1977, 1983), Enelow and Hinich (1981), Aldrich (1983), Palfrey (1984), Bernhardt and Ingberman (1985), Calvert (1985), Aldrich and McGinnis (1989), Cox (1990), Ingberman and Villani (1993), Londregan and Romer (1993), Baron (1994), Snyder (1994), Snyder and Ting (2002), Poutvaara (2003), Schofield (2004), Serra (2005), and Callander (2005). With the exceptions of Snyder (1994) and Snyder and Ting (2002), however, these models either explicitly or implicitly deal with a single district or parties with collective preferences.

strands of theorizing about internal party decisions. The first line of theory assumes competition across many different districts or constituencies. Snyder (1994) analyzes a model similar to ours, but his model is deterministic. That model yields indifference or flat parts of politicians' revealed preferences and multiple equilibria. Under voter uncertainty, though, politicians running for single member districts are never indifferent about what position their party takes. Even a little uncertainty on the part of candidates induces politicians to maximize their probabilities and produces strong preference orderings over the party platforms.

In fact, the relevant literature goes back much farther, at least to Robertson's (1976) informal discussion of the logic of party strategy. He argues that one of the crucial functions of political parties is to establish connections across the potentially unrelated races for different legislative seats, that one means of doing this is to provide national party positions on salient issues, and that natural intra-party divisions will arise in deciding what these positions should be. Austen-Smith (1984, 1986) extends and refines Robertson's insights by analyzing game-theoretic models in which candidates announce individual platforms to appeal to their districts, but must also commit to an aggregate party platform that will be implemented as policy.[3] Calvert and Isaac (1981) model the potential tensions between a party's legislators and the president.[4] This paper assumes a simpler structure. Politicians decide using majority rule within their legislative caucus what platform they wish to implement.

A second strand of the literature assumes that politicians or party activists themselves have policy preferences. Aldrich (1983), Aldrich and McInnis (1989), Poutvaara (2003), Gomberg et al. (2004), and others treat parties as collections of "activists" or

---

[3]Austen-Smith (1984) models two parties that compete for control of an $n$-member legislature by running candidates in each of $n$ single-member districts. Each candidate chooses an individual platform, and these platforms are then aggregated via a "party constitution" into a single position. This is the position that all members of the party will support once elected. Voters know the party constitutions, and vote accordingly. Austen-Smith shows that the parties' positions converge in equilibrium, but individual candidates' platforms do not. He also shows that (pure-strategy) equilibria may fail to exist even if the policy space is unidimensional.

[4]Calvert and Isaac (1981) consider a model in which candidates announce their own platforms, but where candidate's from the president's party must contend with an exogenous party reputation given by the president's record. They analyze the promises made within each district and show that candidates will overcompensate. This model does not endogenize the overall party label.

primary voters with policy preferences, and endogenize both party membership and party positions. Most recently, Roemer (1998, 2002) models internal party decision-making using a solution concept he calls "Party Unity Nash Equilibria." He assumes that parties are composed of three types of actors – militants who care about "publicity," reformists who care about policy outcomes, and opportunists who care only about winning office – and choose their platforms by unanimity rule. In these models there is just one constituency, and internal party conflict is driven by differences in party members' policy preferences (or lack of such preferences). This paper starts with the assumption that politicians solely seek to win their own seats. As a consequence each incumbent wants a party platform equal to their own districts' median. This may look like politicians have their own preferences on policy, but in fact they merely want to do what gets them elected.

## 2.3   Basic Model

There are two parties, $X$ and $Y$, that compete for control of government. The parties choose policy platforms from a policy space $Z \subset R$. Lower case $z$ denotes a generic policy position and $x$ and $y$ denote the platforms of the parties.

Every candidate for office belongs to one and only one party. All politicians in a party run under that party's label and can not distinguish themselves from that label. Each politician seeks to win his or her own seat in the legislature. The utility of politician $j$ in party $X$ is $Q_X^j = Pr(j\ Wins)$, and the utility of politician $j$ in party $Y$ is $Q_Y^j = Pr(j\ Wins)$.[5]

Voters choose between the parties on the basis of two factors. First, voters compare parties' policy platforms. We will make the standard assumption that voters have single-peaked policy preferences, but allow a general functional form. Each voter has an ideal policy platform, which we will denote as $w$. We will write the spatial component of the utility function, using the lower case $u$, as $-u(z - w)$. Thus,

---

[5]It is possible generalize to more complex objectives, such as seeking to be in the majority. This is left to future work.

$-u(d)$ is maximized at 0, is increasing for $d < 0$ and decreasing for $d > 0$. We will assume $u$ is strictly convex, so $-u$ is strictly concave. We will also assume $u$ is continuous. Second, a valence term captures short-term stochastic shocks, such as economic fluctuations or scandals. These shocks arise exogenously; neither party can influence their occurrence. The valence shocks are realized after parties announce their policies. Each party may have separate valence shocks, written $v_X$ and $v_Y$. We will assume that the valence shock is a random variable that the parties cannot control. Let $v = v_x - v_y$ designate the net advantage of party $X$. The cumulative density of the valence shock is $G(v)$.

The overall utility that a voter receives, denoted with an upper case $U$, is the sum of their valence and spatial preferences. The utility a voter receives if party $X$ holds office is $U(x, w, v_x) = v_x - u(x - w)$ and the utility a voter receives if party $Y$ holds office is $U(y, w, v_y) = v_y - u(y - w)$. The differential in utility between party $X$ and $Y$ is denoted $\Delta U(z; x, y, v) = v - u(x - z) + u(y - z)$. A voter chooses party $X$ if $\Delta U$ is positive and chooses party $Y$ if $\Delta U$ is negative.

The policy preference of the indifferent or pivotal voter is of particular interest. We will denote the "cut-point", $c$, as the policy such that all voters with $w < c$ vote for the party on the left along the line $Z$ and all voters with $w > c$ vote for the party on the right. The cut-point depends on the values of $x, y$, and $v$, and can therefore be written as a function $c(x, y, v)$. The cut-point can be thought of as the point along $Z$ such that any voter with an ideal point at that point is indifferent between the parties. Given policy positions, the cut-point depends on the random variable $v$, and the expected value of the cut-point can be written $E[c]$.

Stronger assumptions about the utility function allow for explicit solutions for the cut-point $c$. With quadratic utilities, which are commonly assumed in the literature,

$$c = \frac{x + y}{2} + \frac{v}{2(y - x)}.$$

In examples below we will use this characterization of the cutpoint. Inspection of this result reveals that the cut-point need not lie between the two parties. For example,

20

suppose that $x = 0$ and $y > 0$. If $v > y^2$, then $c > y$.

Three propositions that characterize electoral competition generally follow from the more generic assumptions. These propositions characterize properties of the cut-point that are not dependent on the electoral system. The first proposition guarantees that the cut-point is well defined, or only one party holds seats.

**Proposition 3.1.** For fixed $x$, $y$, and $v$, if $x < y$ then $\Delta U(w; x, y, v)$ is strictly decreasing in $w$. If $x > y$ then $\Delta U(w; x, y, v)$ is strictly increasing in $w$.

The second proposition establishes the monotonicity of the cut-point in terms of the party positions for what one might consider normal politics. If the valence shock is not too large, then if both parties shift in a certain direction, the cut-point between them will shift in that direction. Formally, this is stated as follows.

**Proposition 3.2.** Let $x < y$, $x' < y'$, $x' > x$, and $y' > y$, then there exists $\epsilon > 0$ such that if $|v| < \epsilon$ then $c(x', y', v) > c(x, y, v)$ or all voters vote for party $X$.

Large valence shocks result from dramatic changes in the economy, foreign policy catastrophes, and large scandals, such as Watergate, that might affect an entire party. Normal times correspond to situations when the valence shocks are not too large. Of course normal times are relative to the particular electoral alignment. If the parties are very close to each other then a large valence shock might eliminate most of the disadvantaged party's legislative seats. This is why if $v$ is large, increasing the platforms of both parties (but decreasing the distance between them, or more generally if there is enough asymmetry in $u$) can actually decrease the cut-point.

The tradeoff is easy to see in the quadratic case. Hold $y$ fixed. Moving $x$ $\epsilon$ units to the left has a direct $frac\epsilon 2$ impact on the cut-point, reducing the left party's vote share. However, there is also the additional impact of the second term, $fracv2(y - x)$. The marginal impact of a small change is $frac-v\epsilon2(y - x)^2$, which helps the left party if $v < 0$. That is, the left party is helped by assuming a more extreme platform if valence shock goes against the party. If the valence shock is sufficiently large, the second term can outweigh the first term. For example let $x = 0, y = 100, x' =$

21

$99, y' = 101, v = -400$. Then by the formula given above $c(x, y, v) = 50 - 2 = 48$ while $c'(x', y', v) = 100 - 100 = 0$. Diverged party platforms can help insulate parties (and individual incumbents) from disastrous scandals.

Third, when neither party has an advantage a priori, the expected cut-point is simply the average of the two party's positions. A distribution function is antisymmetric around 0 if $G(v) = 1 - G(-v)$.

**Proposition 3.3.** If $u$ is symmetric and $G$ is antisymmetric around 0, then $E[c] = \frac{x+y}{2}$.

Proposition 3.3 offers an important reference case. Most models of spatial voting assume symmetric spatial preference functions, especially the quadratic, and assume that the mean of the valence shock is zero. When preference functions are symmetric, the parties can anticipate that the ideal point of the voter who is indifferent between the parties lies half way between the two parties' announced policy positions.

The above definitions and results refer to properties of the model without respect to time. In what follows, we will consider the strategies of the parties across two election cycles. Each cycle is, in turn, divided into two parts. In the first part of an election cycle, the platform is determined. In the second part, the valence shock is realized and therefore the cut-point is determined. The two time periods are indexed with the subscript $t = [0, 1]$. They differ slightly. In the first period the parties' policy positions $x_0$ and $y_0$ are exogenously determined. Then, the valence shock $v_0$ is realized, which determines the cut-point $c_0 = c(x_0, y_0, v_0)$ and a division of seats between parties X and Y. The division of seats equals the fraction of voters or districts below the cut-point, $c_0$. In the second time period, the elected officials within each of the two parties choose their parties' policy platforms $x_1^*$ and $y_1^*$ via majority rule. The asterisks indicate that these platforms represent the endogenous, equilibrium outcome of intra-party decision making. After the parties choose their platforms, the valence shock $v_1$ is realized, which yields a cut-point $c_1 = c(x_1^*, y_1^*, v_1)$ as an election result. Wherever possible we will drop the subscript $t$ to avoid clutter.

Further development of the model also depends on the electoral system. We will

distinguish districted systems from proportional representative list systems. Many countries, such as Germany and Japan, now use a mixed system. An interesting extension of this analysis would be to such systems. Throughout we will hold constant the number of parties at two.

## 2.4 Platform Choice in the Single-Member District System

In districted systems, candidates run and individuals vote within particular districts; votes are not transferable across districts; the candidate who wins the most votes in a district wins the seat; and the total number of seats won is the total number of districts won.

Let the median voter of a district be designated by $m$, and let $H(z)$ be the CDF of district medians. We will assume $H$ is continuous. The cut-point for a given election and the median voter of a given district jointly determine which party wins that district for that election. (Since $H$ is continuous one can ignore district medians exactly on the cut-point since they constitute measure zero) For example, if $x < y$ then all districts such that $m < c$ choose the candidate of party $X$ and all districts such that $m > c$ choose the candidate of party $Y$. Given voters' assumed strategies, candidates' expected utility functions are easily described. A candidate, $j$, from party $X$ has expected utility

$$Q_X^j(x; y, m) = Prob(\Delta U(\tilde{z}; x, y, v) \geq 0) \qquad (3a)$$

Similarly, the expected utility of a candidate, $j$, from party $Y$ is

$$Q_Y^j(y; x, m) = Prob(\Delta U(\tilde{z}; x, y, v) \leq 0) \qquad (3b)$$

Let $\tilde{x}(y, m)$ be the most-preferred platform of party $X$'s candidate in a district with median at $m$, and let $\tilde{y}(x, m)$ be the most-preferred platform of party $Y$'s can-

23

didate. As the following proposition shows, $\tilde{x}(y, m) = m$ and $\tilde{y}(x, m) = m$.

**Proposition 4.1.** In equilibrium, if $v$ has full support, the ordinal preferences on party $X$'s platform of an incumbent in party $X$ who represents a district with median voter $m$ are described by $u(x - m)$ (the preferences of a voter with median point at $m$ for policy $x$.) Similarly, the ordinal preferences on party $Y$'s platform by an incumbent in party $Y$ are described by $u(y - m)$. In general, even if $v$ does not have full support, platform $m$ maximizes the incumbent's utility.

Proposition 4.1 immediately implies that it is natural to describe incumbents as having an "induced ideal point" equal to his or her district median. Uncertainty, which enters through the term $v$, produces the well-defined preference function described in Proposition 3.1. Legislators will have strong orders over the platform choices because their probability of winning declines smoothly as the cut-point moves away from the district's median voter's ideal point. The amount of uncertainty affects the shape of the preference function of each legislator, but even a little uncertainty will generate a well-defined preference function in which legislators have strong preference orders over alternative policy platforms.

It is worth noting that candidate *cardinal* utility is not the same as a representative voter's cardinal utility. Candidate utility is filtered through the distribution of potential valence shocks. Only the ordinal preferences are preserved. While this distinction does not matter for purposes of this paper, it may be important when thinking about extensions (such as platforms maximizing average incumbent utility) or welfare analysis of the results. The loss of cardinal information also implies that thinking of the legislature as a "citizen" legislature made of representative citizens with policy preferences is not quite correct. A legislator's ordinal preferences, then, will be the same as the median voter of his or her district, but the legislator and voter won't have the same cardinal utility – that will depend on the probability function, $G$. It is also worth noting that the induced ideal points of candidates are *independent* of the other party's platform. This does not mean, of course, that a candidate is *indifferent* about the other party's platform. In fact, all candidates in a party prefer the

24

other party's platform to be as extreme as possible, as that increases the probability of winning.

This result is distinct from two other lines of thinking. First, uncertainty guarantees a well-ordered preference function and prevents multiple equilibria. In a deterministic model (Snyder, 1994), the induced utility function of incumbents may have "flat spots" or "regions of indifference." The intuition is that, for a given value of $y$, if valence "jumps" at some points then there are regions of $x$ that induce the same probability of winning for a candidate with given $m$. Full support of $v$ *guarantees* that incumbent preferences are well-behaved. The equilibrium derived below would still exist, but there may be others.

Second, Proposition 4.1 differs qualitatively from the assumption that politicians have their own personal preferences over policy, as in Wittman (1977, 1983), Calvert (1985), and Roemer (2002). In those models, the politicians in a party have common policy preferences and want policy to move in that direction. Uncertainty may lead to divergence, but the degree of divergence is a function of the variance of the distribution of the shocks. [6]

Proposition 4.1 ensures that incumbent preferences are well defined, single-peaked, with ideal points equal to the district medians the incumbents represent. The assumption of majority rule within parties yields an explicit characterization of the equilibrium platforms of the parties and leads to the prediction that when parties are not unified teams they will diverge.

We will now make use of the two period structure introduced at the end of section 3. There are two time periods $0, 1$ such that $x_0$ and $y_0$ are determined exogenously and $x_1^*$ and $y_1^*$ are determined via majority rule by the incumbents from period 0. [7] Assume that $0 < H(c_0) < 1$ so neither party was completely wiped out in the period

---

[6]In this model the complete irrelevance of degree of uncertainty relies on the assumption that candidates do not have direct policy preferences. Suppose one extended this model to give incumbents a direct preference for a particular platform in addition to being reelected. The less uncertainty, the more a candidate could afford to support a policy that deviates from the median voter of her district. Whether this would lead to more or less divergence depends on whether incumbents tend to be more moderate or extreme than their districts.

[7]We will assume throughout that $x_0 \neq y_0$. If $x_0 = y_0$, the party that receives the favorable shock will win all of the seats. It is unclear, then, what a party is if it has no seats.

0 election.

Assume the inverse of $H(z)$, $H^{-1}(\theta)$ exists for $\theta \in (0, 1)$. By the median voter theorem, each party's platform in period 1 ($x_1^*$, or $y_1^*$) will be the median of the induced ideal points of candidates of that party who won in period 0.

**Proposition 4.2.** Suppose $x_0 < y_0$, $0 < H(c_0) < 1$, and $v$ has full support. Then $x_1^*(c_0) = H^{-1}(H(c_0)/2)$ and $y_1^*(c_0) = H^{-1}((H(c_0)+1)/2)$ are the unique core outcomes of majority-rule bargaining among incumbents.

A symmetric result holds if $x_0 > y_0$.

Proposition 4.2 has three immediate implications for the platforms chosen by the parties.

- $x_1^* \neq y_1^*$: Party platforms are diverged.

- If $x_0^* < y_0^*$, then $x_1^* < y_1^*$: Party order is preserved over time.

- $x_1^* \leq H^{-1}(1/2) \leq y_1^*$: Parties take positions on opposite sides of the median district.

With the introduction of a stability concept, one can derive a very simple characterization of the party platforms: one party will locate at the 25th percentile of the distribution of voter ideal points and the other at the 75th percentile.

Consider the case when cut-point in period 0 is $H^{-1}(1/2)$. That is, districts left of the median voted for one party while the other districts voted for the other party; each party won 50% of the vote. Then the party platforms for period 1 are $x_1^* = H^{-1}(1/4)$ and $y_1^* = H^{-1}(3/4)$. The left party platform is the first quartile of districts and the right party platform is the third quartile. If $H()$ is anti-symmetric there is a sense in which this outcome is stable.

We will define platforms $x', y'$ as "zero-valence stable" if whenever $v_0 = v_1 = 0$, $x_0 = x'$, and $y_0 = y'$ then $x_1^* = x_0$ and $y_1^* = y_0$.[8]

---

[8]Whether platforms are in any sense dynamically stable over time for a particular stochastic process defining $v$ is an interesting question beyond the scope of this particular paper.

**Proposition 3.3. Quartile Voter Theorem.** For any symmetric $u$, anti-symmetric $H$, $x = H^{-1}(1/4)$ and $y = H^{-1}(3/4)$ are zero-valence stable.

This result provides an interesting analogy to the Median Voter Theorem. When parties are not unified teams but their members must run under a common label, and when the distribution of district median ideal points is anti-symmetric, the cut-point between the parties will divide the electorate equally at $1/2$, but the party platforms will not converge. Rather, one will locate at the ideal point of the 25th percentile voter and the other will locate at the ideal point of the 75th percentile voter, which correspond to the medians within each of the parties.

# 2.5 Comparative Statics and Empirical Implications

The model carries predictions about the equilibrium platforms or policies of the parties, the shares of the votes, and the effects of changes in the exogenous features. We will focus on two such features, the valence shock $v$ and the distribution of district medians $H$.

## 2.5.1 Effects of Short-term forces

Party platform choices depend on the random valence term $v$, which reflects economic times, scandals, and other factors that indicate the ability of the party to produce commonly shared benefits to the voters. Realizations of $v_0$ affect the cut-point between the parties at time 0, $c_0$. One can use changes in $c_0$ to study the effects of short-run national forces on the positions of the parties and their electoral fortunes.

An interesting and subtle implication of the model is that short-term forces and party positioning interact in their effects on aggregate vote shares or seat shares. To see this clearly, consider the case of quadratic utilities. As noted earlier, the cut-point will be of the form $c = \frac{x+y}{2} + \frac{v}{2(y-x)}$. The first term is just the mid-point between the parties, but the second term is the valence shock weighted by the distance between

the parties. Consideration of this last term indicates that the more converged the parties are the more a valence shock will affect changes in the cut-point and, thus, party platform strategies in later elections. However, if parties take highly divergent positions, short-term forces will have muted effects on the cut-point, and thus on fluctuations in the division of the vote.

To the author's knowledge, this point has not been appreciated in the voluminous empirical literature on economic voting. It suggests that the effect of the economy on aggregate vote shares of parties is magnified by the ideological distance between the parties. Depending on party platform choice the economy may matter a lot or little.

One can also analyze the formula for $c$ to study how valence directly affects positioning. Three comparative-static type results deserve emphasis.

- $x_1^*$ and $y_1^*$ are increasing in $c_0$. Past good performance causes a party to moderate; past bad performance causes a party to move to the extreme.

- If $c_0' > c_0$, then $\exists\ \epsilon$ such that if $\mid v_1 \mid < \epsilon$ then $c_1^*(c_0', v_1) > c_1^*(c_0, v_1)$. For small $v_1$, past good performance causes good current performance.

- If $u$ is symmetric and $G$ is antisymmetric around 0, then $E(c_1^*)$ is increasing in $c_0$. In particular, $E(c_1^*) = (1/2)[H^{-1}(H(c_0)/2) + H^{-1}((H(c_0) + 1)/2)]$. Past good performance causes good future performance, on average.

- Suppose that $H$ is continuously differentiable, $v_1 = 0$ and $u$ is symmetric, and denote the first derivative of $H$ as $h$. The derivative of $c_1$ with respect to $c_0$ yields $c_1' = (1/4)h(c_0)[\frac{1}{h(H^{-1}(H(c_0)/2))} + \frac{1}{h(H^{-1}(H(c_0)/2+1/2))}]$.

The first of these results indicates that party platform choices depend on past party performance. If times are good for one party, say the party to the right, both parties will move away from that party's direction, to the left, and the party to the right can expect to gain seats in the next election. Parties that do well tend to moderate, while parties that do poorly tend to move to the extreme. As a party shrinks, the few remaining incumbents tend to come from more extreme districts. In

their pursuit of their own individual re-election, these incumbents pull their party further to the extreme.

The second and third results describes a force that can create longevity for the majority party. As noted in Proposition 3.3 symmetry in $u$ and $G$ means that one can expect an even split in the shares of seats won. But, as the second and third results suggest, the actual division of seats depends on the values of the shocks. If a party enjoyed a positive shock in period 0 and the party platforms were at the first and third quartile, then the advantaged party will win a majority. If $v_1$ is not sufficiently negative, then that party will win a majority again in period 1.

While a party gains an advantage from past electoral success, the fourth result shows that under many circumstances the advantage is less than the original surge. For example, when the distribution of ideal points is uniform, the derivative of $c_1$ with respect to $c_0$ equals $1/2$. Hence, the period 1 cut-point, if $v_1 = 0$, corrects half of the gain from the period 0 shock. While the formula above is not always less than 1 for all distributions, it is for several commonly assumed distributions, such as the uniform, normal, and logistic.

This result is consistent with observed mid-term seat loss observed in the U.S., France, and other nations. Suppose during a typical election $v \approx 0$. However, if a party wins the presidency in a given year, that is evidence that the valence shock was in favor of that party for that year. One should therefore expect a coattail effect of more members of that party winning legislative seats that year. However, in the upcoming election the valence shock will tend back toward its typical value of zero, and some of the legislative gains will be lost. This will be observed as a mid-term seat loss - although, according to the model, the loss will typically be less than the coattail effect.

While this paper does not model long-term strategic consequences of repeated play, the above process suggests that a party could be wiped out by repeated negative shocks. There are several possible reactions to this prediction. First, in many specific cases these shocks would have to be very large. For example, if $H$ is uniform, a party could repeatedly sustain negative valence shocks equal to one quarter

of the seats indefinitely without being wiped out. Second, as a party grows smaller, the assumption that platforms are determined by incumbents and only incumbents becomes more strained. This assumption was made to clarify and simplify the argument of the paper, but could of course be relaxed. An obvious extension to the model is to give some weight to challengers in the party, and have that weight increase as the number of challengers relative to incumbents increases. This will cause a badly beaten party to eventually begin to moderate, and thus bounce back. Finally, one could keep the model as is and embrace the prediction. Two-party democracies are not necessarily perfectly stable for all time. Just as in the long-run we are all dead, perhaps two-party systems, with sufficiently pernicious $H$ and $G$ distributions, are doomed to eventually experience the slow collapse of one of the paries. In particular, a sufficiently perverse gerrymander could lead to the complete and nearly inevitable destruction of one of the parties.

These results should be contrasted with the Downsian model. In that framework, because the platforms are the same, shocks cause one party to win all seats, and shocks will change all seats or none.

The dependence of current party performance on past party performance in the model has wide-reaching implications for many empirical studies of elections. At the very least, one should expect autocorrelation over time of party electoral performance. Studies of elections over multiple years (such as measures of swing-ratios) that ignore such auto-correlation may understate standard errors. Studies that use lagged election results as a proxy for omitted independent variables should observe a positive correlation between current and past elections. However, such a correlation does not imply that the lagged election results are actually correlated with the omitted variable.

The cross time-dependence of cut-points predicted by the model suggests at least two direct empirical tests. At the very least one should look for and measure cut-points, and check to see if they follow an auto-regressive process. If one is willing to take the model very seriously, one could also test the specific formula $E(c_1^*) = (1/2)[(H^{-1}(H(c_0)/2) + H^{-1}((H(c_0) + 1)/2)]$.

Of course to perform the regressions suggested above one would need estimates of $c$ and $H$. Here is one possible approach. Label the parties as $X$ and $Y$ and assume $x < y$. Let $\hat{m}_i$ be the estimated ideology of district $i$.[9] A natural estimator for $H$ would be the empirical distribution of districts $\widehat{H(z)}$; the fraction of districts with ideology less than $z$. To estimate $c$ pick a non-decreasing loss function $L(z)$. Then, $\hat{c} = argmin_z \sum_i (L(|z - \hat{m}_i|)\delta_i(z)$ Where $\delta_i(z) = 1$ if $\hat{m}_i < z$ and district $i$ voted for party $Y$ or $\hat{m}_i > z$ and district $i$ voted for party $X$. Otherwise $\delta_i(z) = 0$. In other words, district $i$ contributes loss if it votes for the "wrong" party. In this stylized model, districts never vote for the wrong party, and so such a correction would be unnecessary.

## 2.5.2 Incumbency Advantages

Incumbency advantages are an important, if somewhat exceptional, feature of American elections. U.S. politicians' vote shares rise from the first time they win a seat, as a non-incumbent, to the second time they contest a seat, as an incumbent. Empirical study of the incumbency advantage has found that this appears to be constant, additive effect, and it does not interact appreciably with ideology and other factors.

A simple, constant incumbency advantage may be added to the analysis without changing the optimal policy of each incumbent. Assume that the valence term includes two components – the national random shock about which candidates are uncertain and an additive incumbency term which is known to the candidates and voters.[10] Let $\alpha_i$ be the advantage of the incumbent in district i; $\alpha_i > 0$ if the incumbent is from party $X$ and $\alpha_i < 0$ if the incumbent is from party $Y$. Now, $\Delta U = \alpha_i + v - (x - z)^2 + (y - z)^2$.

Officeholder advantages change the shape of the politicians' induced utility functions. Adding a constant to the valence term creates a unique cut-point for each politician. In the case of quadratic utilities, the cut-points are: $c_i = \frac{x+y}{2} + \frac{v+\alpha_i}{2(y-x)}$. For

---

[9]Alternatively, one could use a "corrected" estimate for $m$ that accounts for district specific factors other than ideology, such as incumbency.

[10]A random incumbency effect adds another layer of complexity to the analysis.

any given $x$ and $y$, the addition of the incumbency advantage moves the cut-point to the right if the incumbent is of the left party and to the left if the incumbent is of the right party. Because the cumulative distribution of voter ideal points is monotonic, the incumbents' probabilities of winning necessarily increase with the addition of the advantage.

Nevertheless, Proposition 4.1 still holds because the point that maximizes the probability of winning remains the district median. The politician wants a platform that maximizes the probability that he or she wins a seat; that is, the politician maximizes the probability of winning the support of at least half of the voters. The incumbency effect merely adds a constant to the internal part of the maximand in Proposition 3.1, i.e., $W_x(x; y, m) = 1 - G(u(y - m) + \alpha_i - u(x - m))$. The maximum value is obtained at $x^* = m$. A similar argument holds for incumbents in party $Y$.

Because the incumbents' induced ideal points do not change, the collective decisions of the parties in period 1 will still be determined by the median district won by the party in period 0. However, as noted above each candidate has their own cut-point. Therefore, there is no longer a well defined party cut-point such that all districts left of the party cut-point vote for the left party. Hence, there is no longer a simple formula for period 1 platforms. Period 1 platforms are, nevertheless, still easy to calculate given a full description of election outcome of period 0. Simply rank the district medians of all districts won by a party in period 0. The median district from such a ranking will be the platform choice in period 1.

The incumbency advantage does change the comparative statics of the model with respect to $v_0$ and $v_1$. Electoral outcomes will be "stickier." Specifically, the incumbency effect makes each politician in both parties safer. Past shocks will be longer-lived in the presence of an incumbency advantage. Seats that a party won on the margin will be harder to defeat in future elections. By the same token, short-run shocks in the future will have a smaller effect on election outcomes. A good $v_0$ followed by a slightly bad $v_1$ could result in the same outcome as a good $v_0$ followed by $v_1 = 0$.

That said, one should keep in mind that the total independence of proposition 3.1

from incumbency effects depends heavily on two assumptions: candidates care solely about their own reelection and candidates are short-sighted. If candidates cared about other factors, such as the probability their party wins a majority, then the higher the chance a given incumbent has of winning their own seat, the higher relative weight they would assign to gaining a majority. In such a case, a higher incumbency effect would lead to more moderate platforms. Preliminary results on long time horizon versions of the model show candidates tend to be more moderate, in an attempt to shrink the other party and therefore cause it to become more extreme. If a candidate is likely to be re-elected, then they put more weight on the future, and therefore are more willing to accept a more moderate platform today in order to make the other party more extreme tomorrow. The incumbency effect, by making incumbents safer and therefore more patient, would lead to more moderate platforms. Formalizing this argument is left to future work.

In summary, a positive incumbency effect can make valence effects stickier in the basic model, but does not lead to more or less convergence. Candidates caring about the future or party majorities can decrease divergence directly. While there is no direct effect of incumbency on convergence, there is an interaction effect between incumbency and candidate preferences for majorities or the future; a positive incumbency effect can reduce divergence further.

### 2.5.3   Gerrymandering and Bias

One of the functions of gerrymandering is to alter legislative district lines so as to benefit one party over another. Most of the extensive literature on gerrymandering focuses on a particular definition of this concept – electoral system bias. A system is unbiased if, in a hypothetical election in which the parties split the vote evenly, the parties win equal shares of the seats. Deviations in seat shares from 50-50, when the votes are split 50-50, are taken as the degree of bias. This is a characteristic of the function $H$. Empirical research implements this definition by regressing seat shares on vote shares and then examining the predicted seat shares of a party when its vote share equals .5. This idea is used to compare electoral systems broadly.

The model here suggests an alternative way to think about the overall degree of electoral bias produced by gerrymandering. As noted above, most analysts think of partisan gerrymandering only in terms of the direct mapping from votes to seats, or more precisely what is the value of $H(c)$ when the parties divide the electorate evenly.

Spatial models that predict convergence are uninformative about this question. In Downs's model, where parties are teams, gerrymandering has no affect on electoral outcomes. [11] The parties converge to the national median and elections in all districts end in ties, regardless of $H$. The same is true in models where there are no parties and candidates are free to choose whatever position they want within their districts. [12]

In this paper's model, gerrymandering can have three affects. First, it alters the *seat shares* received by each party directly by changing the shape of $H$. Second, altering the shape of $H$ can indirectly affect the share of seats won by a party because it can lead parties to adopt new *platforms*. The full effect of the gerrymander, then, would be the change in seats of the party that resulted from the change in $H$ and the change in platforms. Finally, gerrymandering can cause a mismatch between aggregate median voter and the median voter of the median district, $H^{-1}(0.5)$. Since party platforms only depend on $H$, this can cause platforms to be nonrepresentative of the general public. This third effect is discussed in more detail in the next chapter.

The first effect is the effect most political scientists look for. Does the shape of $H$ created by a districting plan imply a disadvantage for one of the parties, holding the parties fixed at their positions.

The second effect, however, has been missed in the voluminous formal literature on districting and gerrymandering. A gerrymander can force the opposing party to adopt a platform that makes it a permanent minority. Consider the possible effect of racial districting. Racial gerrymandering may change the makeup of Democratic

---

[11]Non spatial models of voter behavior, such as voters with fixed partisan loyalty, can predict electoral outcome changes from gerrymanders. For an example see Shotts (2001).

[12]Note in a Downsian model, Gerrymanders can still effect policy outcomes by moving the median legislator. Shotts (2002) provides a formal model describing gerrymandering given purely ideological legislators. Epstein and O'Halloran (1999) provide an empirical case of the median legislator being moved to accommodate racial minority-majority districts. In their example the median legislator becomes, perversely, less friendly to minority rights.

incumbents by creating a disproportionate number of extremely left-wing districts in that party. That, in turn, might change the platform of the Democratic party as well, most likely shifting it to the left. Thus, the party balance can depend on the shape of how districts are distributed, in addition to the median (or average) district.

To illustrate this more precisely, consider the following examples.

**Example 5.0.** Let voter utility be symmetric and let $H(z) = z$ where $z \in [0, 1]$.

The uniform distribution implies that the zero-valence stable cut-point is $1/2$, the parties will locate at $1/4$ and $3/4$, and the parties will each win half of the seats.

**Example 5.1.** Let voter utility be quadratic, and let $H(z) = z^2$ where $z \in [0, 1]$.

As shown in the appendix, the cut-point corresponding to the unique zero-variance stable platforms (given $x < y$) is approximately $.653$. The left party locates at $.462$; the right party is at $.845$; and the right party wins approximately 57% of the seats (i.e., $H(c) \approx 0.427$).[13]

Assume, for the sake of simplicity, that the underlying distribution of *voters'* ideal points is the same in both of these circumstances but that a clever political cartographer managed to reshape district lines so that the initial uniform distribution of district medians became quadratic. Doing so has several effects.

First, it changes the quantiles of the distribution of district medians. In example 4.1., the median district is at $.707$, instead of $.5$, the 25th percentile district is at $.5$, and the 75th percentile is at $.866$.

Second, relative to these quantiles, the left party is more extreme and the right party is more moderate. The left party is $.207$ units below the median district, while the right party is $.159$ units above the median.

---

[13]Example 5.1 is by no means unique. Another good example is a piecewise uniform district density with support between 0 and .75, median at 0.5, and density of 1 in $[0, 0.5)$ and 2 in $(0.5, 0.75]$. In other words, example 5.0 with the right most districts "compressed." Analogous examples can be constructed where the left most districts have been "expanded." In this example, it is easy to show that the unique zero valence stable platforms for quadratic utility are 0.2 and 0.6, the corresponding cut-point is 0.4, and $H(0.4) = 0.4$. That is, the left party with its long tail of districts becomes relatively more extreme, the right party relatively more moderate, and the left party receives only 40% of the vote.

Third, the right party gains an electoral advantage. The cut-point between the two parties lies to the left of the median. The left party becomes something of a permanent minority, as its zero-order stable seat share fell from 50 percent to 43 percent. This, then, is the effect of redistricting on the electoral system.

This analysis also highlights some difficulties with the traditional seats-votes measure of bias. Interpreting the relationship between seats-votes requires mapping the distribution of voters as well as districts. Assuming that mapping is straightforward (chapter 3 shows it is not), there is the further problem with interpreting the hypothetical case when the vote is evenly divided. Suppose the parties have taken diverging and asymmetric positions around the median. Then, only a large and asymmetric shock would produce the 50-50 division of the vote, but such as shock would necessarily lead the parties to alter their policy platforms (in equilibrium). Hence, the hypothetical 50-50 division of the electorate would be out of sample in two respects: not observed in the data, and very unlikely to be observed as it is not in equilibrium.

An example demonstrates a situation where this is clearly the case. Assume that voters' ideal points are uniform, as in Example 4.0, and unchanged by redistricting. Call the distribution of voters' preferences $F(z)$. However, redistricting alters the distribution of district medians from uniform on $Z$ to quadratic, $H(z) = z^2$. In the pre-redistricting state the parties locate at the quartiles and each party wins half of the vote and half of the seats. Under the gerrymander, as already shown, the cut-point between the parties is .653, and the left party will receive 43 percent of the seats. This would be an extreme gerrymander: party $X$ wins 65 percent of the vote and 43 percent of the seats.

Now consider the traditional definition of party bias. Holding fixed the parties' positions under example 4.1, what is the division of the seats occurs if the division of the votes equals .5? In the terms of the model, there must be a valence shock that makes the vote division equal, holding the parties at platforms $x = .43$ and $y = .84$. This would require a large negative valence shock. Again assuming uniform voter ideal points, a cut-point of .5 would produce vote shares of $F(.5) = .5$, and party $X$'s seat share would fall to just .25. This would mean an enormous "bias" of .25. The

zero-order stable definition leads to a much smaller bias of .07.

One could argue that the measure offered above – the expected or zero-order seat loss associated with the change in $H$ – is preferable to the traditional definition of bias. The enormous bias predicted by the traditional measure is not usually in empirical analyses an in-sample value. But, as the discussion highlights, there is a conceptual problem as well, it does not account for the fact that the parties have changed their positions to accommodate the new districting map.

As discussed above, the model also provides a specific formula for the effects of gerrymandering: $E(c_1^*) = (1/2)[H^{-1}(H(c_0)/2) + H^{-1}((H(c_0) + 1)/2)]$. In this case $H$ is what is being varied. One could either use the formula as a way to predict the consequences of a proposed future gerrymander, or test the model by testing the formula against historical gerrymanders. Note that in any given election with a given cut-point in the prior election, only certain parts of $H$ impact the current cut-point. For example, if a gerrymander only tinkered with districts in the extreme left tail while leaving the rest of the distribution the same, the model would only predict an impact when the previous election went strongly against the left party. A gerrymander that has a small impact one year might have a major impact the next. Or, districting plans that only operate in the tail of the distribution of one party's districts and change the party median little, may have no substantial effect on electoral competition.

The gerrymander suggested by these examples is severe in another sense. It substantially altered the policy outcomes relative to the preference of the median voter. Assuming, for the sake of argument, that the median is at $z = .5$. In example 5.0, policy will be either $z = .25$ or $.75$, depending on which party wins. In example 5.1, the right party will win and policy will be at $z = .845$. This is very extreme relative to the median voter's preferences, and all voters but those to the right of the $z = .8(= (.845 + .75)/2)$ will be worse off.

This contrast raises a deeper welfare issue. In assessing gerrymandering, one should first ascertain whose benefit is at stake. If one cares first and foremost about elites – about elected officials and parties – then the definition of bias as seat loss

may make sense. However, if the ultimate goal is representational bias and one cares about voter utility, then the relevant comparison is not between seats and votes or even seat loss but how far party platforms deviate from the median. [14]

## 2.6 Extension: Platform Choice in the Party List System

In the list system, the parties offer a list of candidates running under their label, and the entire national electorate votes for one of the two parties. Parties win shares of seats equal to their shares of the vote. The number of seats won by the party equals the number of seats times the share of seats it deserves. An individual candidate within a party wins a seat if that candidate's rank on the list is higher than total number of seats the party won in the election.

Party list systems are often viewed as having a very different sort of politics than districted systems because the candidates face very different electorates and are chosen through a very different sort of mechanics. Surprisingly, an analogous logic and results characterize politicians' induced preferences and party positioning in party list and districted systems.

Slightly different notation is required to describe the electoral competition in list systems. Parties win seats in proportion to their votes in the national electorate, rather than the fraction of districts won. As before, the cut-point $c$ defines the ideal point along $Z$ of the voter indifferent between the two parties, $X$ and $Y$. The cumulative density of the voters' ideal points evaluated at the cut point $c$ determines the fraction of seats won by the left party $X$, and the right party, $Y$, wins the remaining seats. We will denote the cumulative density of voters ideal points as $F(z)$, and assume $F(z)$ is continuous and has inverse for $\tau \in (0, 1)$ $F^{-1}(\tau)$.[15] The

---

[14]For an example of thinking about bias in terms of voter welfare, see Coate and Knight (2005). They show that in some circumstances a biased seats/votes curve can actually increase voter welfare. However, in their model party platforms are fixed. The model indicates that to fully assess the welfare implications of bias, one must also account for how the shape of districts impacts party platforms.

[15]Unless the policy space $Z$ is bounded, $F^{-1}(\tau)$ may be undefined at $\tau = 0$ and $\tau = 1$. Since these represent measure zero worth of candidates on the party list one can safely ignore these cases.

left party wins a share $s$ of votes (and of seats) equal to $s \equiv F(c)$ and the right party wins $1 - F(c)$ share of votes.

In the party list case, $F(z)$ plays the same role as $H(z)$ in the districted case. Nevertheless it is important to remember the distinction. $F(z)$ is a description of voter ideal points across an entire nation; $H(z)$ is a description of median voter ideal points, district by district.

We will represent the positions on each party's list as the interval $[0, 1]$, where 0 denotes the top of the list and 1 denotes the bottom. A candidate at the $\lambda th$ position on the list wins a legislative seat if and only if the candidate's party receives a vote (and seat) share larger than the candidate's position on the list, i.e., $\lambda < s$. The $\lambda th$ candidate on party $X$'s list has the following utility function

$$Q_X = Prob(s \geq \lambda) \tag{7a}$$

The $\lambda th$ candidate on party $Y$'s list has the following utility function

$$Q_Y = Prob(1 - s \geq \lambda) \tag{7b}$$

These equations correspond to the generic formulation on page 7. Further, may also write these probabilities in terms of $F$ using the equality $s = F(c(x, y, v)))$.

Equations (7a) and (7b) define candidates' preferences over party platforms. As in the single-member district model, candidates' expected utilities can be written in terms of voters' utility functions. For $x < y$, $\lambda \in (0, 1)$:

$$Q_X(x, y, \lambda) = Prob(\Delta U(x, y, v, F^{-1}(\lambda)) \geq 0) \tag{8a}$$

and

$$Q_Y(x, y, \lambda) = Prob(\Delta U(x, y, v, F^{-1}(1 - \lambda)) \leq 0). \tag{8b}$$

Symmetric equations hold for $x > y$. Note the similarity between these pairs of equations and equations (3a)-(3b), which describe the districted case.

Let $\tilde{x}(y, \lambda)$ be the most-preferred platform of a candidate who is at position $\lambda$

on party $X$'s list, and let $\tilde{y}(x, \lambda')$ be the most-preferred platform for a candidate at position $\lambda'$ on party $Y$'s list. As the following proposition shows, $\tilde{x}(y, \lambda)$ and $\tilde{y}(x, \lambda')$ are well-defined.

**Proposition 6.1.** Suppose $y \geq 0$. For all $\lambda \in (0, 1)$ and $y$, $Q_X(x, y, \lambda)$ is either double-peaked or single-peaked in $x$, and has a unique global maximum. This global maximum is attained at $\tilde{x}(y, \lambda) = max(F^{-1}(s), y)$.

A symmetric result holds for the case $y \leq 0$, and an analogous proposition holds for $Q_Y(x, y, \lambda)$. The preferences of the person exactly at the top and bottom of the list are somewhat pathological, but they constitute a set of measure zero and so are ignored.[16]

This result has an interesting implication for the understanding politicians' preferences under list systems. There is a seeming divide among political scientists studying districted and list systems. Those studying districted systems, such as the United States, frequently assert that politicians just want to win office, while those studying list systems often maintain policy-oriented politicians. Proposition 5.1 suggests that the difference is more apparent than real. Election oriented politicians' positions on the list induce policy-oriented preference functions.

Candidates near the top of their party's list have the strongest preference for adopting extreme policies. The intuition is straightforward, and reflects the trade-off between a party's expected vote-share and the variance of the party's vote-share. Consider a candidate at the 25th percentile on the party $X$'s list ($\lambda = 1/4$). This candidate wins a seat as long as $X$ wins at least 25% of the vote. If $X$ chooses a platform that diverges considerably from $Y$'s platform, then there will be a subset of voters with a strong policy-based preference for party $X$. If $x < y$, for example, then voters with $z < (x + y)/2$ all prefer $X$ to $Y$ on policy grounds, and this preference grows more intense the smaller is $z$. Voters with a strong policy-based preference for party $X$ are very likely to vote for $X$, since it will require a large, negative valence

---

[16]Unless one requires the support of $F$ to be bounded, the person on the top wants infinite divergence; the person at the bottom wants full convergence, and is indifferent among all other platform choices.

shock to cause them to change their minds and vote for $Y$. Thus, even if party $X$ is unlikely to win a majority of the vote, it might be *very* likely to win 25 percent or more of the vote. On the other hand, if $X$'s platform is close to $Y$'s, then there is a higher degree of uncertainty about $X$'s vote-share, since the valence shocks dominate voter choices. Although $X$ expected vote-share may be higher, the variance of it's vote-share will also be higher, and the probability that $X$ wins at least 25% of the vote may fall. As shown in Proposition 5.1, if $y > 0$ then the probability that $X$ wins at least 25 percent of the vote is maximized when $x$ is equal to the 25th percentile of the voter ideal-point distribution.

By contrast, candidates low on their party's list all want the same platform, a "me too" platform equal to that of the other party. Such candidates actually *welcome* some degree of electoral uncertainty, since their only chance of winning a seat in the legislature lies in a favorable valence shock for their party.

Suppose that each party's incumbent legislators are placed at the top of the party's list. Recall that each party's platform is chosen by simple majority rule among the party's incumbents. Denote the current period as period 1, and let $s_0$ be party X's seat-share from the election in period 0. Assume that $s_0 \in (0,1)$, so neither party was completely wiped out in period 0's election. Although some incumbents do not have single-peaked preferences, their preferences satisfy a weaker condition that guarantees the existence of a unique majority winner. The equilibrium platforms are simply characterized, as follows.

**Proposition 6.2.** Suppose $F^{-1}$ and $u$ are differentiable, $s_0 \neq 0.5$, and $0 < s_0 < 1$, Then the equilibrium core platforms for the current election satisfy (i) $x_1^* = F^{-1}(s_0/2)$ and $y_1^* = F^{-1}(s_0/2 + 1/2)$, or (ii) $x_1^* = F^{-1}(1 - s_0/2)$ and $y_1^* = F^{-1}((1 - s_0)/2)$.

According to Proposition 6.2 each party's platform for the election at time 1 is equal to the median of the ideal points of the voters who supported the party in the election at time 0. In case (i) $x_1^* < 0 < y_1^*$, and in case (ii) $y_1^* < 0 < x_1^*$. Note that the equilibrium platforms are always on opposite sides of the overall median ideal point.[17]

---

[17] If the cut-point is at the median, there exists a convergent equilibrium as well as two divergent

Proposition 6.2 is nearly identical to Proposition 4.2 (as long as the cut-point is never *exactly* at the median,) and so most the conclusions that derive from Proposition 4.2 (divergence, past performance effects current platforms and performance, and performance depends on the entire distribution of voters) apply, with one major exception: there is no guarantee that party ordering will be the same. If one exogenously imposes party ordering then the standard predictions would apply. If one does not impose the restriction exogenously, then there is nothing about the model that prevents party's in part-list systems from "swapping" left-right positions.

Although similar results on party-positioning hold for list and districted systems, the model also helps clarify the differences between them.

First, induced preferences are well-defined and concave for all legislators in the districted system, but oddities arise in the list system. Most notably, those legislators on the bottom of the list want perfect convergence exactly, as that plus a favorable shock is their only hope of winning a seat. The analogous legislators in districted systems are those in moderate districts or districts leaning toward the other party. Those in moderate districts want moderate platforms, while those in the "wrong districts" (left party politicians in districts right of the median and vice versa) want their party to be more conservative than the median district.

Second, electoral competition in the list system depends only on the distribution of voters' preferences, not the intermediary mapping of voters' preferences into districts and then into seats. Propositions 3.2 and 5.2 show that the platforms adopted by the parties depend on the relevant electoral distribution functions – either of voters' ideal points or of district median ideal points. If a particular country were to switch from a list system to a districted system or vice versa, party platforms may change significantly depending on the mapping of voters into districts. If, for example, voter ideology were driven by income, the distribution of ideology in the entire nation may have a thick right tail. A districted system, on the other hand, might be, district median by district median, more balanced. As example 3.1 illustrates, a right-skewed

---

equilibria. The intuition is that the pivotal incumbent has a 50 percent chance of winning in both the convergent and divergent equilibria.

distribution can result in a small right-party. This result may give some intuition for why socialist parties are relatively weak in districted systems and strong in party list systems.

## 2.7   Conclusion

The Median Voter Theorem has provided a focal point for the spatial theory of politics. Two parties, acting as teams to control government, will converge to the ideal point of the median voter in the electorate. But, parties are not teams - at least not in the strict sense that Downs and others have expressed it. Rather parties are collections of politicians motivated by their personal desire to hold office as well as grander ideological goals. Legislators benefit from and enjoy personal benefits from holding their own seat in the legislature. Starting from this assumption, the model comes to a strikingly different conclusion.

When parties are not teams, parties will represent the median of their elected officials, rather than the median of the electorate as a whole. This result applies equally to single member district systems and party list systems, which suggests that party divergence is a universal phenomenon driven by the personal interests of politicians. Politicians' desire to hold office will lead parties to diverge. The degree of divergence expected is approximately the inter-quartile range within the electorate.

Electoral systems themselves determine whether parties are teams. The relevant question is this. If all politicians seek to win office, are their induced preferences over policy identical within each party? In neither districted nor list systems was this true. Hence, in neither system should the team assumption be taken lightly. There are, however, some electoral systems that would produce parties as teams. Consider, for example, a system in which all legislators are elected at-large from the nation as a whole, voters have as many votes as positions, and voters must vote for all candidates. The theory predicts parties as teams in this circumstance because all candidates will have the same *induced* policy preferences.

The model is arguably a good starting point for the empirical and theoretical

43

development of a model of parties that does not assume unified times. Empirically, the model generates several basic predictions in agreement with the basic facts of electoral competition. Most notably, elections are not ties in all districts, and the parties are diverged. The simple spatial model in which parties are teams or in which candidates compete district-by-district do not generate these predictions, but instead predict that the patently false prediction that all elections end in ties.

As a theoretical matter, the model has been simplified in several respects in order to clarify the importance of the team assumption. Many analytical extensions are possible and desirable. They produce a wider range of predictions, but are cases of the general results here. First, we have assumed only two parties, following much of the literature. Multiple parties or a system with entry is much more complicated, and requires assumptions about coalition formation and voter sophistication. Second, one can expand the party beyond the incumbent members of the legislature to incorporate Presidents, non-incumbent candidates, multi-body legislatures, and activists. Some of these extensions, such as a President, will create pressures for convergence, while others create centripetal forces. Third, politicians might have objectives beyond winning their own seats, including their own policy preferences and desire for more power within the legislature. Again, some of these extensions, such as winning a majority, push the model in the direction of the team assumption, while others create more dissention within the party. Fourth, a variety of technical complications are likely of interest, including longer time horizons and multiple dimensions.

While we have focussed primarily on positive implications of the model, there are normative implications as well. For example, in the world of the model, most incumbents are significantly better off than they would be if parties were forced to maximize seat shares, or if parties did not have to run under a common banner. In those cases, in equilibrium, each incumbent would only have a 0.5 chance of winning each election. In the model here, all but the most marginal incumbents will have a greater chance of winning their own personal seat.

There is a welfare implication for voters as well. When parties are not teams, the parties diverge, and one of the two parties' policy platforms will win and become

law. The winning platform will deviate considerably from the median, and thus a majority of the voters would prefer a policy that is closer to the median. There is little voters can do to change this situation. The majority of politicians within each party support the divergent positions that the parties take: those positions maximize the politicians' chances of winning office. Voters cannot achieve the same degree of coordination. They are left with a policy that is less than optimal for a large majority.

While the predictions of this model have direct impact on fundamental concerns of democracy, such as election outcomes and representation, one of the primary inputs to many of the predictions is the distribution of voters across districts and the nation. The remaining chapters discuss how to find these distribution.

## 2.8 Appendix: Proofs

**Proof of Proposition 3.1.** We will only show the case $x < y$ here. Consider two values $z$ and $z'$, where, without loss of generality, $z' > z$. Then either $z' - x > z - x \geq z' - y > z - y$ or $z' - x > z' - y > z' - y > z - y$. We need to show that

$$\Delta U(z; x, y, v) > \Delta U(z'; x, y, v) \,,$$

that is,

$$v - u(z - x) + u(z - y) > v - u(z' - x) + u(z' - y)$$

$$u(z' - y) + u(z - x) < u(z - y) + u(z' - x) \,.$$

Define $\alpha$ and $\alpha'$ implicitly by

$$z' - y = \alpha(z - y) + (1 - \alpha)(z' - x)$$

$$z - x = \alpha'(z - y) + (1 - \alpha')(z' - x) \,.$$

Adding the two relations above and manipulating yields

$$(\alpha + \alpha' - 1)(z - y) = (\alpha + \alpha' - 1)(z' - x) \,.$$

Since $(z' - x) > (z - y)$ this implies $\alpha + \alpha' = 1$. By strict convexity

$$u(z' - y) < \alpha u(z - y) + (1 - \alpha)u(z' - x)$$

$$u(z - x) < \alpha' u(z - y) + (1 - \alpha)u(z' - x) \ .$$

Adding the two relations above and using the result $\alpha + \alpha' = 1$ yields

$$u(z' - y) + u(z - x) < u(z - y) + u(z' - x)$$

as desired. ■

To prove Proposition 3.2 we first prove two lemmas.

**Lemma 3.1a** Let $x < x' < y$. $\exists \epsilon > 0$ such that if either $c(x, y, v) > x'$ or $|v| < \epsilon$ then either $c(x', y, v) > c(x, y, v)$ or all voters vote for party $X$.

*Proof of Lemma 3.1a* There are two cases. Case one is $c(x, y, v) > x'$. Then the voter with ideal point $c(x, y, v)$ strictly prefers $x'$ to $x$ and therefore $\Delta U(c(x, y, v); x', y, v) > 0$. By proposition 2.1 $\Delta U$ is strictly decreasing in ideal points, and therefore the solution $z^*$ to $\Delta U(z^*; x', y, v) = 0$ must be greater than $c(x, y, v)$ or not exist. But if $z^*$ exists it is by definition the cut-point $c(x', y, v)$. If it does not exist then $\Delta U(z; x', y, v) > 0$ for all $z$. That is all voters prefer party $X$.

Case two is $c(x, y, z) \leq x'$. If $v = 0$ a voter with ideal point $x'$ strictly prefers party $X$ to $Y$. That is $\Delta U(x'; x', y, 0) > 0$. By continuity of $\Delta U$ with respect to $v$, there exists $\epsilon > 0$ such that $\forall |v| < \epsilon$, $U(x'; x', y, v) > 0$. Then for such $v$ by a similar argument as case one, $c(x', y, v) > x'$ or all voters support party $X$. But in this case $c(x', y, z) > x' \geq c(x, y, z)$.■

**Lemma 3.1b** Let $x < y < y'$. If $c(x, y, v) < y$ then either $c(x, y', v) > c(x, y, v)$ or all voters vote for party $X$. Furthermore, $\exists \epsilon > 0$ such that if $|v| < \epsilon$ then $c(x, y, v) < y$ and therefore either $c(x', y, v) > c(x, y, v)$ or all voters vote for party $X$.

*Proof of Lemma 3.1b* First consider the case where $c(x, y, v) < y$. Then the voter

46

with ideal point $c(x, y, v)$, given $v$, strictly prefers $y$ to $y'$, is indifferent between $x$ and $y$ (by definition of cut-point), and therefore strictly prefers $x$ to $y'$. Therefore $\Delta U(c(x, y, v); x, y', v) > 0$. By proposition 2.1 $\Delta U$ is strictly decreasing in ideal points, and therefore the solution $z^*$ to $\Delta U(z^*; x, y', v) = 0$ must be greater than $c(x, y, v)$ or not exist. But if $z^*$ exists it is by definition the cut-point $c(x, y', v)$. If it does not exist then $\Delta U(z; x, y', v) > 0$ for all $z$. That is all voters prefer party $X$.

For the second part, if $v = 0$ and party $X$'s platform is at $x$ and party $Y$'s platform is at $y$, then a voter with ideal point at $y$ strictly prefers party $Y$ to $X$ and a voter with ideal point $x$ strictly prefers party $X$ to $Y$. Therefore by proposition 2.1 $c(x, y, 0)$ exists and $x < c(x, y, 0) < y$. By continuity of $\Delta U(y; x, y, v)$ and $\Delta U(x; x, y, v)$ with respect to $v$, $\exists \epsilon$ such that $\forall |v| < \epsilon$ a voter with ideal point at $x$ still strictly prefers $x$ to $c(x, y, v)$ and a voter with ideal point at $y$ still strictly prefers $y$ to $c(x, y, v)$. Therefore $x < c(x, y, v) < y$. ∎

Note, Lemmas 3.1a and 23.1b imply that as long as the original cut-point lies between the two platforms, and platform shifts do not cross the cut-point, then the cut-point is monotonic in platforms. Roughly speaking, we need either small $v$ combined with potentially large platform changes, or small to medium sized $v$ (such that the cut-point stays between the parties) combined with small changes in platforms to ensure monotonicity.

**Proof of Proposition 3.2** Case one is $y' - y < x' - x$. Let $\delta = y' - y$. Since $\Delta U(z; x, y, v) = \Delta U(z + \delta; x + \delta, y + \delta, v)$, $c(x + \delta, y + \delta, v) = c(x, y, v) + \delta$. Therefore $c(x + \delta, y', v) > c(x, y, v)$. For this case $x' > x + \delta$. Then apply Lemma 2.1a to show $\exists \epsilon$ such that $\forall v$ such that if $v > -\epsilon$ then $c(x', y', v) > c(x + \delta, y', v)$ or all voters vote for party $X$.

Case two is $x' - x < y' - y$. Let $\delta = x' - x$. Since $\Delta U(x, y, v, z) = \Delta U(x + \delta, y + \delta, v, z + \delta)$, $c(x + \delta, y + \delta, v) = c(x, y, v) + \delta$. Therefore $c(x', y + \delta, v) > c(x, y, v)$. For this case $y' > y + \delta$. Then apply Lemma 2.1b to show $\exists \epsilon$ such that $\forall v$ such that if $v < \epsilon$ then $c(x', y', v) > c(x', y + \delta, v)$

Case three is $x' - x = y' - y$. Let $\delta = x' - x$. Since $\Delta U(x, y, v, z) = \Delta U(x +$

47

$\delta, y + \delta, v, z + \delta)$, $c(x + \delta, y + \delta, v) = c(x, y, v) + \delta$. Therefore for this case $c(x', y', v) > c(x, y, v)$.∎

**Proof of Proposition 3.3.** Recall that a cut-point is defined as the solution to $-u(z - x) + u(z - y) + v = 0$. Since $u$ is symmetric, when $v = 0$ the solution is $z = (x + y)/2$. For $v \neq 0$, write the solution as $c(x, y, v) = (x + y)/2 + \delta_+$. Then $-u((x - y)/2 + \delta_+ + u((y - x)/2 + \delta_+) + v = 0$. But by symmetry of $u$ the following must hold

$$-u((y - x)/2 - \delta_+) + u((x - y)/2 - \delta_+) + v = 0$$
$$u((y - x)/2 - \delta_+) - u((x - y)/2 - \delta_+) - v = 0$$

But this implies that the cut-point for $-v$ is $c(x, y, -v) = (x + y)/2 - \delta_+$. Therefore $c(x, y, v) + c(x, y, -v) = x + y = 2c(x, y, 0)$. If $G$ is antisymmetric around zero, then the expected value of the cut-point is:

$$
\begin{aligned}
E(c(x, y)) &= \int_{-\infty}^{\infty} c(x, y, v) dG \\
&= \int_0^{\infty} c(x, y, -v) dG(v) + \int_0^{\infty} c(x, y, v) dG(v) \\
&= 2c(x, y, 0) \int_0^{\infty} c(x, y, -v) dG(v) \\
&= 2c(x, y, 0)(1/2) \\
&= c(x, y, 0) .∎
\end{aligned}
$$

**Proof of Proposition 4.1.** The expected utility of a candidate from party $X$ with district median $m$ is

$$
\begin{aligned}
Q_X(x, y, m) &= Prob(\Delta U(m; x, y, v) \geq 0) \\
&= 1 - G(u(x - m) - u(y - m))
\end{aligned}
$$

Since in equilibrium $y$ is treated as a constant, $u(x - m) - u(y - m)$ represents the same preferences as $u(x - m)$. $G$ is a CDF and therefore is monotonic. Any maximum of

48

$-u(x-m)$ is a minimum of $u(x-m)$, a minimum of $G(u(x-m))$ (by monotonicity), and a maximum for $1 - G(u(x-m))$. Since $m$ is a maximum for $-u(x-m)$, it is also a maximum for $Q_X(x, y, m)$. If $v$ has full support, then $G$ is strictly increasing. Now, $u(x-m)$ represents the opposite preferences to $-u(x-m)$, $G(u(x-m))$ represents the same preferences as $u(x-m)$ by strict monotonicity, and $1 - G(u(x-m))$ represents the opposite preferences to $G(u(x-m))$. Therefore, $Q_X(x, y, m)$ represents the same preferences as $-u(x-m)$. Similar arguments apply for $Q_Y(y, x, m)$. ∎

**Proof of Proposition 4.2.** Since $0 < H(c_0) < 1$, each party has positive vote share. Since $x_0 < y_0$ every district with median less than $c_0$ votes for party $X$ in period 0. Since $v$ has full support, by Proposition 3.1 party $X$ ideal points correspond to district medians in equilibrium. Therefore, incumbent members of party $X$ have ideal points given by the CDF $H(z)/H(c_0)$ for $z < H(c_0)$. The median incumbent ideal point is defined by the equation $H(z_m^*)/H(c_0) = 1/2$. Solving yields $z_m^* = H^{-1}(H(c_0)/2)$. Consider any alternative platform $x'$. If $x' > z_m^*$ then there exists $\epsilon$ that defines a majority coalition strictly preferring $z_m^*$ to $x'$. This coalition consists of all incumbents with districts from $H^{-1}(0)$ to $H^{-1}(H(c_0)/2 + \epsilon) < x$. A similar coalition exists if $x' < z_m^*$. However, there does not exist a majority coalition that prefers any $x \neq z_m^*$. Therefore, $z_m^*$ is the unique core point. A similar argument applies to show $y_1^*(c_0) = H^{-1}((H(c_0) + 1)/2)$. Strict monotonicity of $H$ and $H^{-1}$ implies $x_1^* < y_1^*$. ∎

**Proof of Proposition 4.3.** The result is immediate from Proposition 3.2 and the definition of zero-valence stable. ∎

**Proof of claims in Example 5.1.** Suppose $z \in [0, 1]$ and $H(z) = z^2$. Then the PDF is $h(z) = 2z$, and the inverse is $H^{-1}(\theta) = \sqrt{\theta}$. The average is $\bar{z} = \int_0^1 zh(z)dz = \frac{2}{3}$. The median is determined by the equation $\frac{1}{2} = z_m^2$, so $z_m = \frac{\sqrt{2}}{2}$.

By Proposition 4.2, the definition of zero-variance stable platforms, and the formula for cut-point with quadratic utility,

$$c = \frac{1}{2}(\sqrt{\frac{c^2}{2}}) + \frac{1}{2}(\sqrt{\frac{c^2}{2} + \frac{1}{2}})$$

After some algebra this simplifies to

$$c = \frac{\sqrt{2}}{2}\left(\frac{1}{\sqrt{4 - 2\sqrt{2}}}\right) < \frac{\sqrt{2}}{2} = z_m \ .$$

The formulas in the text follow immediately. ∎

**Proof of Proposition 6.1.** First, suppose $x = y$. Then $s_X = 1$ if $v > 0$, $s_X = 0$ if $v < 0$, and $s_X = 1/2$ if $v = 0$. Thus, $Q_X(x, y, s) = Prob(v \geq 0) = 1/2$ for all $s$.

Next, consider $x < y$. Then

$$
\begin{aligned}
Q_X(x, y, s) &= Prob(\Delta U(F^{-1}(s); x, y, v) \geq 0) \\
&= Prob(v > u(x - F^{-1}(s)) - u(y - F^{-1}(s))) \\
&= 1 - G(u(x - F^{-1}(s)) - u(y - F^{-1}(s))) \qquad (A1)
\end{aligned}
$$

Note that $\lim_{x \to y} 1 - G(u(x - F^{-1}(s)) - u(y - F^{-1}(s))) = 1 - G(0) = 1/2$, so $Q_X(x, y, s)$ is continuous in $x$ on $(-\underline{a}, y]$ for all $s$. By a similar argument (since the utility function is the same except instead of $m$ we have $F^{-1}(s)$) as in the proof of Proposition 3.1, $Q_X(x, y, s)$ is single-peaked in $x$ over the interval $(-\underline{a}, y]$. If $s \leq F(y)$ then $Q_X(x, y, s)$ attains its unique maximum on $(-\underline{a}, y]$ at $\tilde{x}(s, y) = F^{-1}(s)$, and if $s > F(y)$ then $Q_X(x, y, s)$ attains its unique maximum on $(-\underline{a}, y]$ at the "corner," $\tilde{x}(s, y) = y$.

Next, consider $x > y$. Then

$$
\begin{aligned}
Q_X(x, y, s) &= Prob(\Delta U(F^{-1}(1-s); x, y, v) \geq 0) \\
&= Prob(v > u(x - F^{-1}(1-s)) - u(y - F^{-1}(1-s))) \\
&= 1 - G(u(x - F^{-1}(1-s)) - u(y - F^{-1}(1-s))) \qquad (A2)
\end{aligned}
$$

Again, $Q_X(x, y, s)$ is continuous in $x$ on $[y, \underline{a})$ for all $s$, since $\lim_{x \to y} 1 - G(u(x - F^{-1}(1-s)) - u(y - F^{-1}(1-s))) = 1 - G(0) = 1/2$. Again, as by a simular argument as Proposition 4.1, $Q_X(x, y, s)$ is single-peaked in $x$ over the interval $(y, \underline{a})$. If $s \leq 1 - F(y)$ then $Q_X(x, y, s)$ attains its unique maximum on $[y, \underline{a})$ at $\tilde{x}(s, y) = F^{-1}(1-s)$,

and if $s > 1 - F(y)$ then $Q_X(x, y, s)$ attains its unique maximum on $[y, \underline{a})$ at the "corner," $\tilde{x}(s, y) = y$.

Since $y \geq 0$ by assumption, $1 - F(y) \leq 1/2 \leq F(y)$. Divide the interval $[0, 1]$ into $[0, 1 - F(y))$, $[1 - F(y), F(y)]$, and $(F(y), 1]$. These intervals correspond to cases (i), (ii) and (iii), respectively.

The characterizations for cases (ii) and (iii) are easily derived from the results above. For case (ii), where $s \in [1 - F(y), F(y)]$, $Q_X(x, y, s)$ reaches an interior maximum at $x = F^{-1}(s)$ on the interval $(-\underline{a}, y)$, and it reaches a corner maximum at $x = y$ on the interval $[y, \underline{a})$. Thus, $Q_X(x, y, s)$ is single-peaked, attaining its global maximum at with $\tilde{x}(y, s) = F^{-1}(s)$. For case (iii), where $s > F(y)$, $Q_X(x, y, s)$ reaches a corner maximum at $x = y$ on both $(-\underline{a}, y]$ and $[y, \underline{a})$. Thus, $Q_X(x, y, s)$ is single-peaked and attains its global maximum at with $\tilde{x}(y, s) = y$.

For case (i), where $s < 1 - F(y)$, $Q_X(x, y, s)$ is double-peaked, reaching an interior local maximum at $x = F^{-1}(s) < y$ and at $x = F^{-1}(1 - s) > y$. All that remains is to show that $Q_X(F^{-1}(s), y, s) > Q_X(F^{-1}(1 - s), y, s)$; then $\tilde{x}(y, s) = F^{-1}(s)$ is the global optimum for all $s$ in case (i), and $x = F^{-1}(s)$ is only a local optimum. Substituting, we have

$$Q_X(F^{-1}(s), y, s) = 1 - G(-u(y - F^{-1}(s)))$$

and

$$Q_X(F^{-1}(1 - s), y, s) = 1 - G(-u(y - F^{-1}(1 - s))) \ .$$

These two equations imply that $Q_X(F^{-1}(s), y, s) > Q_X(F^{-1}(1 - s), y, s)$ if and only if $-u(y - F^{-1}(s)) < -u(y - F^{-1}(1 - s))$. Since $F$ is symmetric about zero, $1 - F(y) = F(-y)$. Also, $s < 1 - F(y)$ for the case under consideration, so we have $F^{-1}(s) < -y < 0 < y < F^{-1}(1 - s)$. Again using the symmetry of $F$, $-y - F^{-1}(s) = F^{-1}(1 - s) - y$. Thus, $y - F^{-1}(s) = 2y + F^{-1}(1 - s) - y > F^{-1}(1 - s) - y$. Thus, since $-u$ is symmetric and single-peaked, $-u(y - F^{-1}(s)) < -u(y - F^{-1}(1 - s))$. Thus, $Q_X(F^{-1}(s), y, s) > Q_X(F^{-1}(1 - s), y, s)$ as desired.

Parts (i)'-(iii)', for the case $y \leq 0$, are proved in a similar fashion. ∎

**Proof of Proposition 6.2.** From equations A1 and A2 of the proof of Proposi-

tion 6.1, and the fact that ordinal utility functions are preserved from monotonic transformations, the preferences of candidates in party $X$ can be written as

$$-u(x - F^{-1}(s)) + u(y - F^{-1}(s))$$

if $x < y$ and

$$-u(x - F^{-1}(1-s)) + u(y - F^{-1}(1-s))$$

if $x > y$. Assume $y > F^{-1}(1/2)$. Consider $x_1^* \equiv F^{-1}(s_0/2)$. We will show there exists a majority coalition preferring $x_1^*$ to any other platform $x'$. There are four cases:

(i) if $x' \leq x_1^*$, then all candidates in party $X$ with $s' \geq s_0/2$ prefer $x_1^*$ to $x'$

(ii) if $x_0 < x' \leq y$, then all candidates in party $X$ with $s' \leq s_0/2$ prefer $x_1^*$ to $x'$

(iii) if $y < x' \leq F^{-1}(1 - s_0/2)$, then all candidates in party $X$ with $s' \leq s_0/2$ prefer $x_1^*$ to $x'$

(iv) if $x' > y$ and $x' > F^{-1}(1 - s_0/2)$, then all candidates in party $X$ with $s' \geq s_0/2$ prefer $x_1^*$ to $x'$.

Cases (i) and (ii) are immediate from single peakedness (since $x' \leq y$ and $x_1^* < y$). For cases (iii) and (iv) a candidate in party $X$ at position $s$ strictly prefers $x_1^*$ to $x'$ if

$$-u(x_1^* - F^{-1}(s)) + u(x' - F^{-1}(1-s)) - u(y - F^{-1}(1-s)) + u(y - F^{-1}(s)) > 0 \ .$$

In the proof of Proposition 6.1 we showed $-u(y - F^{-1}(s)) < -u(y - F^{-1}(1-s))$. Therefore a sufficient condition for the above relation to hold is

$$-u(x_1^* - F^{-1}(s)) + u(x' - F^{-1}(1-s)) \geq 0 \ .$$

At $s = s_0/2$ the above expression is satisfied. Taking derivative with respect to $s$ yields

$$\frac{dF^{-1}}{ds}(s)u'(x_1^* - F^{-1}(s)) + \frac{dF^{-1}}{ds}(1-s)u'(x' - F^{-1}(1-s)) \ .$$

Note that $dF^{-1}/ds > 0$ and $u'(z) < 0$ if $z < 0$ and $u'(z) > 0$ if $z > 0$.

For case (iii) for the proposed coalition $s < s_0/2$, the derivative is negative. Since the condition above is satisfied at $s_0/2$, and the condition is decreasing in $s$, it must be satisfied for all $s < s_0/2$. Therefore everyone in the proposed coalition strictly supports $x_1^*$ over $x'$. For case (iv) the derivative is positive for $s > s_0/2$ and therefore everyone in the coalition supports $x$ over $y$.

By a similar argument, one can show if $y < F^{-1}(1/2)$ then $x_1^* = F^{-1}(1 - s_0/2)$ beats all other $x'$. The proof for $y_1^*$ is also similar. ∎

# Chapter 3

# Seats and Votes

## 3.1 Introduction

The empirical relationship between the shares of votes and the shares of seats that a party wins is a basic tool for evaluating electoral systems.[1] Political scientists estimate the seats-votes curve for specific systems and assess whether there are systematic biases against some parties. In legal cases challenging districting maps, for example, seats-votes curves are often introduced as evidence that a plan creates biases against one group or another. In comparing different sorts of electoral systems or laws, seats-votes curves are used to measure how responsive elections are to changes in public sentiment.

Despite its wide use in evaluating electoral systems, the theoretical foundations of the seats votes curve are not well developed. The seats-vote relationship is an empirical relationship, first described by F.Y. Edgeworth (1898) who noted that the cumulative normal distribution characterized the distribution of votes across constituencies in Britain. See also Kendall and Stuart (1950). This relationship has since been formalized several different ways. One approach (March 1957-1958) is to fit multiple pairs of seats and votes for different elections to a curve, such as a line or a logistic curve. Another approach (Butler and Stokes 1969) is to measure the seats-votes relationship using the cumulative distribution of votes across districts in

---

[1]See King and Browning (1987) for example studies and court cases.

a given year and consider hypothetical alternate vote shares by shifting vote shares equally in all districts. Gelman and King (1994) modify the uniform swing model above to correct for incumbency and allow for different district variances.

The exact interpretation of these empirical relationships is unclear. The goal of the analysis of seats and votes is to study the properties of the electoral system - the rules through which votes are translated into seats. However, some existing methods conflate aspects of voting behavior, such as the incumbency advantage, and aspects of the system, especially the geographic dispersion of partisans or the median voters across districts. As a result, it is unclear what the "partisan bias" and "responsiveness" parameters estimated from such empirical methods mean.

To interpret the seats/votes curve, one requires a theory of voters. In this chapter, I apply a simplified version of the model presented in chapter 2 to districts and voters to generate a theory of seat shares and vote shares. The model is intended to be illustrative of an argument about how to think about vote swings, and as a tool to clarify and inspire new estimation strategies. Like all models, the model is NOT intended to be comprehensive or perfectly realistic. Nevertheless, despite its simplicity, the model is more general then one commonly assumed model: the uniform vote swing hypothesis. Using the model one can conclude that the parameters of the electoral system are indeed not identified; they cannot be untangled from the parameters of voters' preferences and some standard techniques of measuring and interpreting seats/votes relationship are flawed. Finally, we will consider several methods, using additional information than just seats and votes, for identifying the parameters of the electoral system, as distinct from the parameters of the distribution of voters' preferences. Once the distribution of voters' preferences in the nation is estimated, the seats votes curve can be transformed to reveal the underlying electoral system.

## 3.2 Framework and Model

In this section I present a simple voting model and derive implications for the seats/votes relation. The setup is a simplified and modified version (although in some ways more general) of the setup for the model from chapter chapter two.

### 3.2.1 Parties and Voters

Let there be two parties, labeled $X$ and $Y$. There is a one dimensional policy space, with $z \in R$ denoting a typical element of the policy space. Each party represents a platform, $x$ and $y$ respectively. Assume $x < y$. For this chapter, we shall treat $x$ and $y$ as exogenous. ($x$ and $y$ are determined endogenously in the previous chapter.)

Let there be a continuum of voters, with a given voter designated by $i$ and let each voter be defined by an ideal policy $w_i$. Let $v_t$ describe a time-varying net "valence" preference every voter has for party $X$. $v_t$ can be negative, indicating a net preference for party $Y$. [2] Let voter utility for party $X$ be described by $u(|x - w_i|) + v$ with $u(z)$ strictly decreasing, maximized at $u(0)$, and $u$ strictly concave. That is, a voter has a stronger preference for party $X$ if the party's platform is closer to her ideal point and if $v$ is large. Let voter utility for party $Y$ be described by $u(|y - w_i|)$. Assume voters vote for the party that grants them more utility in any given election. (That is, there is no strategic abstention, attempts to manipulate future party platforms, and so forth.) Therefore a voter votes for party $X$ if $u(|x - w_i|) - u(|y - w_i|) + v_t > 0$ and party $Y$ if $u(|x - w_i|) - u(|y - w_i|) + v_t < 0$.

The voter who is indifferent between the two parties defines a "cut-point" $c$ between the parties. For example, if $u(z) = -(z)^2$ then $c = \frac{x+y}{2} + \frac{v_t}{2(y-x)}$. In general, it is easy to show (see chapter two) that the term $u(|x - w_i|) - u(|y - w_i|) + v_t$ is strictly increasing in $w_i$ and thus there is a well defined cut-point $c(x, y, v_t)$ such that all voters with ideal point $w_i > c$ vote for party $Y$, and all voters with ideal point $w_i < c$ vote for party $X$.

---

[2]It is also possible that party platforms and voter preferences can change over time. This possibility will be discussed below.

Note, the purpose of the spatial component of the model is to generate the concept of a cut-point. Many alternate models that also imply there being a single variable which divides voters into supporters of one party or another would work for the analysis that follows. For example, suppose voters have a "partisan intensity" for party $X$, $q_i$. This term could be negative, indicating a tendency to vote for party $Y$. Suppose further, like in the spatial model, there is a universal valence factor, $v_t$, indicating an election-specific tendency for all voters to vote for party $X$. We can then define a voter utility function as $q_i + v_t$. Voters with $q_i$ such that $q_i + v_t > 0$ vote for party $X$; voters with $q_i + v_t < 0$ vote for party $Y$. Therefore the cut-point is $-v_t$. In effect, the ideological voter model can be mapped into a party-loyalty model by equating $q_i = u(|x - w_i|) - u(|y - w_i|)$. As long as party loyalty and party platforms are exogenous the two models are isomorphic. [3]

### 3.2.2 Vote Share and Seats

Given a model of voter behavior that implies a cut-point, $c$, we can now analyze the aggregate seats/votes relationship. Let $F(z)$ be a cumulative density function describing the fraction of voters with ideal points $w_i < z$. $F$ is defined with respect to an electorate. There may be one $F$ for presidential voters, another $F$ for aggregate district votes, separate $F$'s for each district, and so forth. Assume $F$ is continuously differentiable with derivative $f(z)$ and inverse $F^{-1}(\theta)$. Furthermore assume $f(z) > 0$ for all relevant $z$. With this assumption we can ignore the choice of voters with ideal points exactly at the cut-point, since they constitute measure zero. By the definition of $F$, the fraction of voters, $V$, voting for party $X$ is then $F(c)$.

How about seats? For party $X$ to win a majority of seats in any given district, $j$, they must win a majority of votes from that district. To do this, the median voter of the district, $m_j$, must have an ideal point left of the cut-point. Therefore to characterize district behavior we only need to know the preferences of the median voter in each district. [4]

---

[3]See Ward (2000) for an example where voter preferences change with platforms. Such a model would be difficult to merge with the model here.

[4]Note this result relies heavily on the assumption that voter behavior depends only on a one-

Let $H(z)$ be a cumulative density function describing the fraction of districts with district median voters such that $m_j < z$. Where necessary we shall assume $H$ is continuously differentiable with derivative $h(z)$. The seat share party $X$ wins, $S$ is then given by $H(c)$.

Up to this point we have treated party platforms, voter preferences and identities as fixed over time. To model possible changes, one can simply think of $F$ and $H$ changing with time. Furthermore, if voters are held constant in the general population, but district boundaries change, $F$ must remain constant but $H$ might change. In this sense $H$ represents district boundaries. However, it is possible for $H$ to change if district borders remain fixed but voters move from one district to another. One might argue such a change represents a kind of passive, natural redistricting. Furthermore legislatures might choose not to redistrict at all and allow natural redistricting to do gerrymandering for them. It is also possible to redraw borders without changing $H$. Such a change will not change the seats/votes relation, as will be shown below. Therefore, it is unclear whether such a change would have any political impact. For example, in the model described in chapter two electoral outcomes are completely determined by $H$ and so such a neutral gerrymander would have no effect. [5]

## 3.2.3 Seats/Votes Relationship

We now have expressions for the votes for party $X$ and the seats of party $X$ in terms of one variable, $c$. Let the function $G$ be defined as the relationship between seats and votes $S = G(V)$. Because $F$ is invertible, we have the very simple but powerful relation between seats and votes:

$$S = G(V) = H(F^{-1}(V)).$$

---

dimensional variable. That is, there are no district specific effects, such as incumbency. In section 3.4.2 we will consider an extension accounting for incumbency effects.

[5]Such a result depends on the assumption that there is no incumbency effect, or at least incumbency is not influenced by neutral gerrymanders. If incumbents build long-run constituencies that get disrupted by changing borders, even by a neutral gerrymander, then election outcomes will be impacted. See Ansolabehere, Snyder, and Stewart (2000) and Fenno (1978). Such issues are beyond the scope of this model.

The seats votes curve is therefore made up of two parts - the distribution of voters in the general population, $F$, and how voters are distributed across districts, $H$. One can immediately see that the two functions are intertwined. As noted above, $H$ can be thought of as representing the political impact of district boundaries. Therefore, there is no way to separate the impact of districting from changing voter positions or strength of partisan identification from just seats and votes data alone. To put it another way, just knowing $G$ is insufficient to make claims about either $F$ or $H$. In particular, one can not, without further assumptions, make definitive claims about the impact of gerrymandering based on $G$ alone. $G$ might shift after a redistricting, but this may be entirely due to changes in $F$.

Since $F^{-1}$ and $H$ are increasing, the model predicts $G$ is increasing as well. While this may seem natural and obvious, it does represent another testable prediction of the model. If voters were not purely ideologically motivated, one could imagine strange scenarios where aggregate vote share increases but seat share actually decreases.

### 3.2.4 Linear Approximation Example

Consider the first order linear approximation of $G$. The first order Taylor expansion of $G$ around $V^*$ is

$$S_t \approx H(F^{-1}(V^*)) + \frac{h(F^{-1}(V^*))}{f(F^{-1}(V^*))}(V_t - V^*)$$

Since $G$ could be non-linear, we will define the term "responsiveness" at the value $V$ in this model as the first derivative of $G$. Again, since $G$ may be non-linear, there is no single universal responsiveness measure - responsiveness must be defined at a point.[6]

The marginal impact of an increase in vote share on a party's seat share depends on two quantities. First, as is commonly asserted by empirical analysts, the slope

---

[6]In Tufte's (1973) paper, he finds in his samples that the seats/votes curve is, in practice, approximately linear. We will revisit the question of whether the raw seats/vote curve is linear in practice in chapter five. It is worth noting, however, that even if the raw seats/votes curve is linear, $H$ need not be.

depends on the density of district medians. A high density of district medians near $V^*$ means the slope will be higher - the more nearby pivotal districts, the more easily seats change hands. Second, the slope depends on the density of voters near $V^*$. More moderate voters reduces the vote-swing. If there are many moderate voters, a small change in the underlying cut-point causes a large number of voters to shift. So for any given seat swing, the number of voters changing parties is large, and so the number of seats changing relative number of voters changing is small.

For example, consider adding infinitely extreme partisans to every district in equal proportions. Doing this has no effect on party chances of victory, no effect on expected vote shares, no effect on district medians, and could not reasonably be described as a gerrymander. However, it will make the swing ratio look larger, because now the total FRACTION of moderates has decreased.

This point has not been well-understood in a widespread way in the literature. Researchers repeatedly act as if responsiveness is entirely a description of the district distribution, i.e. $H$. For example, Tufte (1973) discusses the normative value of a high responsiveness versus a low responsiveness. He claims that low responsiveness indicates that incumbents are trying to create more safe seats. This model calls such an exercise in question, since responsiveness is to some degree arbitrary, unless $F$ is known to be uniform and constant. Tufte also notes that responsiveness is higher during presidential years. He suspects this might be because in those years elections are more nationally focussed. The model suggests an alternate possibility: that $F$ may be different in presidential years. Jacobson (1987) studies how responsiveness has changed with time, and claims that since responsiveness has not significantly changed, there must be no wide-spread changes in the health of the system. Again, responsiveness is the wrong variable of interest. $H$ is the relevant term.

The above relation can also be used to reinterpret Edgeworth's (1898) central limit theorem type argument for why responsiveness tends to be greater than one when the cut-point is near the median district. Suppose that districts were generated by random draws from the general population. The distribution of districts, $H$, would then approach a normal distribution centered at the median of $F$. For a

large population, the variance of $H$ would shrink. Therefore, the density of $h$ near the median would become large, generating a large responsiveness. As noted by Edgeworth, in practice, actual districts are probably not formed by true random draws of single individuals from the general population, as such a large number of persons would result in tremendous responsiveness values.

### 3.2.5   Logistic Example

As another example of the lack of identification, consider the logistic functional form.

The Cube Law states that the odds that a party wins a seat are approximately equal to the cube of the odds of winning a seat. March (1957-1958) formalized the Cube Law using the logistic function (see also Tufte (1973) and King and Browning (1987)).

$$\frac{S}{1-S} = A \left( \frac{V}{1-V} \right)^{\gamma}$$

Taking logarithms of both sides reduces to a linear formula commonly used in empirical study of districted-system:

$$ln \left( \frac{S}{1-S} \right) = \gamma_0 + \gamma_1 ln \left( \frac{V}{1-V} \right).$$

where $\gamma_0 = ln(A)$. It is also the case that the logistic function is widely used to study the distribution of voter preferences and vote probability functions.

In this example, we assume that the logistic distribution characterizes both $H$ and $F$. Specifically, the log of the odds of winning a seat and the log of the odds of winning a vote can be written as functions of the cut point $c$ as follows:

$$ln \left( \frac{S}{1-S} \right) = \alpha_0 + \alpha_1 c$$

$$ln \left( \frac{V}{1-V} \right) = \beta_0 + \beta_1 c.$$

With these assumptions we may now express the log of the odds of winning a seat

in terms of the log of the odds of winning a vote, as in March (1957).

$$ln\left(\frac{S}{1-S}\right) = \left(\alpha_0 - \frac{\alpha_1\beta_0}{\beta_1}\right) + \left(\frac{\alpha_1}{\beta_1}\right)ln\left(\frac{V}{1-V}\right).$$

Hence, under the logistic functional form, the commonly used method for estimating the relationship between seats and votes arises immediately from the spatial model. Like the linear approximation model, we see the seats/votes relationship is a mixture of coefficients from both $H$ and $F$ - we can not identify the coefficients separately. This observation also implies that the Cube Law, $\frac{\alpha_1}{\beta_1} = 3$, is not directly informative about the shape of districts.

## 3.3  Interpreting Bias

One of the basic properties students of the seats/votes curve check for is "unbiasedness." The model presented above can help us define and reinterpret this property of electoral systems. [7]

### 3.3.1  Median Unbiased

One natural definition of an unbiased electoral system is that whenever $V = 0.5$, then $S = 0.5$. We will call this median unbiased. Our model gives a succinct characterization of median unbiasedness. A system is median unbiased if and only if the median voter in the general population, $w_i$ such that $F(w_i) = 0.5$, has the same preferences as the median voter of the median district $H(w_i) = 0.5$. If one knows $F$ and $H$ (from, for example, survey data) one can directly tell if a system will be unbiased without looking at seats/votes curves. This property of the model can be used as a test of the model, or as a specification check, when using any empirical seats/votes curve that

---

[7]In this section I define conditions for an absolute absence of bias, but I do not attempt to define bias for positive bias. For example, Tufte (1973) distinguishes between the seat share at $V = 0.5$ and the vote share such that $S = 0.5$ as two different metrics of bias. The lack of identification problem makes choosing between these two different measures even more difficult. It may be possible to utilize these metrics as a measure of degree of misrepresentation. Such an approach is beyond the scope of this paper.

reports to show zero bias. [8]

The definition and model also suggests two natural unbiased systems. One possibility is to make $H = F$. For every type of voter, left, right and center, make a corresponding district. [9] Another possibility (albeit stretching the standard assumptions of the model) is to create one giant district, or make all districts have the same median voter equal to the median voter in the general population. As pointed out by Tufte (1973), the former system has a small responsiveness (equal to one everywhere) while the latter system produces an infinite responsiveness at $V = 0.5$ and zero responsiveness elsewhere. But both systems are unbiased.

### 3.3.2  Global Unbiased

Define global unbiased as a system such that $G(0.5 + x) + G(0.5 - x) = 1$ for all $x \in [0, 0.5]$. [10] For example, suppose whenever party $X$ receives 60% of the vote, they receive 70% of the seats. Arguably a perfectly "fair" system should require whenever party $Y$ receives 60% of the vote, they would also receive 70% of the seats. Median unbiased is a special case of global unbiased.

In our model, a sufficient condition for global unbiasedness is both $f$ and $h$ to be symmetric around $V = 0.5$. However, if only one or the other of $f$ and $h$ is symmetric, but not both, then the system will not be global unbiased. Furthermore, for generic $F$ and $H$, the system will not be global unbiased. Essentially global unbiasedness requires strong symmetry conditions. These conditions are stronger then one might realize without the model, because even if, say, districts are symmetric, asymmetric voter preferences will yield bias. In particular, one can not conclude that districts have a strange, asymmetric and therefore perhaps gerrymandered shape merely because the seats/vote relationship is biased or has a strange shape. That is one can not infer

---

[8]Note unbiasedness is not a property of $F$ alone or $H$ alone - it is a property of the interaction between the two. Informally speaking, unbiasedness tries to characterize how well districts and voters mesh together. In particular, statements such as "the distribution of voters created bias" or "the distribution of district medians created bias" may be a bit misleading. It is the interaction that matters.

[9]This construction also satisfies the very strong alternate definition of unbiasedness requiring $S = V$ for all $V$. See for example Gudgin and Taylor (1974).

[10]If $G$ has a derivative, it would be symmetric around $v = 0.5$.

evidence of a gerrymander solely from observation of seats/vote bias.

Another way to see the importance of global symmetry is to consider the second-order Taylor-series expansion of $G$. The second term is:

$$(h'(c^*) - f'(c^*)) * \frac{h(c^*)}{f(c^*)} * (\frac{v - v_0}{f(c^*)})^2$$

where $c^* = F^{-1}(V^*)$ is the cut-point implied by the value of $V$ one is expanding about. This term will typically be non-zero, unless one is expanding around a point where both $h$ and $f$ have extrema. This would happen if one were expanding around $V = 0.5$ and $h$ and $f$ were symmetric and sufficiently smooth, but will not happen in general.

Another sufficient condition for global unbiasedness is median unbiasedness plus constant responsiveness. Or in other words, the seats/votes curve is a line passing through $V = 0.5$, $S = 0.5$. If one assumes seats/votes curve is always linear, one might lose sight of the fact that there is a distinction between median unbiasedness and global unbiasedness. Hence while some of the literature assumes there is one natural definition of unbiasedness, our model makes it clear that there are at least several possible definitions. [11]

### 3.3.3 Skewed Electorate.

Picking which definition of bias to use is a normative question, but there is a sense in which neither of the above definitions are appropriate. That is, bias by the above definitions is not a good measure of party competitiveness. Consider the following example.

Assume that valence shocks are symmetrically distributed around zero. Assume further that the density of district ideal points along the dimension $X$ is uniform from 0 to 1. Hence, the median district is at 1/2. This is a competitively fair districting

---

[11]The two definitions of unbiasedness are not meant to be exclusive. For example, recall in the model that vote share and seat share implicitly depend on valence shock, and therefore are themselves random variables. One could define the seats/votes relationship as "ex-ante unbiased" if the expected value of $S$ equals the expected value of $V$. If one believes that a system is fair as long as each party gets, on average, a fair number of seats, this definition may be normatively appealing. If valence shocks are drawn from a symmetric distribution around zero, then global unbiased is a sufficient condition for ex-ante unbiased. But if valence shocks are not symmetrically distributed (even if they have mean zero) then global unbiased would not imply ex-ante unbiased.

distribution because if the two parties are equidistant from that point, say, at 1/4 and 3/4, then each party has an equal chance of winning a majority. The model in chapter two of this dissertation implies such a distribution of platforms would be natural for such an $H$.

However, assume also that the density of voters' ideal points is skewed. For simplicity assume that the density of ideal points is triangular with highest density at $X = 0$, i.e., $f(x) = 2(1 - x)$, $F(x) = (1 - x)^2$, $F^{-1}(V) = 1 - \sqrt{1 - V}$

Using the results above, the relationship between seats and votes becomes

$$S = 1 - \sqrt{1 - V}$$

When the vote share is evenly divided between the two parties, the predicted seat share heavily favors one of the parties. That is $S = 1 - \sqrt{1 - .5} \approx .29$.

This example reveals a fundamental problem with seats-votes analyses commonly done in the English and Australian contexts (e.g., Butler and Stokes 1969) in which researchers use the distribution of votes and consider uniform shifts in the distribution until the division of the seats is equal. The example reveals that such an hypothetical calculation may not reflect the distribution of district medians but the distribution of voter ideal points. Indeed, the parties likely would never compete for seats in the way supposed by such hypothetical calculations. What the calculation above measures is just this: in order to have electoral competition in which the left party wins 50 percent of the votes, they would have to adopt a platform in which they would win only 29 percent of the seats.

Observing bias might very well mean parties are not representing the median or average voter in the general population. But bias does not imply parties are non-competitive. If one believes that gerrymandering is associated with non-competitiveness, searching for bias as evidence for gerrymandering is beside the point. Biased systems can be competitive (as shown in this example) and examples can be made with unbiased systems that are non-competitive. Interpreting bias as a measure of "an unfair partisan differential in the ability to win seats" (King And Browning 1987) may be

66

inappropriate.

That said, there is an argument that bias is evidence for something being artificial with a system. Suppose districts were generated by random independent samples from the population of voters. The distribution of district medians would then approach a normal distribution, centered around the population median. That is, the system should be median unbiased. A biased system would then indicate the possibility of gerrymandered districts. However, a biased system might also simply mean districts are not made up of random individuals, but rather have demographic, historical, cultural and other traits that make each district unique and shape the preferences of its residents. Again, since bias does not imply non-competitive, one can not assume malign intent simply from observing bias.

## 3.4   Estimation Methods

The seats votes relationship, defined above as $G$, is what is oftentimes measured by scholars, but is not necessarily the true object of interest. Typically what one really wants to know is $H$, the distribution of district medians. Changes in $H$ are indicative of possible gerrymanders, or changes in the constituencies incumbents need to face. $H$ may also have a significant impact on party platforms, as shown in chapter two. If the goal of parties is to win seats, then competing for the median district may be more important than competing for the median voter in the aggregate population.

However, as shown in 3.2 above, $H$ is not independently identified from $F$ using just seats and votes data. To find $H$ we must either estimate or make assumptions on $F$. Even if one is willing to make assumptions on $F$, trying to estimate $F$ is still useful as a way to test those assumptions.

In section 3.4.1 I describe how to find $F$, assuming no incumbency effect. In section 3.4.2 I discuss how to control for incumbency. In section 3.4.3 I discuss how to find $H$ using regression techniques. In section 3.4.4 I discuss the uniform swing method, and an alternative that takes advantage of having detailed knowledge of district $F_i$.

### 3.4.1 Finding $F$

How does one find $F$? One needs a measure of how ideology is distributed in the electorate. Possible approaches include:

1. Survey or exit poll data with direct questions on ideology or partisanship.

2. Survey or exit poll data with a battery of questions on various issues. Use the questions to construct a measure of ideology.

3. Find the distribution of demographics from one data source (such as the census). Use survey data to find a mapping from demographics to ideology and/or voting behavior. Combine the two sources to find distribution of ideology.

Technique number 2 is described and implemented in the next two chapters. Technique number 1 is also explored in the next chapter, but is shown to have complications. Technique number 3 is left to future work. However, it is worth noting that Gelman et al. (2005) have shown that the relationship between income and voting behavior might not be consistent from region to region and state to state.

Some survey data will be too grainy to estimate $F$ smoothly. Even with grainy data we can still approximate another item of interest: the density $f$. For example, we might only know voter ideologies on a three point scale (liberal, moderate, conservative.) For this case we can still construct the measure: set $f$ equal to the number of moderates divided by total number of voters. While implementing techniques based on knowing only $f$ is left to future work, in the next chapter several possible proxies for $f$ are considered (self identified moderates and political independents) and checked to see if they are good proxies for the number of centrists. Self-identified party independents with no lean toward either party turn out to be the closest proxy to centrists. Interestingly, self-identified moderates are problematic as proxies for centrists, because they are distinctively left-of-center.

### 3.4.2 Incumbency Correction for $F$

A simple solution to the incumbency effect is to use presidential vote instead of district vote. Recall as noted in the definition, $F$ refers to a particular electorate. If the survey

sample was explicitly presidential voters, then presidential vote share is particularly appropriate. A drawback, however, is that we need to construct an artificial seat share measure if we wish to compare the analysis to standard seats/votes techniques. One possibility is to assume that every district with more than 50% support for party $X$'s presidential candidate supports party $X$. However, this assumes a very strong tie between presidential support and district support.

Alternatively, we could use district vote data and subtract off the incumbency effect.[12] The model presented in section two assumes that voters are purely ideological. Naturally this assumption is a bit strong, and one may wish to weaken it by considering district specific effects. Many empirical models treat incumbency as a direct boost to vote share. A natural extension to the current model would then be to think of each observed district vote share, $\hat{F_i}(c)$, be made up of an ideological component $F_i(c)$ and an incumbency term $\delta * I_i$, where $\delta$ is the incumbency effect, and $I_i$ is an indicator variable set to zero if the district is an open seat, one if the district is held by party $X$, and negative one if held by party $Y$. Assume the terms are related by $\hat{F_i}(c) = F_i(c) + \delta * I_i$. Let $n_i$ be the number of voters voting in a district. Then the aggregate observed vote share, $\hat{F}(c) = \frac{\sum n_i * F_i(c)}{\sum n_i} + \frac{\sum n_i * I}{\sum n_i}$. The first term is just $F(c)$, the aggregate vote share assuming purely ideological voters, or equivalently, the hypothetical vote share if all seats were open. We can then find the hypothetical aggregate vote share from the observed aggregate vote share by subtracting off the incumbency correction: $F(c) = \hat{F}(c) - \frac{\sum n_i * I}{\sum n_i}$. We need to estimate the incumbency effect separately using standard techniques, but (mercifully) we do not need to know the actual $F_i$'s. To find the hypothetical seat shares to trace out hypothetical seat shares we can simply subtract off the incumbency effect for each district. This gives a hypothetical seats and votes shares assuming every district were open.

Another possibility is to explicitly estimate the relationship between presidential vote share and district vote shares to approximate hypothetical district vote shares. Suppose the observed vote share in a district is related to presidential vote share,

---

[12]This method could be extended to include other district specific effects, such as challenger quality. This approach is left to future work.

$F_i^P(c)$: $\hat{F_i}(c) = \alpha + \beta F_i^P(c) + \delta * I$. Then from the equation above, $\alpha + \beta F_i^P(c) = F_i(c)$ or after aggregating, $\alpha + \beta F^P(c) = F(c)$. This technique requires us to find or assume $\alpha$ and $\beta$. These values can be estimated by regressing observed vote shares on presidential vote and incumbency. Of course this method also finds the incumbency effect, allowing us to perform the technique given in the previous paragraph as well as a check. This method may also be useful for answering certain counterfactuals, such as "If presidential vote share went up by one percent, how many seats would change hands" while the previous method is better suited for questions such as "If all seats were open seats, what fraction of seats would party $X$ win?"

The above two methods maintain the standard assumption that incumbency acts directly on vote share. However, what if incumbency acts on voter utility like a bonus valence shock in favor of incumbents? Using the notation given above, suppose $\hat{F_i}(c) = F_i(c + \delta * I_i)$. That is, incumbency moves the relevant cut-point within a district. The first order Taylor series approximation gives us $\hat{F_i}(c) \approx F_i(c) + f_i(c) * \delta * I_i$ where $f_i(c)$ is the density of voters in a district with ideal points near $c$. Here we see that incumbency effect in a district must be weighted by the number of moderates in a district. In retrospect this is obvious. If a district is made up of extreme partisans, incumbency will not move very many votes. However, if a district is made up primarily of moderates, a small voter utility boost will cause a large change in votes. This observation might explain the mystery of the increased incumbency effect observed in the 60s. If the fraction of strong partisans declines, the incumbency effect measured in terms of votes should increase, even if the effect measured in terms of voter utility stays constant.

### 3.4.3   Regression to Find $H$

The regression method posits a relationship between seats and votes over time $S_t = H(F^{-1}(V_t)) + \epsilon_t$ with $\epsilon_t$ independent of $V_t$. One can then either assume a particular functional form for $H$, such as assume $H$ is linear or logistic, estimate a $H$ non-parametrically using a kernel density, or approximate $H$ with a polynomial expansion. The party position model of chapter 2 informs us that non-symmetry in

$H$ has important electoral consequences. Recall a skewed $H$ can cause one party to choose a platform closer to the median district relative to the other party. This in turn can lead to a permanent advantage for that party, even with ideological voters and flexible party platforms. In other words, unlike responsiveness and bias of the raw seats/votes curve, a skewed $H$ is actual evidence for a gerrymander that undermines competition. Therefore testing, rather than assuming, whether $\beta_3 = 0$ is very important. [13] Therefore, regression models that allow for non-symmetrical $H$ should be favored. A 3rd-order polynomial approximation around $V = 0.5$, $(S_t - 0.5) = \alpha + \beta_1 * (F^{-1}(V_t) - F^{-1}(0.5)) + \beta_2 * F^{-1}(V_t - F^{-1}(0.5)))^2 + \beta_3 * F^{-1}(V_t - F^{-1}(0.5)))^3$ would arguably be a better choice than a linear or simple logistic model. Like the linear model, the parameters can be estimated using ordinary least squares and $\alpha$ and $\beta_1$ can be interpreted as a measure of bias and the marginal density of districts near $F^{-1}(V = 0.5)$.

Naturally, the above method requires a detailed measure of $F$. If a detailed measure is not available, there are two other possible approaches. The first is to simply assume $F$ is uniform and constant. Then the raw seats/votes curve, $G$, is $H$ (up to a scale factor) and the polynomial regression can be interpreted as a direct measure of district shapes. Implicitly this is what most other scholars who use seats/votes regressions over time are doing - they find $G$ and assume it is informative about $H$. In this case, responsiveness can be interpreted in terms of the relative density of districts to the (fixed) relative density of voters.

Another possibility is to assume $F$ may have an arbitrary shape, but is constant over time. Studies that examine seats/votes parameters over time and make claims about changes in electoral system (e.g. Jacobson 1987) must either implicitly be making this assumption, or else risk being flawed. With this assumption, any changes to the raw seats/vote curve $G$ over time must be due to changes in $H$. In this case responsiveness is difficult to interpret. One could have constant responsiveness over $V$, but the density of districts could be changing, as long as $h/f$ is constant. However, we

---

[13]Brady (1988) also examines skewness, but not for the reasons given here, and he studies skewness of the raw seats/votes curve, derived through the uniform-swing hypothesis, and not through $H$.

can still see if $H$ is changing over time. We could regress $S$ on $V$, $t$, the interaction $V * t$, and higher order terms. If the coefficients on the interactions are non-zero, then combined with the assumption of constant $F$, we know $H$ must be changing. Another possibility is to identify some special time, such as *Baker* v. *Carr*, such that one expects districting to change significantly from before to after. Let the dummy variable $d_t = 1$ for all $t$ after the special event, and $d_t = 0$ otherwise. Then regress $S$ on $V - 0.5$, $D$ for change in bias, the interaction $(V - 0.5) * D$ for change in marginal vote swing, plus as many higher order terms as desired.

Finally, if we have $f(F^{-1}(V^*))$ at a given point but not $F$, one can still estimate the density of moderate districts, $h(F^{-1}(V^*))$ by regressing seats on $V_t/f(F^{-1}(V^*))$. For example, if using fraction of moderates as a proxy for $f$, instead of regressing $S$ on $V$, regress $S$ on $V/f$. This is justified by the linear approximation of $G$ given in section 2.4 above. $h(F^{-1}(V^*))$ can then be used to approximate $H$ near $V^*$ through the first order Taylor series expansion $H(F^{-1}(V^* + \Delta V)) \approx H(F^{-1}(V^*)) + h(F^{-1}(V^*)) * \Delta V$

### 3.4.4 Uniform and Non-uniform Swing to Find $H$

The regression method has several drawbacks. First, it requires a series of elections and a reasonably stable electoral system. Second, it throws away a fair amount of information, relying only on aggregate vote share and seat share. The uniform vote-swing method makes some additional assumptions on districts to allow one to estimate a seats/votes relationship for each year using data from that year only. The standard uniform vote-swing method simply takes each vote share of each district, and assumes that for each unit increase $\delta$ in aggregate vote share, $v$, each district's vote share, $v_i$, increases by $\delta$ as well. This in turn increases the seat share as district vote shares cross 0.5.

What does the uniform-swing assumption look like in the context of our model? Let $F_i(c)$ describe the vote share in district $i$. The uniform vote swing model then implies a relationship between $F$ and $F_i c$. For each district there exists a constant initial value $a_i$ such that $F_i(c) = F(c) + a_i$. $H(c)$ can be found by finding the number of districts with $F_i(c) > 0.5$ for any value of $c$.

Obviously the uniform swing method requires very strong restrictions on the behavior of each district. We could assume the restriction holds and then test it with outside data. One possibility is to assume both the function $F$ and parameters $a_i$ are constant over time and test whether the $a_i$'s are indeed constant over multiple elections. Another possibility is to measure $F$ and $F_i$ directly and see if they correspond to the formula above. This test would work even with rough measures of $F$ and $F_i$.

However, if one has direct measures of $F_i$, one can improve upon the uniform swing model. If one knows $F_i$, one can directly find the seats/votes curve. For any given cut-point, the vote share is simply $F(c)$, and the seat share is the fraction of districts with $F_i(c) > 0.5$. This technique is implemented in chapter 5 for the US 2006 congressional elections. See chapter 5 for more discussion.

If one only has measures of $f_i$ (for example, fraction of moderates in every district) one can modify the uniform swing method with weights. Districts with large numbers of moderates will have a larger vote swing than districts with small numbers of moderates. Instead of moving every district by 1 unit when total vote share moves 1 unit, weight each district by the number of moderates. This technique is very closely related to Gelman and King (1994)'s technique. Gelman and King weight districts by their historical variance; I propose weighting districts by the number of moderates.

These techniques also suggest a different manner of handling incumbency, if one assumes incumbency acts on cut-points instead of votes (see section 4.2 above.) Divide districts into open seats, seats with party $X$ incumbents, and seats with party $Y$ incumbents. These groups define three different group aggregate vote shares, $F_O$, $F_X$, and $F_Y$. Each of these subgroups define an initial cut-point $F_O^{-1}(v_O)$, $F_X^{-1}(v_X)$, $F_Y^{-1}(v_Y)$. With these three cut-points one can determine which districts will vote for which party using $F_i(c)$. To consider alternate vote shares, one can alter the cut-points (holding the differences between the group cut-points constant).

## 3.5 Conclusion

This chapter has shown that there is a fundamental identification problem with using the raw relationship between seats and votes to learn about the distribution of districts. Seats/votes bias does not imply one party has an electoral advantage over the other. It merely implies there is a mismatch between the preferences of the median voter of the median district and the median voter. Seats/votes responsiveness is even less useful as an indicator of the district distribution, since it can be arbitrarily manipulated simply by changing the number of moderate voters.

This does not mean the seats/votes relation is useless. If one is willing to assume the distribution of voters in the population is uniform, or at least constant, changes in the seats/votes curve will imply changes in the distribution of district medians.

However, in order to use seats and votes data to find the full distribution of seats for arbitrary voter distributions, one needs to break the identification problem. By using polling or demographic data, one can transform vote data. One can then relate the transformed vote data to seats to find the distribution of districts.

There is a sense in which accounting for voter ideologies is nothing more than returning to the original reasoning of why some, particularly the courts, were drawn to studying the seats/votes curve in the first place. That is, what is really at stake is voter interests, and studying whether a particular legal framework for apportionment (or, by extension, districting) "did not substantially deprive some element of the citizenry of their ability to compel through the electoral process governmental responsiveness to their interests." (*The Yale Law Journal* 1963)

Chapter two of this dissertation provided a model implying the critical importance of the distribution of districts. For example, the distribution of districts determines party platforms and an asymmetric distribution of districts can lead to uncompetitive elections. This chapter has provided a list of methods of finding that distribution. The following two chapters will implement one technique: direct measurement of ideology (chapter 4), which will then be used to construct seats/votes curves and find the distribution of districts (chapter 5).

# Chapter 4

# Measuring Ideology

## 4.1 Introduction

The goal of this chapter is to describe the distribution of voter ideology throughout the country. The measures developed and described here will be used in the next chapter to discuss implications of the party position paper from chapter two, and finding seats/votes curves discussed in chapter three.

The chapter is organized as follows. The next section, 4.2, describes the primary ideology measure, principal factor rank scores. Section 4.3 describes the dataset, the common content of the Cooperative Congressional Election Study.

The bulk of this chapter presents information about the distribution of ideologies for sub-populations of interest. The subpopulations described here serve two purposes. First, they can help check robustness of the measure. Second, they relate to the party position paper discussed in chapter two or the seats/votes model discussed in chapter three. The sub-populations we will examine are self-identified moderates (versus liberals and conservatives) in sections 4.4.1 and 4.4.2, self-identified independents (versus partisans) in section 4.4.3, and states (i.e. Senate districts) in section 4.4.4. Some of the methods in chapter 3 rely on having a good proxy for the number of centrists. It turns out that some measures (non-party-leaning independents) may be better than others (self identified moderates). Furthermore, the measured distributions should (and generally do) match up with intuitive expectations. Self

identified liberals should be measured as liberal. Democrats should be more liberal than Republicans. Massachusetts should be measured as having more liberals than Utah.

The last few sections discuss the robustness of the measure. Section 4.5 discusses a drawback of linear principal factor analysis - the linearity. Section 4.6 considers an alternative non-parametric, non-linear measure and compares it to the linear factor analysis. It turns out the linearity assumption is not necessary for the results described in the rest of the chapter. The concluding section, 4.7, will highlight particularly interesting results from the chapter.

## 4.2   Principal Factor Rank Scores

The primary measure discussed here is rank-ordered principal factor scores. I generated these scores with the following procedure:

1. Find 1st-dimension principal factor regression coefficients and scores for the subset of individuals who answer every chosen policy question.

2. Using the factor regression coefficients above, impute the scores of persons who only answered some questions.

3. Find the score ranks; that is, for each person, their new score is equal to the fraction of people more liberal than they are. Ties are permitted.

The imputation method is as follows. Let answer for person $i$, question $j$ be $y_{ij}$. Then $y_{ij} = b_j * x_i + e_{ij}$ with $b_j$ a question specific regression coefficient, $x_i$ a person specific score, and $e_{ij}$ be the residual. Estimate $b_j$ using principal factor analysis on those subjects that answered every question. Then for all individuals estimate $x_i = \sum \frac{b_j * y_{ij}}{b_j^2}$, using only available questions. By construction estimated scores for individuals that answer all questions will be the same as the original principal factor scores.

Why use principal factor rank scores and not the raw scores themselves? First, there are reasons to be skeptical about the assumptions underlying the linear factor analysis model. These reasons will be discussed in section 4.5 below. In deference

to these concerns, this paper does not attempt to interpret any cardinal information from the scores. For example, if comparing scores to a baseline of zero, a score of two is not "twice" as far from zero as a score of one. Instead this paper will concentrate solely on ordinal interpretations. All we will hope for is a score of two is more than a score of one which is more than a score of zero.

Another reason to disregard cardinal information from factor scores is that the cardinal information they might provide is arguably uninteresting for the purposes of this project. That is, under certain (dubious and perhaps violated, see section 4.5) assumptions, linear factor analysis scores could be interpreted as describing the probability a subject would give the conservative answer to questions. However, this project is not directly interested in how subjects answer questions. The project is focused on vote choice. A more useful cardinal measure of ideology would describe how a voter would choose between the two parties. Such an approach is considered in the next chapter in section 5.2.

The other benefit of ranked scores is interpretation of subpopulation distributions. When one uses a ranked measure, one fixes the distribution of voters to be uniform with respect to the measure. For this case, I calculated the ideological ranks with respect to the national sample (country as a whole). Therefore, voter ideologies will be, by construction, uniform at the national level. The following graph shows the distribution of estimated scores, along with bootstrapped confidence intervals, for the entire sample.

Since the scores were constructed to be uniform at density one, they should be (and are) close to uniform and density of one. Future distribution graphs depicting subpopulations will be of a similar form, with a solid line indicating the median bootstrapped density estimate, and the dotted lines indicating the 25th and 975th (out of a thousand iterations) highest and lowest density estimates at each rank score. That is, the band between the dotted lines is an estimate of the 95% confidence interval of the density function at each rank score.[1]

---

[1]Footnote: because the bootstrap method used holds the scores as fixed, the confidence intervals presented are probably too small. Ideally one would recalculate scores for each iteration. Scores were held fixed in order to reduce computation time. Furthermore, as usual, the confidence intervals

**Whole Sample**
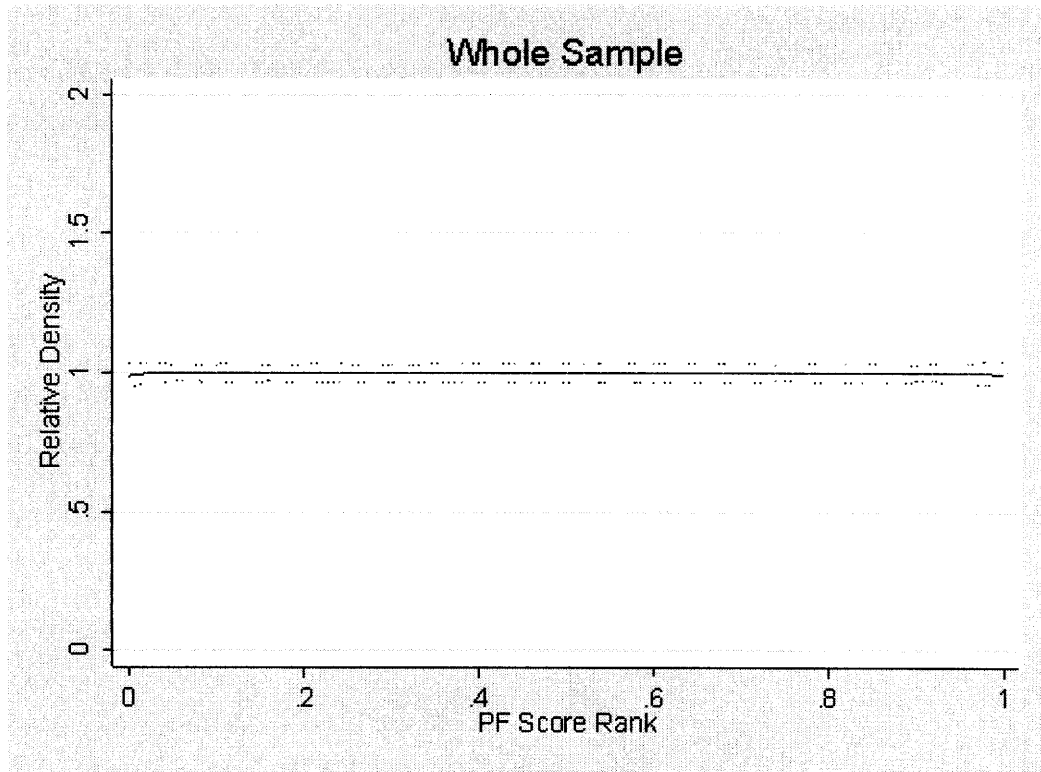
Subsamples need not be uniform with respect to the general population. This allows for straightforward interpretations of densities. If a density is less than one at a point, then the fraction of people within the subpopulation at that ideology is smaller than the fraction of people in the general population at that ideology.

## 4.3 Data

The primary data set examined here is the common content of the Cooperative Congressional Election Study. 36051 unique subjects were available. This large sample size gives traction to examine the full distribution of subpopulations.

Questions used were 14-18 and 28-33 on the pre-election survey and 16 and 17 of the post election survey. The topics these questions deal with include increasing the minimum wage, ending the Iraq war, amnesty for immigrants, stem cell research,

---

assume that the one-dimensional ideology model is a good approximation of subjects' responses to questions. Adding in "model uncertainty" would further expand the confidence region.

banning "partial birth" abortions, environment, social security, capital gains tax, union influence, tax rates (two questions), and use of military (for defense of oil supply, spread democracy, destruction of terrorist camps, defense of allies, and help the UN uphold international law)

Questions from a preliminary survey on minimum wage and union influence were also used.[2] There were therefore two differently worded and differently timed questions on minimum wage. The Pearson correlation between the minimum wage questions is 0.885; the tetrachoric correlation is 0.988.

A question on the Central American Free Trade Agreement (CAFTA) was considered for inclusion in the analysis, but was dropped. The factor loading on the first dimension was only 0.10. The next smallest loading is military use for allies at 0.26. The rank correlation between CAFTA responses and principal factor scores is 0.11. The next lowest rank correlation is 0.28 for military use of allies. (Although, of course, the scores are based in part on military use for allies responses. By construction they should be positively correlated). The Loevinger H of CAFTA combined with the other variables used in non-parametric Mokken scores (see section 4.6 below) is a tiny 0.07. A value of 0 would mean complete independence. The suggested absolute minimum value for inclusion in analysis is 0.3. Essentially CAFTA support or opposition seems to be nearly completely independent of ideology, as measured by responses to other questions. In addition, CAFTA also has the wrong sign correlation with the immigration question. If all questions were driven by the same underlying single dimension, then all questions should be (up to multiplying some questions by negative one) positively correlated with all other questions.

### 4.3.1 Self Placed Ideology, 5 point scale

The first set of sub-populations we will examine is self-reported ideology. If rank scores derived from question responses are truly measuring ideology, and if self-reported

---

[2]Minimum wage wording was "As you may know, the federal minimum wage is currently $5.15 an hour. Do you favor or oppose raising the minimum wage to $7.25 an hour over the next two years, or not?" Union influence wording was "How much influence would you like unions in the United States to have?" with possible responses of more, same, or less.

ideology also truly measures ideology, then the two measures should be strongly correlated. In fact, within the survey they are. The spearman correlation of the rank principal factor scores and the self-reported 5-point scale is 0.73. As can be seen in the following graphs, self-identified liberals tend to answer policy questions one way; self-identified conservatives the other.

### Self Identified Ideology
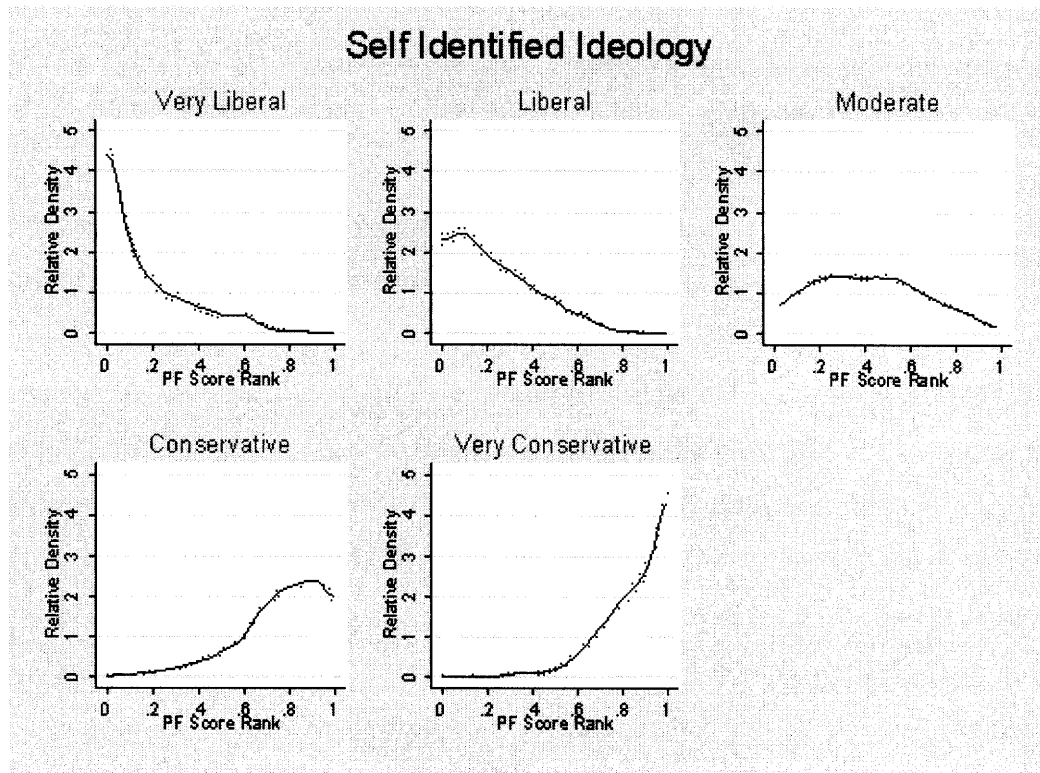


Self identified moderates are worth a closer examination. As is immediately apparent, they are not "centrists" per se, but left of center in terms of relative position on issues. There are people who hold the same policy positions as self-identified liberals, but choose to call themselves moderates. There are not many people who hold conservative positions but call themselves moderates.

Part of this can be explained by the fact that more people identify as conservative rather than liberal:

| Very Liberal | Liberal | Moderate | Conservative | Very Conservative |
|---|---|---|---|---|
| 7.0% | 17.6% | 39.1% | 26.1% | 10.3% |

Suppose that moderates were distributed between liberals and conservatives. Their median ideological rank would therefore be one half of (7.0+17.6) + (1-10.3-26.1). That is, rank 0.441. If there were a perfect correspondence between rank score and self-identified category, moderates would have rank scores between 0.246 and 0.637.

In truth, the median moderate has rank 0.412, (bootstrapped, 1000 iteration, 95% confidence interval: .4054, .4193). There are positive numbers of self-identified moderates with rank scores ranging all the way from 0.000087 (5th most liberal scoring person) to 0.999898 (3rd most conservative person). However, the distribution still leans left. The first quartile score of self identified moderates is 0.234, and the third is 0.595.

Even if the left-of center nature of self-identified moderates were entirely due to a shortage of self-identified liberals and a surplus of self-identified conservatives, the number of self-identified moderates might not be a good measure of the number of centrists (and hence swing voters). Recall in chapter 3 that one of the techniques to find the distribution of district ideologies is to weight by the number of ideologically marginal voters. One might be tempted to use fraction self-identifying as moderates as a proxy for number of marginal voters. However, this might be inappropriate. The median voter might self-identify as moderate, but the median moderate is actually quite solidly left of center relative to the whole country.

Just as self-identified moderates tend to give liberal answers, self-identified liberals tend to give moderate answers to questions, relative to self-identified conservatives. This can be seen in the density graphs above. The exact density estimates (with lower and upper 95% confidence bounds) for each subpopulation at principal factor rank score of 0.5 are as follows:

| Self Reported Ideology | Density at Rank 0.5 | Lower Bound | Upper Bound |
|---|---|---|---|
| Very Liberal | 0.457 | 0.372 | 0.540 |
| Liberal | 0.815 | 0.737 | 0.901 |
| Moderate | 1.384 | 1.317 | 1.447 |
| Conservative | 0.620 | 0.558 | 0.683 |
| Very Conservative | 0.193 | 0.142 | 0.247 |

Conservatives are more pure in their response patterns; to first order, one self-reports as conservative if and only if one answers the policy questions conservatively. Self-identified moderates and liberals however blur into each other in terms of question responses.

To some extent the lack of centrist, self-identified conservatives is simply a different manifestation of the result of a lack of right-of-center self-identified moderates. If self-identified liberals have moderate to left of center positions, and self-identified moderates have moderate to left of center positions, then the remaining group, self-identified conservatives, must contain those who hold right-of-center positions.

## 4.3.2 Self Placed Ideology, 101 point scale

Subjects were asked to rate themselves on a 0 to 100 scale, 0 meaning most liberal, 100 most conservative. This score is also strongly correlated with principal factor rank scores. The spearman rank correlation coefficient with principal factor scores is 0.78; the spearman rank correlation between 101 point scale and the 5 point scale is 0.81. The data is consistent with all three variables measuring the same underlying trait.

One might ask, why not just use self-reported scores instead of the rank principal factor scores? One reason is there are a few oddities with people's self-reported scores. The fraction of people who self identify at 0, 50, and 100, and people who identify less than 50 and greater than 50 is given below:

| Identify: | 0 | 50 | 100 | Less than 50 | More than 50 |
|-----------|------|------|------|--------------|--------------|
| Percentage: | 0.66% | 9.99% | 3.61% | 40.25% | 49.75% |

The first thing we notice is a huge lump of people at 50, and a medium sized lump at 100. There is no corresponding lump at 0, but there is a lump at 1:

| Identify: | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|-----------|------|------|------|------|------|------|------|
| Percentage: | 0.66% | 2.01% | 0.70% | 0.74% | 0.35% | 0.63% | 0.28% |

Apparently subjects are averse to identifying themselves as a "zero."

What about the lump at fifty? Is it "real?" Is there really a huge concentration of centrists? Let us consider nearby values to the left:

| Identify: | 44 | 45 | 46 | 47 | 48 | 49 |
|---|---|---|---|---|---|---|
| Percentage: | 0.68% | 1.96% | 0.84% | 0.96% | 1.69% | 1.29% |

And nearby values to the right:

| Identify: | 51 | 52 | 53 | 54 | 55 | 56 |
|---|---|---|---|---|---|---|
| Percentage: | 1.49% | 0.98% | 1.23% | 0.65% | 1.32% | 0.58% |

Compare these values to the fraction at 50, 9.99%. There is a sudden jump at 50, and otherwise a fairly rapid drop toward 1%.

There is some evidence of smaller lumps at 45 or 55. The fraction of self placed ideology values divisible by 5 (except for 0 , 50, or 100) is 32.1%. If values were uniformly distributed this number would be (just under) 18%. The fraction of self placed ideology values divisible by 10 (except for 0, 50, or 100) is 16.2%. Again, if values were uniformly distributed this would be (just under) 8%. There seems to be an attraction to round numbers divisible by 5 or 10.[3] This partly explains the concentration at 100. It makes the lack of people at zero particularly interesting.

Another way to see these artifacts is to regress the number of (sample weighted) persons at each value of self placed ideology (101 observations) on dummies for divisible by 5, divisible by 10, and equal to values 0,1,50, or 100. The results of such a regression are as follows:

| Dummy: | Divisible by 5 | Divisible by 10 | =0 | =1 | =50 | =100 |
|---|---|---|---|---|---|---|
| Coefficient: | 0.94 | 0.42 | -1.36 | 1.35 | 7.97 | 1.59 |
| Standard Error: | 0.12 | 0.16 | 0.36 | 0.04 | 0.36 | 0.36 |

For example, on average an entry divisible by 5 (but not divisible by 10) contains 0.94% more of the surveyed population than baseline entries (not divisible by 5 and

---

[3]this does suggest one possible alternate measure of self-placed ideology would be to collapse the 101 point scale down to a 21 point scale. This approach is left to future work

not 1). All coefficients are statistically different from zero at the 95% confidence level. In particular, the surplus of people reporting at exactly 50 or 100 can not be explained simply as the tendency to pick values divisible by 5 or 10.

I suspect part of the reason that there is such a large surge at exactly 50 is that people have an artificial, culturally induced desire to self-identify as centrists. People's aversion to report an ideology of zero, and a propensity to report ideologies divisible by 5, is evidence that people are choosing what ideology to report for reasons, such as aesthetics, other than their "true" ideology.

An alternative explanation is that some survey participants are simply trying to get through the survey as fast as possible, and thus mindlessly report the middle value. One possible measure of this tendency is the number of people who report both parties at being at the 50 position ideologically. The fraction of people who say both parties are at the center is 6.55% among those who self-place at 50, but at 0.25% for the rest of the population. On the other hand, 8.15% of the entire survey population assigns the same ideology value to themselves and both parties in the entire survey. For example, someone who ranks themselves at 40 in this category would also ranks both Democrats at 40 and Republicans at 40. This is either more evidence of lazy respondents, or evidence of confused respondents, or evidence of respondents seeing political harmony.

Regardless of which explanation is correct, there is reason to be suspicious of those who self-report at 50. Let us consider the distribution of principal factor rank scores for those who self-report at 50:

Similarly to self-identified moderates, survey respondents who identify at 50 are left of center. Again, part of this is because more people identify as more conservative than 50 (49.75%) than liberal (40.25%). If there were a perfect correspondence between rank score and self placement, the median person in the subsample of people with self place at 50 should have rank 0.450. Instead the median rank score is 0.417. Rank scores should, in the perfect correspondence case, vary from 0.403 and 0.502. 0.403 is approximately the 48th-percentile of the distribution; .502 is approximately the 63rd-percentile of the distribution. That is, about 48% of the subsample is "too
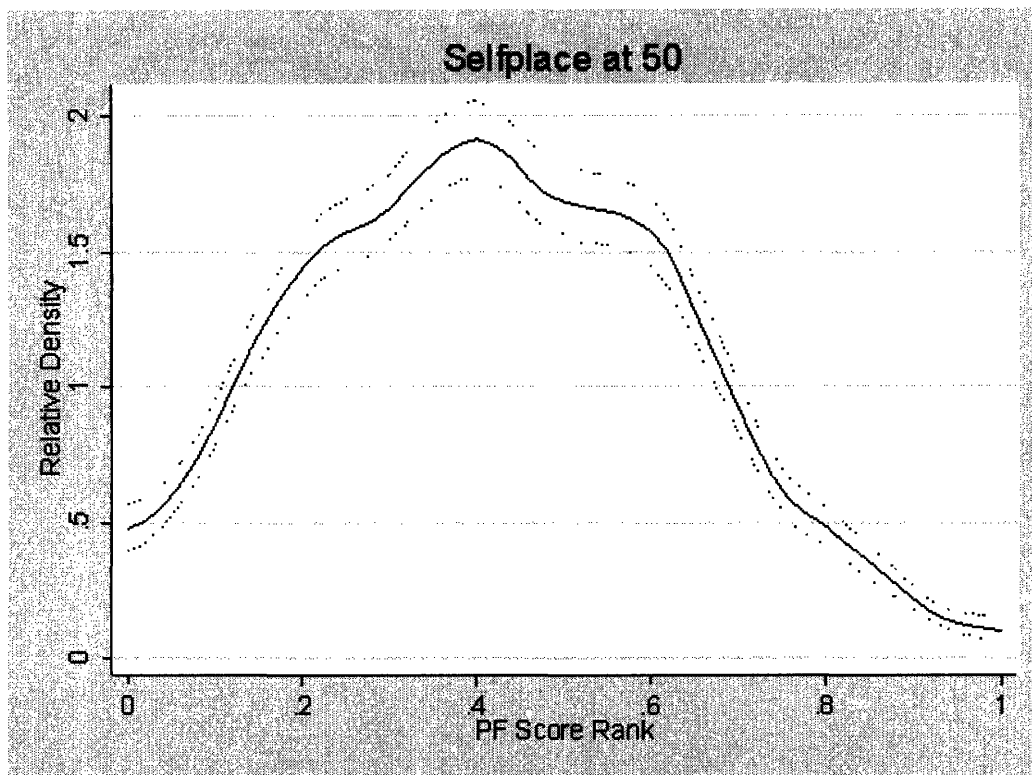
Selfplace at 50

liberal" and 37% are "too conservative."

Just as in the five-point scale, self-identified liberals score as centrists more than self-identified conservatives in the 100-point scale. Consider extreme liberals (selfplace = 1) and extreme conservatives (selfplace = 100). In the following graph, the solid line represents distribution of rank scores for people who self-identified as extreme conservatives. The dashed line represents one minus the rank scores of those who identified as 1 on liberal scale.

I have left off confidence intervals for clarity. The pointwise confidence intervals cross outside the interval (.4, .6). In the following graph, solid lines are the estimated density functions, while the dotted lines represent the bootstrapped 95% confidence intervals. The higher curve represents self-identified mirrored liberals, the lower line represents conservatives.

The pointwise density estimates are statistically distinguishable at median rank. The point estimates are dramatically different. At the median, the point estimate density of self-identified liberals is 0.53. For self-identified conservatives it is 0.17.

85

**Selfpace 1 (mirrored) and 100**



**Selfplace at 1 and 100**
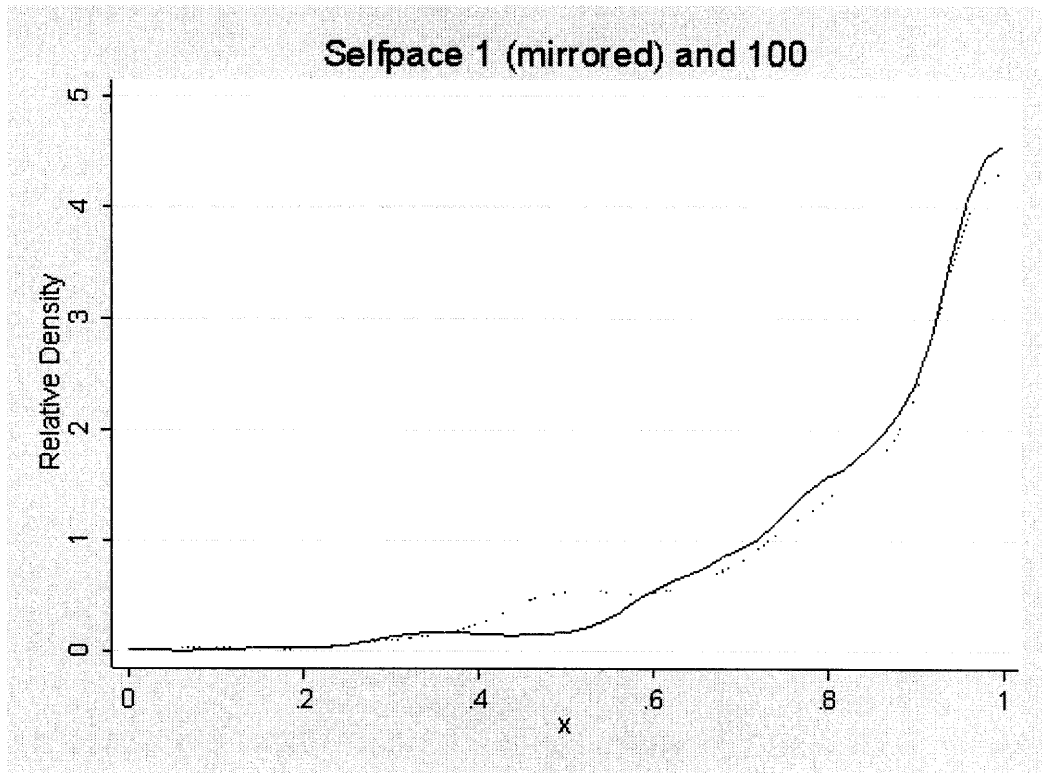
Of self-identified liberals, there is a higher fraction of self-identified liberals who are really centrists then self-identified conservatives who are really centrists.

### 4.4.3: Party ID

In this section we will look at ideology scores by party ID. The first measure of party ID allows the respondent four answers: Democrat, Republican, Independent, and Other.
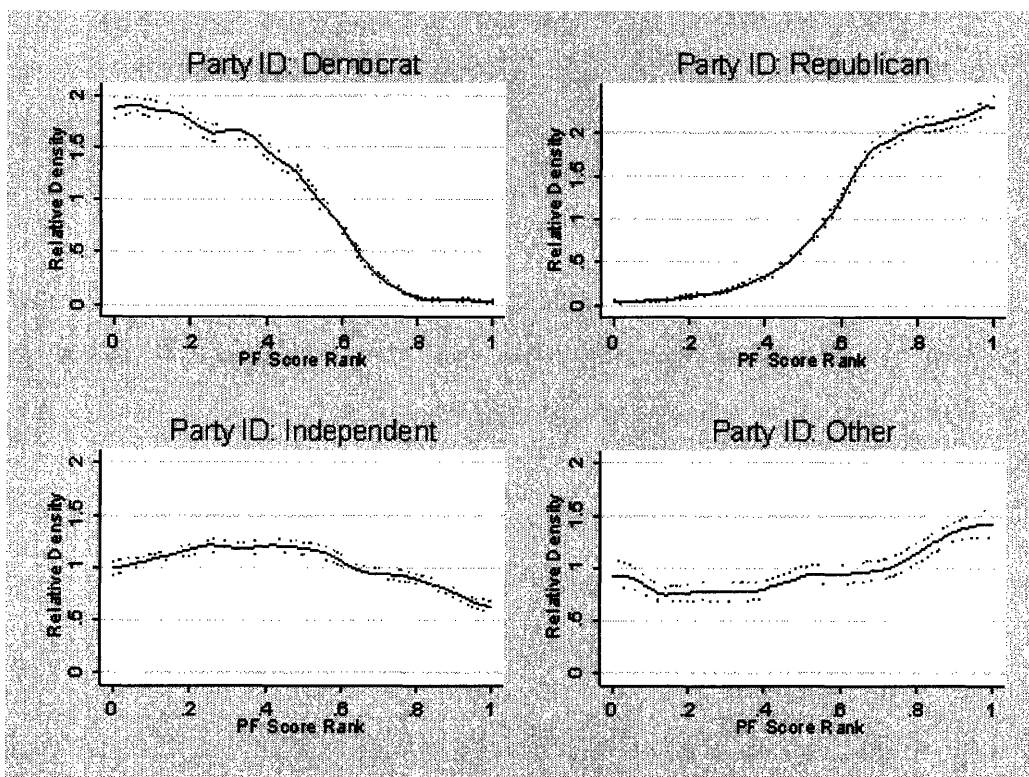


One immediate observation from the above table of graphs is that self identified "independents" on a simple party ID scale is not a good proxy for policy centrists – like self-identified moderates they are slightly left-of-center. The median independent has score 0.45. Furthermore, in this case one can not argue that this is an artifact of more identifying in the conservative category than the liberal category; 33.1% identify as Democrats but only 30.9% identify as Republican. Even more troublesome, the distribution of self-identified independents is relatively flat – there are substantial numbers of independents with far-left policy views and with far right policy views.

A random independent is hard to distinguish from a random person from the general population.

As is the case with self-identified ideology, Democrats are more likely to be policy centrists than Republicans.

| Party | Density at rank 0.5 | Lower Bound | Upper Bound |
|---|---|---|---|
| Democrat | 1.177 | 1.100 | 1.250 |
| Republican | 0.674 | 0.620 | 0.728 |
| Independent | 1.179 | 1.121 | 1.240 |

Interestingly, Democrats have a very high representation in moderate ideology scores. The fraction of Democrats near rank 0.5 is indistinguishable from the fraction of Independents near rank 0.5.

In another set of questions, respondents who claimed affiliation with Democrats or Republicans were asked if their affiliation was "strong" or "weak." furthermore, those who identified as independent were asked whether they leaned Democrat or Republican, or had no lean.

The next graph shows that self-identified independents on a 7-point scale are a much better proxy for centrists. Nevertheless independents still lean slightly to the left. The median self-identified independent has rank score of 0.474; first quartile 0.300 and third quartile is 0.636.

Another phenomenon is people who identify as independent but then lean Democratic (Republican) are more liberal than people who identify as Democratic (Republican) but only "weakly." The same phenomenon extends to voting behavior (see chapter 5). Even controlling for ideology, an independent who leans Democratic is more likely to vote Democratic than a weak Democrat.

The phenomenon of liberals and Democrats having more centrists than conservatives and Republicans persists in the 7-point scale:

However, the "weak" partisan identifiers - even weak Republicans - are likely to be centrists. The tendency for weak partisan identifiers to be centrists outweighs the tendency of Republicans to not be centrists.

## Self Identified Ideology
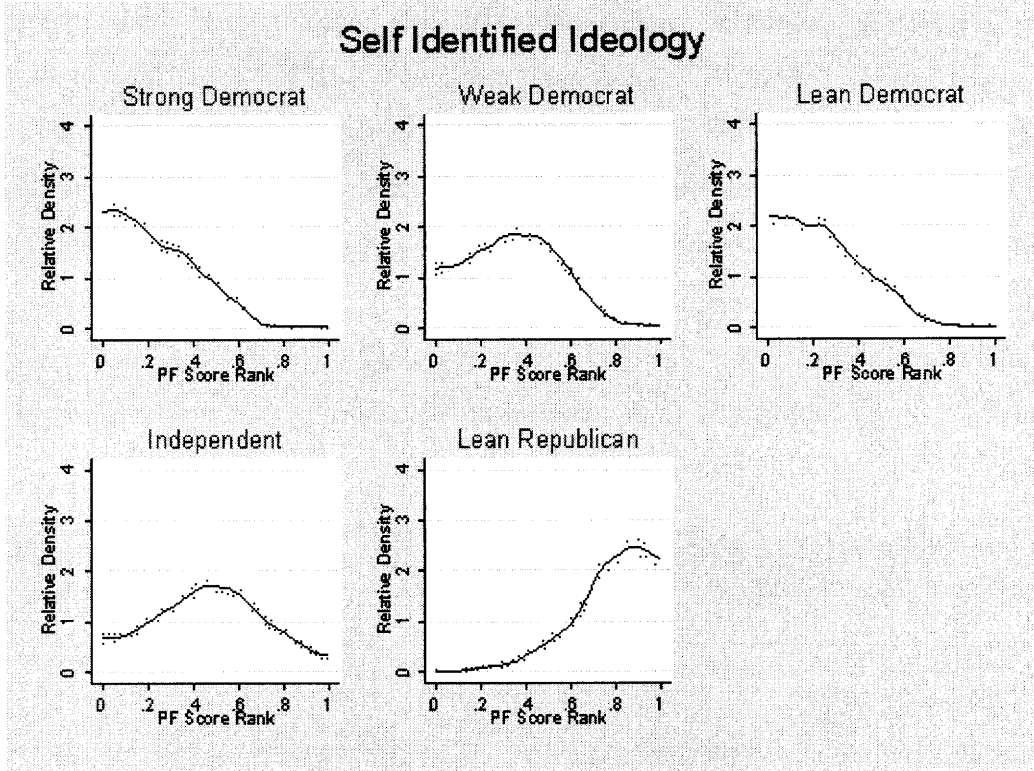


| Party ID | Density at rank 0.5 | Lower Bound | Upper Bound |
|----------|---------------------|-------------|-------------|
| Strong Democrat | 0.890 | 0.807 | 0.969 |
| Weak Democrat | 1.650 | 1.543 | 1.772 |
| Lean Democrat | 0.890 | 0.799 | 0.983 |
| Independent | 1.697 | 1.586 | 1.805 |
| Lean Republican | 0.520 | 0.601 | 0.687 |
| Weak Republican | 1.366 | 1.257 | 1.485 |
| Strong Republican | 0.260 | 0.311 | 0.364 |

### 4.3.3 States

In this section we will examine the distribution of ideologies by state. As discussed in the introduction, one of the primary goals of this section is to double-check the scores by making sure state distributions roughly comport with state distributions. The other goal is to check for variation of the number of centrists between states. According to the seats/votes model of chapter 3, if the number of centrists varies significantly state by state, traditional uniform swing methods of evaluating the seats/votes curve will be flawed.

Another important trait of interest for states (and House districts) is the distribution of state medians. However, analysis of this characteristic will be deferred to chapter 5.
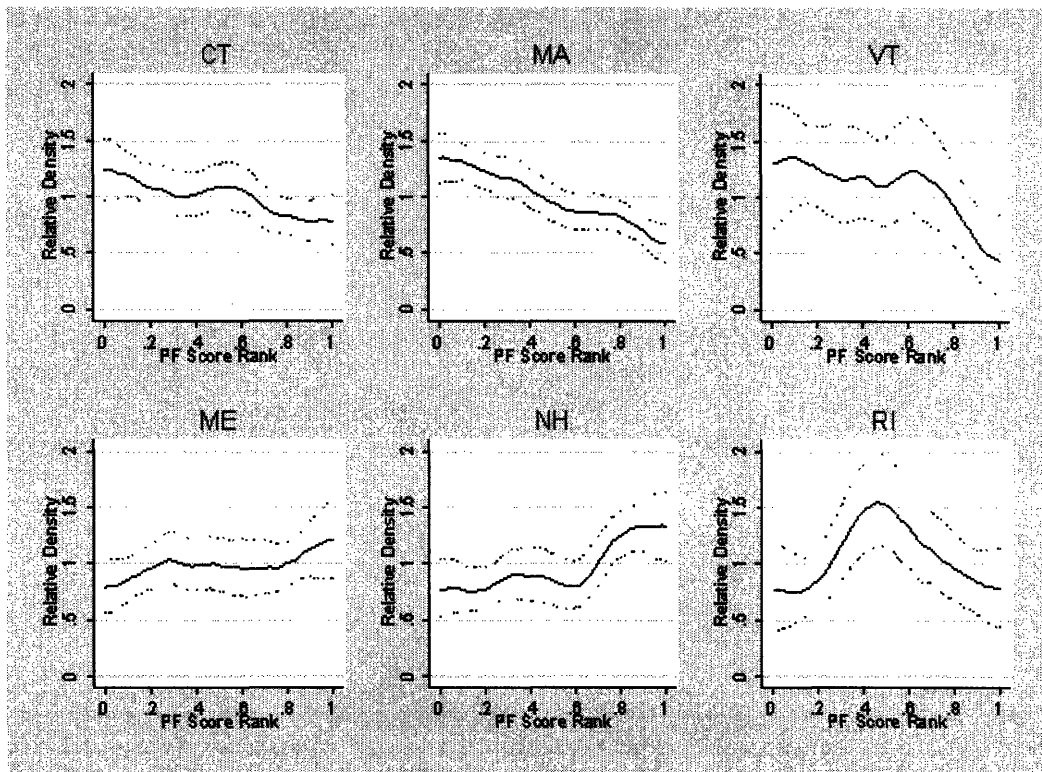


As expected Connecticut, Massachusetts, and Vermont have a high concentration of liberals and a low concentration of conservatives. Perhaps less obviously, they also have a fair share of policy centrists. It is also worth noting that there are indeed

conservatives in these states. If we compare these graphs to, say, self-identified conservatives or self-identified Democrats, there is a marked difference. This observation is fairly obvious - in any large demographic area such as a state there will be a variety of opinion on policy questions. however, it is important not to lose sight of this fact when interpreting the distribution of ideology in the country. If we consider the 20% most conservative citizens, more than 10% of the citizens of the state of Massachusetts are as conservative.

New Hampshire and Maine, in the survey, are measured as MORE conservative than the country as a whole. This makes some sense, since their senators are all Republican. However, it does add focus on the solid support of New England for Kerry in 2004. The mystery is not why Maine and New Hampsire have Republican senators; the mystery is why they voted for a Democratic presidential candidate.
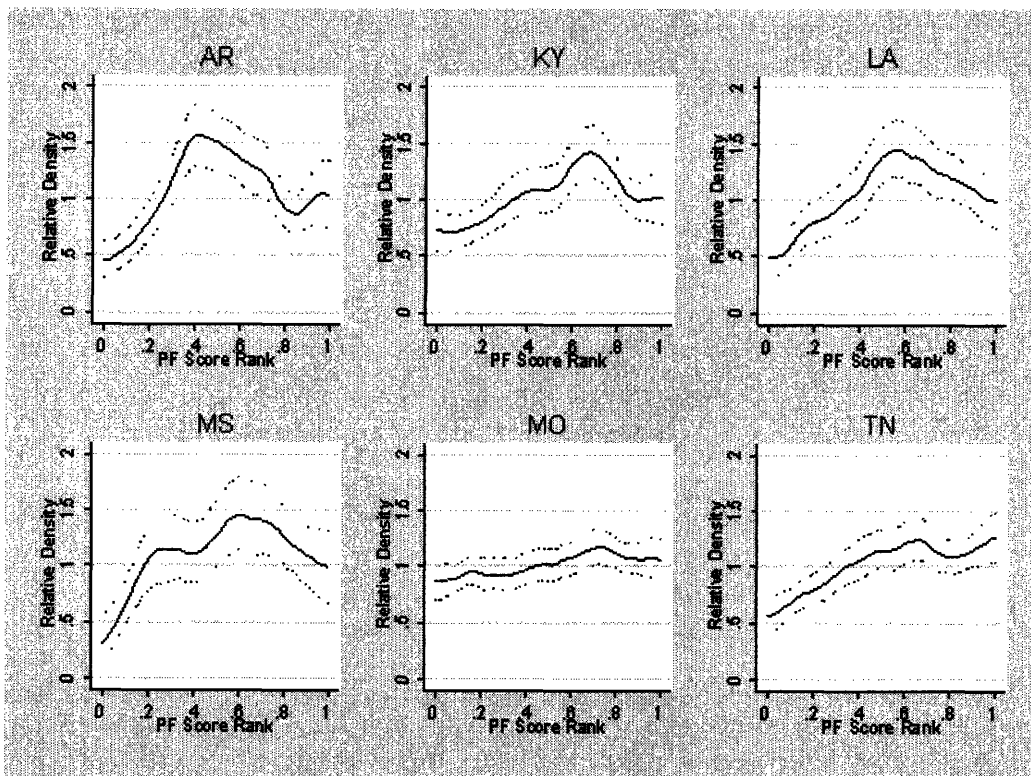


Florida and Virginia reflect the nation as a whole remarkably well. Most other southern states lack very strong liberals.

Continental west coast states have a very strong number of extreme liberals. For
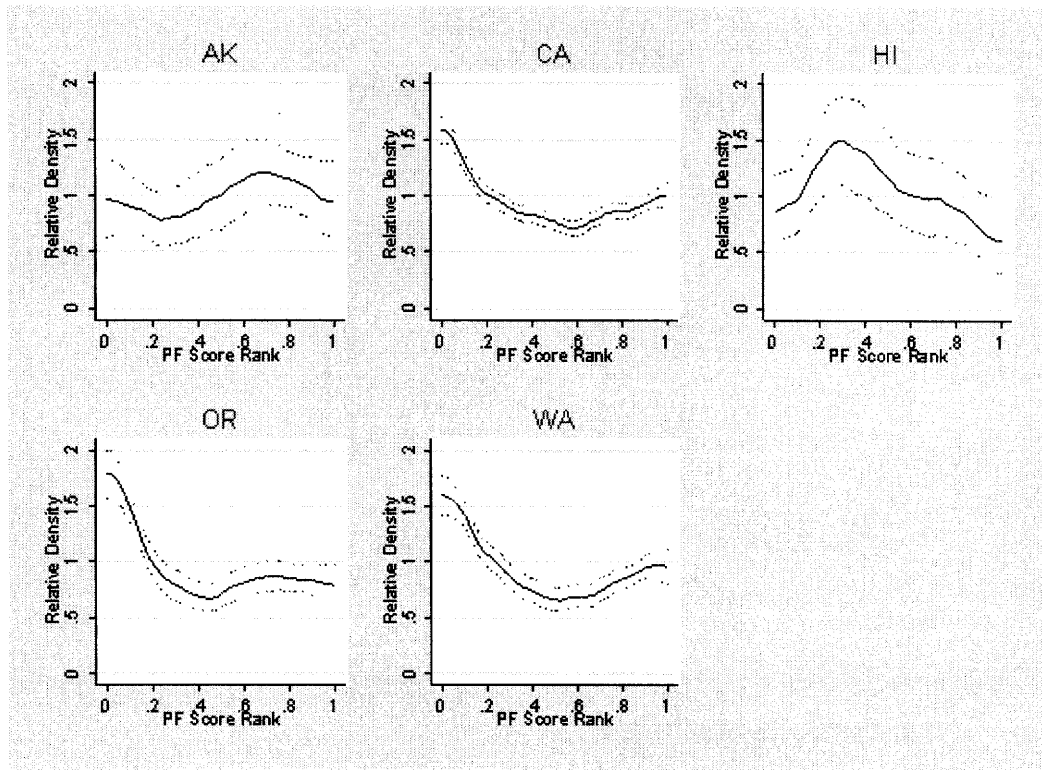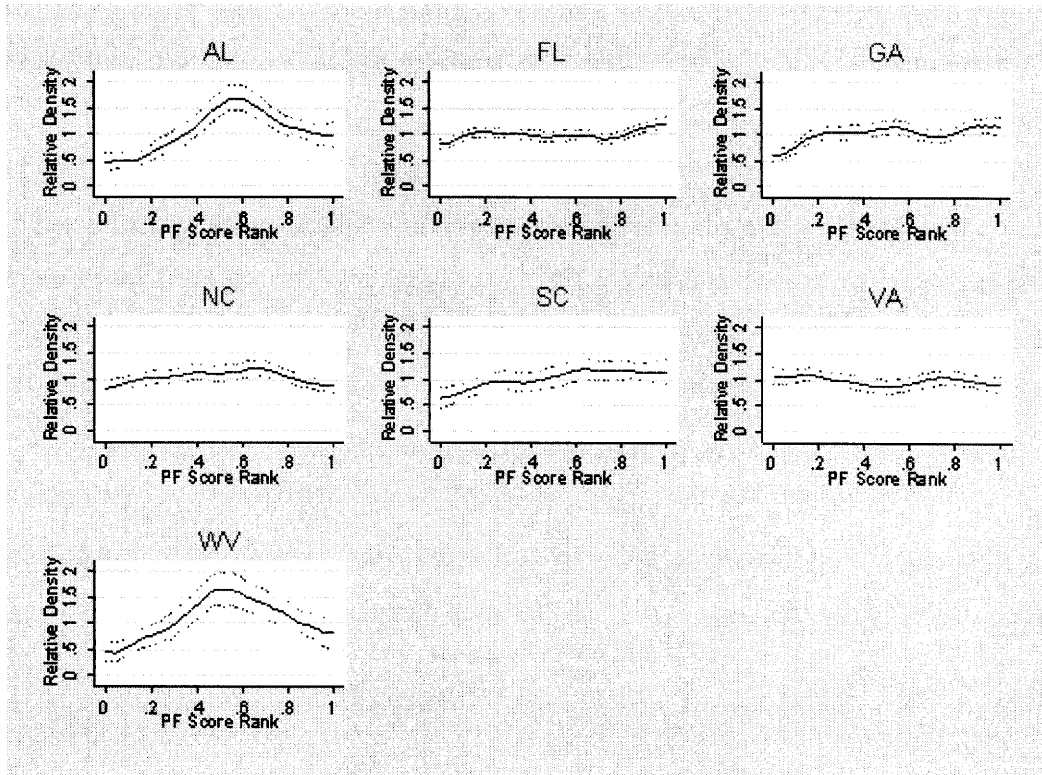
example, consider the state with the left-most median, Massachusetts. While the median voter in Massachusetts is left of the median voter of California, Oregon, and Washington, the density of extreme liberals is higher in the pacific states than Massachusetts. The pacific states also have their fair-share of extreme conservatives (density near one in California and Washington). What makes these states unique is a relative absence of centrists.

The sample in Hawaii is small, so the following must be taken with a grain of salt. Nevertheless, there is some evidence the modal ideology of Hawaii is right of zero. This contrasts with liberal states such as Massachusetts or California, where there is a mode near zero. It compares to Utah, described below, which is also a strongly conservative state, but with modal ideology less than 1. Despite the differences between Utah and Hawaii, perhaps there is something similar about the politics of both states. Exploring such a possibility is left to future work.
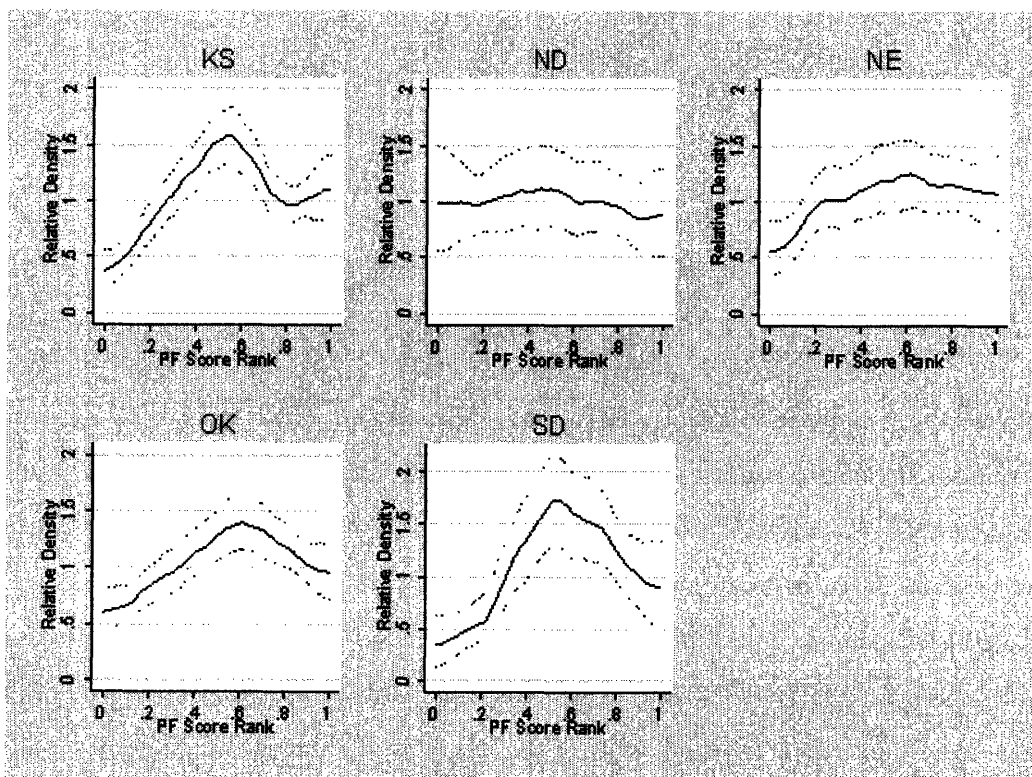


As per conventional wisdom, plains states lack strong liberals. The exception is North Dakota; however the small number of observations ensures large amount of

error. On the other hand, North Dakota is an odd state. Democrats control house and senate seats, while Republicans control state government. Whether North Dakota is in fact a liberal state, a conservative state, a representative state, or a moderate state remains unclear.
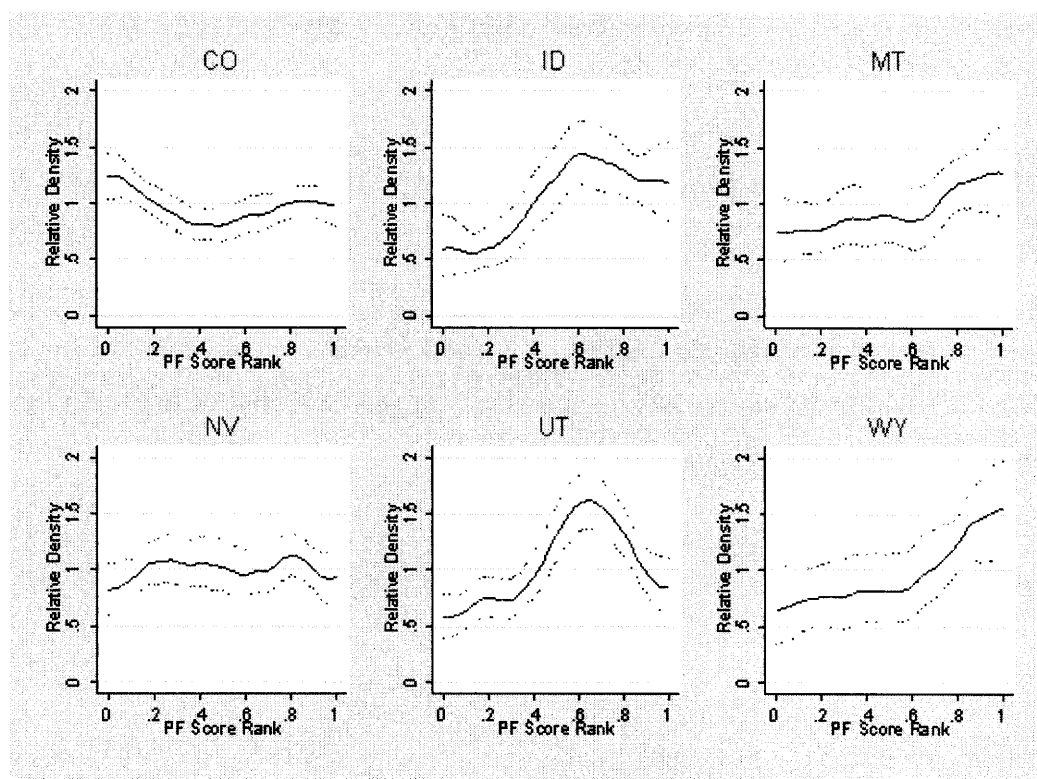


Nevada and Colorado are remarkably representative of the country as a whole. Colorado is actually left of center in the survey (median rank score is 0.487). Considering 2006 elections (Democratic governor, four out of seven house seats Democratic) the ideology distribution is probably reasonable. As discussed above regarding Hawaii, Utah is a conservative state, but the modal rank score is left of 1.

Perhaps contrary to popular expectation, Texas and Arizona have a representative number of liberals (in terms of responses to questions) compared to the nation as a whole. As expected, however, both states have a shortage of moderates and a surplus of conservatives. However, overall Texas and Arizona are still fairly representative, at least compared to the pacific states. That is, the concentration of extreme conservatives in, say Texas, is less than the concentration of extreme liberals in, say,

94
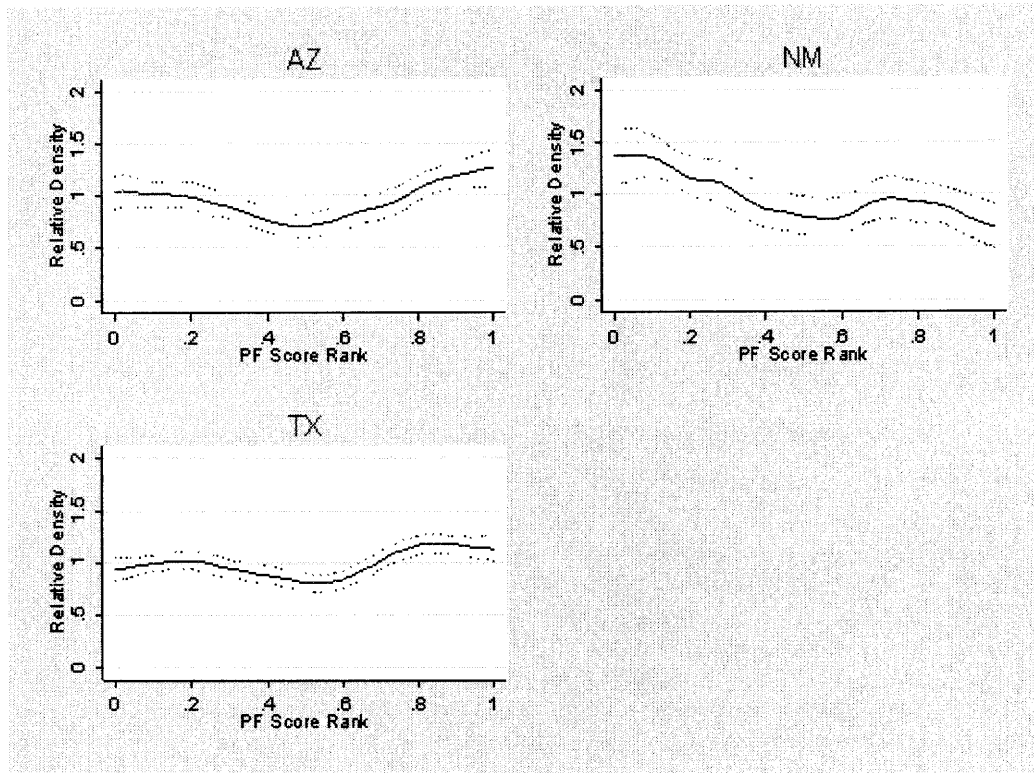
Oregon.

The midatlantic states do not fit any fixed pattern, despite them all being "blue states." Pennsylvania is actually right of center in the survey (median = 0.516). In contrast, New York looks similar to Massachusetts.

Fitting the stereotype, midwest states mostly mirror the country as a whole. The exceptions (Minnesota having a few extra extreme liberals; Indiana short on extreme liberals) also fit stereotypes for those states.

This section has focused on Senate districts (states) because for many states the sample sizes are large enough to make entire distribution graphs meaningful. However, one could pick out a few particularly interesting districts and check one particular part of their distributions. One set of interesting districts are districts held by House leadership members. John Murtha is included in the following table since he was, at one point, a serious contender for the majority leader position.

As has been observed elsewhere for past House leaderships (Cox and McCubbins 1993), there is no single pattern for leadership districts. Some represent extreme

DE  MD  NJ

NY  PA

IA  IL  IN

MI  MN  OH

WI

Relative Density

PF Score Rank
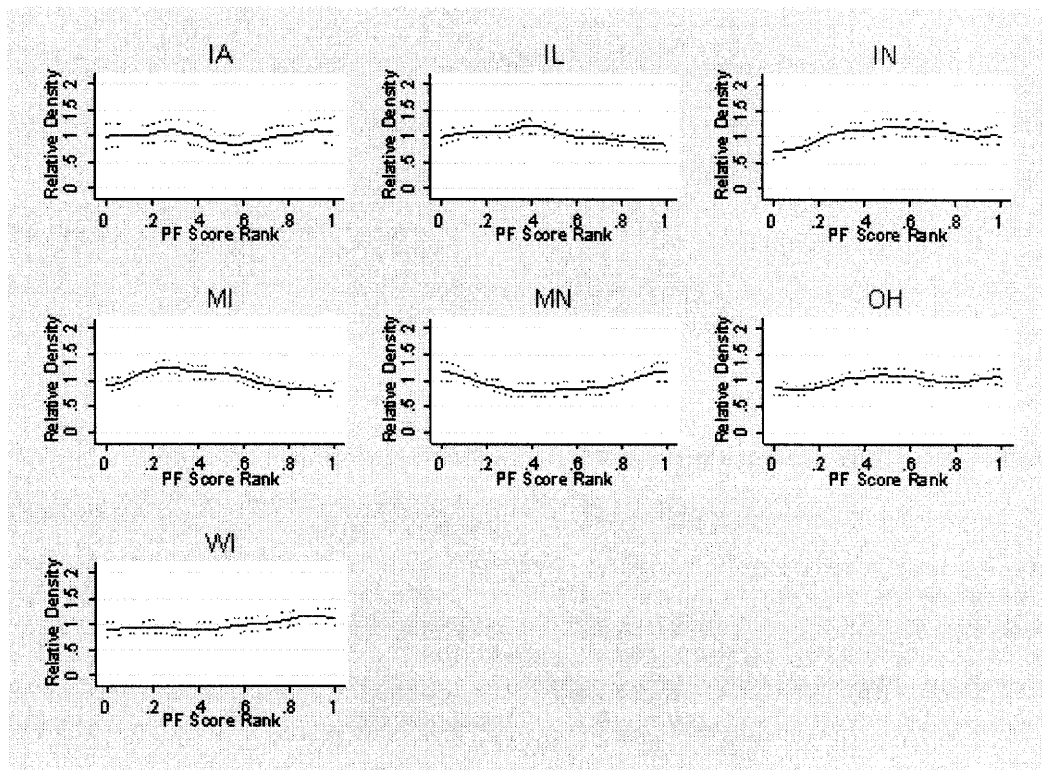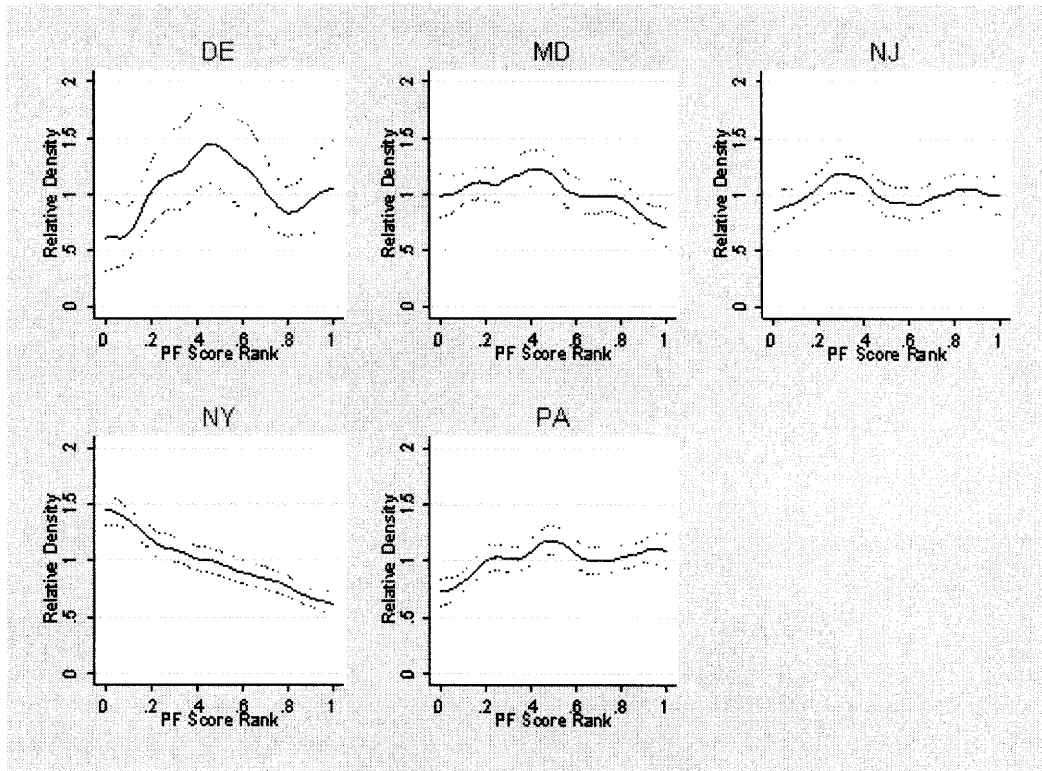
| Name | Position | Median | Lower Bound | Upper Bound |
|------|----------|--------|-------------|-------------|
| Pelosi | Speaker | 0.16 | 0.10 | 0.23 |
| Hoyer | Maj. Leader | 0.465 | 0.40 | 0.52 |
| Clyburn | Maj. Whip | 0.51 | 0.44 | 0.55 |
| Murtha | N/A | 0.49 | 0.46 | 0.56 |
| Boehner | Rep. Leader | 0.57 | 0.53 | 0.68 |
| Blunt | Rep. Whip | 0.62 | 0.50 | 0.68 |

districts (Pelosi) some moderate districts (Hoyer and Clyburn) and others districts are near the center of their party caucus (Boehner). All three possibilities have been described as reasonable in the literature (extremists are electorally safe and can therefore focus on duties to the House, caucus medians are the natural choice of strong party caucuses, and House medians are the natural choice for maximizing the size of the party in a Downsian sense.)

## 4.4 A Critique of Linear Factor Scores

One problem with ordinary factor analysis, especially with non-continuous variables, is the assumption of a linear relationship between underlying scores and the question answers may be violated. For binary questions, the linear model amounts to a linear probability model - the chance a subject gives the conservative answer to a given question should increase linearly with her score. Unfortunately, it appears the linear probability model is indeed violated for at least some of the questions used.

The preceding graphs were constructed by fitting a probit model to principal factor scores for two different questions; one on the minimum wage, one on immigration reform. One outlier observation was deleted. The right graph roughly fits the assumptions of a linear probability model. The linear regression of immigration on principal factor score has a coefficient of 0.04986 (0.0003). A person one point more conservative than another person has about a 5% greater probability of being against amnesty.

The left graph does not fit a linear model very well. All scores below zero almost always support the liberal position on minimum wage. There is essentially a
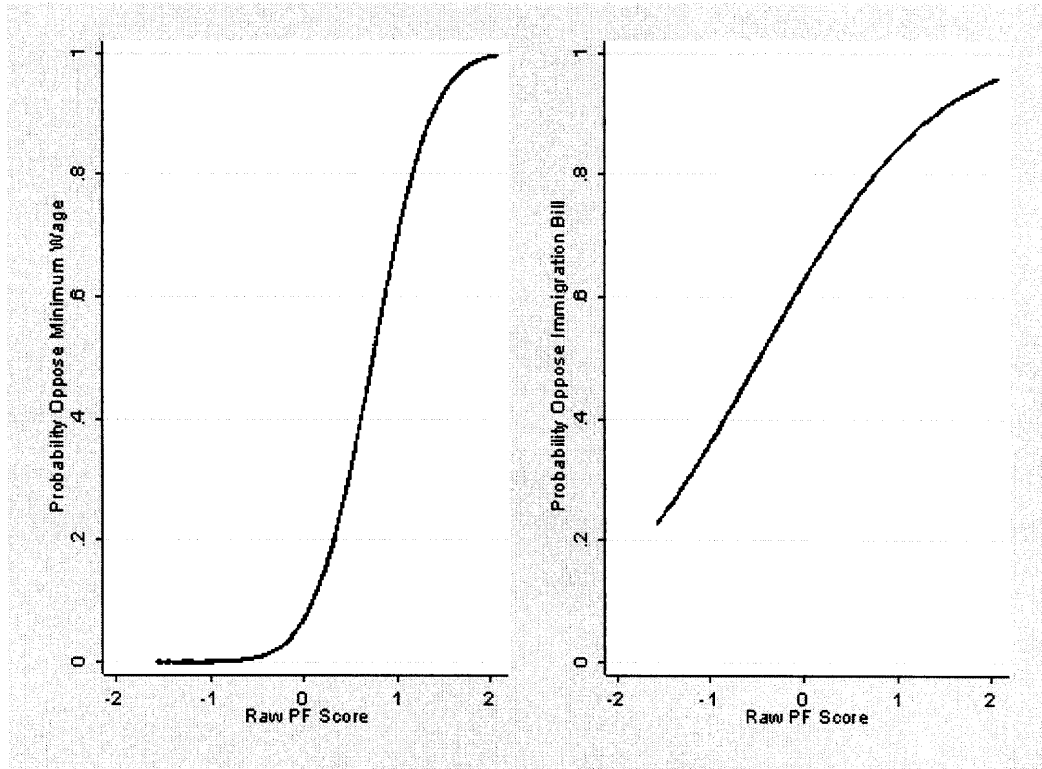
cutpoint where everyone below the cutpoint votes liberally; everyone above votes conservatively.

What will happen if one tries to force the linear factor model onto binary, cutpoint-behaving questions? Suppose the true underlying score distribution was uniform. Suppose we started with a battery of questions which correctly generated uniform scores. Then suppose we added one more question which had a perfect cutpoint like behavior, with a cutpoint at value c.

The factor analysis method will assign a coefficient to this question; after all, it is strongly correlated with the underlying factor. Suppose the model assigns a coefficient of $\beta$ to such a question. To first order no-one will have an estimated factor score within $\beta$ of the cutpoint c. For any $0 < \varepsilon < \beta$, anyone with a true score of c $+ \varepsilon$ will be given an assigned score of approximately c $+ \varepsilon + \beta$ (assuming the other estimated coefficients do not change much).

In actual data there will be noise and changes in other coefficients to dampen the effect, but we should expect to see dips in the distribution of scores around the cut-

points. The following graph compares two kernel density plots - one of the principal factor scores using all the questions (solid line), and one with all the questions except for minwage (dashed). Including the minwage question exaggerated the bipolarity of the scores. The number of persons with scores around 1 (roughly where the cut-point between yes's and no's is, according to the previous graph) is depressed when minimum wage is included.
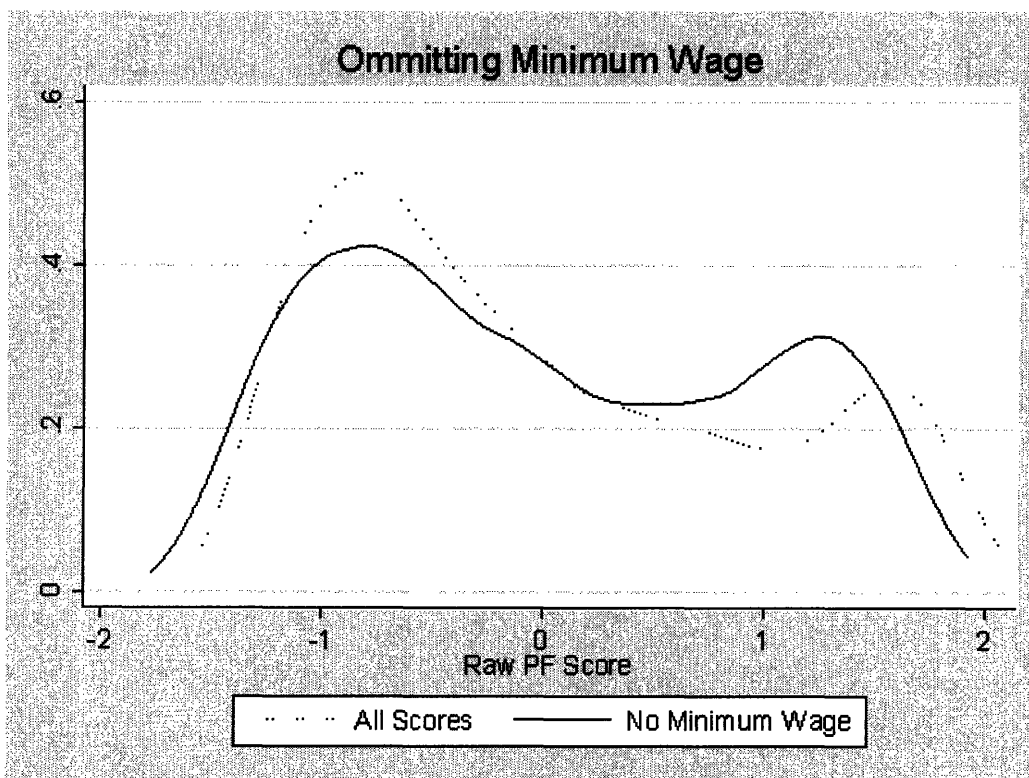


In contrast, the following graph compares full question scores and scores omitting immigration question. There is no substantial change:

## 4.5    A Non-Parametric Alternative Score

To assess the problem described in the previous section, we will consider an alternative, more robust measure: Mokken scores (Mokken 1971, Mokken and Lewis 1982).[4]

---

[4]There are other approaches that could of been taken rather than Mokken scores. Factor analysis essentially works with correlation matrix between variables. One proposed solution to the binary

**Ommitting Immigration**

Legend: ...... All Scores ——— No Immigration
x-axis: Raw PF Score
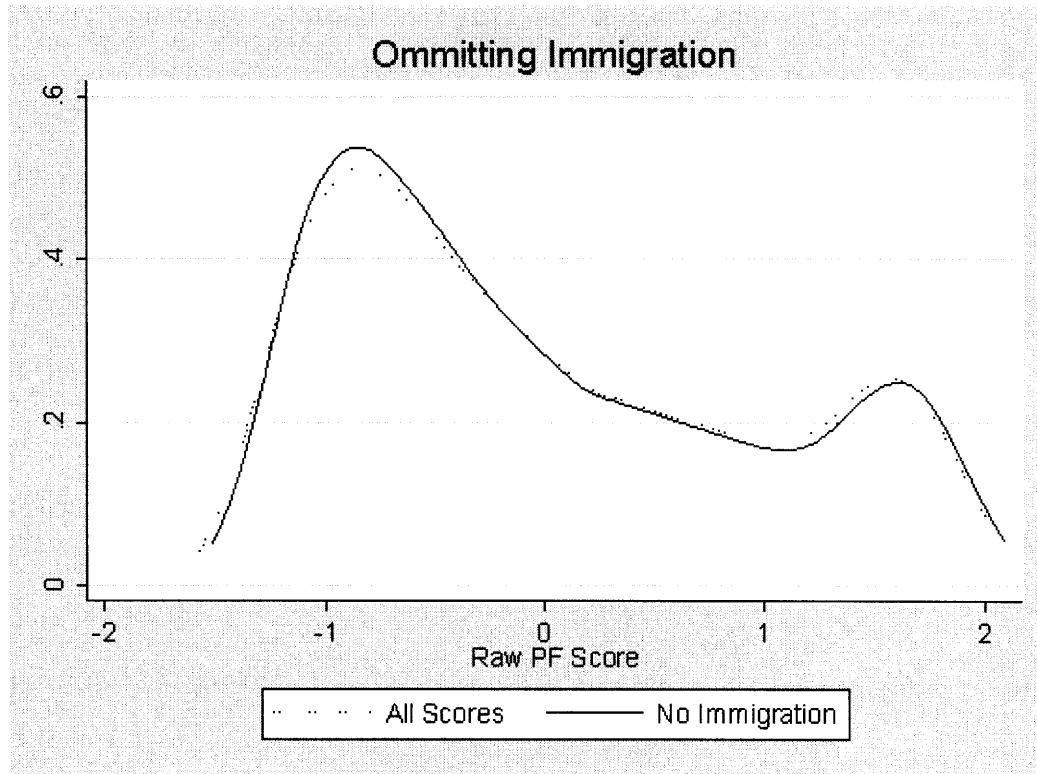
Mokken scores are a generalization of a basic scoring technique: the Guttman (1944) scale. Suppose one had a set of questions such that extreme conservatives answered "yes" to every question, slightly less extreme conservatives answered yes to every question except question number one, even less conservatives answered yes to every question except one and two, and so forth, until the most liberal person said no to every question. If there were such a perfect correspondence between answers and ideology, then one could easily construct a ideology measure. Simply use the number of "yes" responses for each subject.

Naturally, this is only an ordinal measure of ideology, but we have already restricted our attention to ordinal measures in this chapter.

Of course, in practice people do not always line up so perfectly in their responses. A Mokken scale is similar to a Guttman scale, but allows people to make random

---

variable solution to this problem is to use a different correlation concept than Pearson correlation, such as tetrachoric correlation. Another possibility is to abandon factor analysis entirely and use conditional logit, or more general Rasch models. Implementing these approaches is left to future work.

"errors" on occasion. Unlike some methods, like conditional logit, Mokken scaling does not try to interpret the errors as being a measure of distance between questions. The Mokken scale simply requires the number of errors to be relatively small, and remains otherwise agnostic about the distances between question values. The Mokken scale is therefore more general than some parametric models.

To see what is meant by an error, consider two binary questions. One can construct a two-by-two table of possible response patterns. Consider the following example:

|  | Question 2: Yes | Question 2: No |
|---|---|---|
| Question 1: Yes | 400 | 0 |
| Question 1: No | 200 | 1000 |

With a Guttman scale, at least one of the entries in the table must be zero. Suppose we had a priori reason to think Yes responses for both questions were associated with conservatism. Then in the above example, 400 people would be put in the most conservative category, 1000 in the most liberal, and 200 in the middle. One can also rank the questions in this case: Question 2 is more conservative in the sense that only the most conservative say Yes to it.

Now suppose there were errors. In particular, suppose 1 of 100 responses were opposite (e.g. yes-yes to no-no) and 9.5 of 100 responses had the wrong answer (yes-yes to no-yes) on the first question, and 9.5 of 100 responses were wrong on the second question (yes-yes to yes-no). Then the observed response pattern would be:

|  | Question 2: Yes | Question 2: No |
|---|---|---|
| Question 1: Yes | 349 | 135 |
| Question 1: No | 293 | 823 |

The 135 entries in the yes-no position would be "Guttman errors". The Mokken scale procedure has us essentially ignore these errors, as long as there are not "too many" of them. The standard rule of thumb for determining if the number of errors is sufficiently small is determined by Loevinger's H. To find H, take the number of Guttman errors and divide by the number of items that would be in the error cell if

responses were made independently. One minus the above quantity is H. In the above example, the chance of saying yes to question one is (349+135)/1600 = 0.365. The chance to saying no to question two is (135+823)/1600 = 0.59875. If responses were independent, then the chance of saying yes to question one and no to question two is 0.365*0.59875 = 0.218. Therefore, the expected number of entries in the yes-no box under independence would be 0.218*1600=349.67. So H = 1 - 135/349.67 = 0.61. The recommended heuristic is for H to be at least 0.3, preferably 0.4 or 0.5. Therefore, the above pattern would be considered acceptable for the two questions to be included in a Mokken scale.

The above logic can be extended to multiple questions, and non-binary questions. See Sijtisma (2002) for discussion. However, two questions that were included in the principal factor scores were dropped for the Mokken scores. These questions on tax and spending policy allowed 101 different possible responses, which could of caused them to dominate the Mokken score.

The H scores for the questions used (calculated using persons that answered every question only) are as follows:

| Question Topic | H |
|---|---|
| minimum wage bill | 0.60567 |
| minimum wage increase | 0.59862 |
| iraq withdrawal | 0.57672 |
| stem cell research | 0.56418 |
| captal gains tax | 0.52047 |
| social security | 0.49022 |
| "partial birth" abortion | 0.46473 |
| military: democracy | 0.45773 |
| union influence | 0.45434 |
| abortion | 0.44844 |
| affirmative action | 0.42177 |
| environmentalism | 0.41267 |
| military: terrorist camp | 0.39729 |
| military: oil | 0.37788 |
| immigration amnesty | 0.3614 |
| military: allies | 0.32141 |

Another important heuristic for a Mokken score is that all traits should be moving

in the same direction. That is, if all answers represent the same unifying underlying one-dimensional scale, then every question should be positively correlated with every other question. For the set of questions of questions considered that were also used in principal factor analysis, only military use for UN purposes was negatively correlated with another variable (military use for allies) and was therefore dropped. All other variables (when recoded so that the conservative response has a positive value) were all positively correlated with each other in every possible pairing.

It makes sense that military use for the UN and might cause problems. From our survey responses, conservatives generally support military use more than liberals. On the other hand, conservatives tend to have a dimmer view of the United Nations than liberals. For example, in one 2004 poll[5] 60% of self-identified liberals have a favorable view of the UN, while only 18% of self-identified conservatives have a favorable view of the UN. This presumably makes the question cut across ideology, making it a bad predictor of ideology. (Principal factor analysis can still try to make use of the question by finding the degree in which the question partly loads onto first dimension. Mokken score analysis can not do this, however.)

Standard Mokken scales require all subjects to answer every question. At the very least, if the distribution of the sample is to be representative of the distribution in the population, missing answers need to be accounted for. to impute missing scores, the double imputation method was used. See Bernaards, C.A. and Sijtsma, K. (2000). To implement double-imputation, take the matrix of all available observations. For each missing answer corresponding to person i and question j, find person's i's average score for the questions they do answer, add the average score across all available persons for the given question j, and then normalize by subtracting off the average score over all available questions answered by all persons.

If we code yes's as 1's and no's as 0's, then Alfred's average available answer is $\frac{1+1}{2} = 1$; the average value for question 2 is $\frac{0+1+0}{3} = \frac{1}{3}$ and the global average response is $\frac{5}{11}$. (5 yes's, 11 total responses). So Alfred's question 2 imputed response would be $1 + \frac{1}{3} - \frac{5}{11}$, or about 0.879.

[5]See http://www.rasmussenreports.com/public_content/politics/38_have_favorable_opinion_of_u_n__1

| Person | Question 1 | Question 2 | Question 3 |
|--------|------------|------------|------------|
| Alfred | Yes | Missing | Yes |
| Betty | Yes | No | No |
| Carl | Yes | Yes | No |
| Dora | No | No | No |

The imputed Mokken scores provide some support for the previous principal factor analysis. First, the scores are positively correlated. The spearman rank correlation between the two scores is 0.9668. Second, many sub populations of interest correspond in both measures. In the following, the Mokken scores have been transformed into rank Mokken scores, ranging from 0 to 1, with 0.5 being the median score in the general population. The following table lists first quartile, median, and 3rd quartile score ranks for each subpopulation for each measure, Mokken and principal factor.

| Subpopulation | 25% M | 25% PF | 50% M | 50% PF | 75% M | 75% PF |
|---------------|-------|--------|-------|--------|-------|--------|
| | | | | | | |
| Very Liberal | 0.054 | 0.045 | 0.148 | 0.128 | 0.316 | 0.305 |
| Liberal | 0.109 | 0.101 | 0.224 | 0.213 | 0.390 | 0.369 |
| Moderate | 0.232 | 0.234 | 0.413 | 0.412 | 0.600 | 0.595 |
| Conservative | 0.651 | 0.650 | 0.781 | 0.781 | 0.888 | 0.889 |
| Very Conservative | 0.749 | 0.751 | 0.873 | 0.862 | 0.944 | 0.946 |
| | | | | | | |
| Democrat | 0.141 | 0.133 | 0.273 | 0.276 | 0.451 | 0.435 |
| Independent | 0.238 | 0.237 | 0.454 | 0.450 | 0.687 | 0.685 |
| Republican | 0.637 | 0.641 | 0.779 | 0.773 | 0.892 | 0.890 |
| | | | | | | |
| Strong Democrat | 0.105 | 0.104 | 0.224 | 0.223 | 0.378 | 0.377 |
| Weak Democrat | 0.198 | 0.202 | 0.359 | 0.351 | 0.511 | 0.496 |
| Lean Democrat | 0.118 | 0.114 | 0.245 | 0.235 | 0.406 | 0.375 |
| No Lean Independent | 0.293 | 0.300 | 0.469 | 0.474 | 0.636 | 0.636 |
| Lean Republican | 0.657 | 0.672 | 0.786 | 0.795 | 0.891 | 0.895 |
| Weak Republican | 0.541 | 0.542 | 0.677 | 0.670 | 0.809 | 0.799 |
| Strong Republican | 0.718 | 0.712 | 0.840 | 0.827 | 0.924 | 0.922 |
| | | | | | | |
| Self Place = 50 | 0.268 | 0.263 | 0.417 | 0.417 | 0.582 | 0.576 |

The two measures give very similar descriptions of the distributions of subpopu-

lations.

However, there is one subpopulation of special note that the two distributions differ substantially on. That is, self identified moderates who answered every question used in the principal factor score calculation. (N=3130)

| Subpopulation | 25% M | 25% PF | 50% M | 50% PF | 75% M | 75% PF |
|---|---|---|---|---|---|---|
| | | | | | | |
| Moderate: Answered All Questions | 0.268 | 0.251 | 0.450 | 0.421 | 0.613 | 0.600 |

In this subgroup, the median Mokken score is much closer to the median in the general population. As discussed in section 4.3.1, if one accounts for the fact that more people describe themselves as conservative than describe themselves as liberal, the predicted median moderate should score at 0.441. The notion that "moderates are liberals too" is therefore not supported when we restrict attention to those who answer every question, and use Mokken scores. Of course, restricting the sample to only those who answer every question is inappropriate if we wish to interpret the sample distribution as approximating the distribution of the whole population. Questions are not skipped completely at random; some questions are systematically skipped by moderates and conservatives. Nevertheless, there is reason for caution in future work when applying imputation. To a certain extent some of the results here are depending on the use of imputation.

## 4.6 Conclusion

In this section I will go over some of the more significant findings of this chapter.

First, self-identified liberals and self-identified moderates answer policy questions in a similar manner. As a practical matter, this raises concerns that the number of self-identified moderates may not directly correspond to the number of policy centrists. As far as interpreting the result, more evidence is needed to make any conclusive claims. However, one can speculate. Perhaps there are policy liberals who, for cultural reasons, are unwilling to identify as liberals. Or at least, the number in

105

that category is greater than the number of policy conservatives who fear identifying themselves as conservatives.

There are some other possibilities if one looks beyond the pure ideological voter model. Suppose perhaps liberals (relative to conservatives) are more likely to have a mix of views on issues. To the extent that voters have a pool of views and thoughts on an issue and give answers based on a view chosen at random (see Zaller 1992), liberals will be more likely to give mixed answers and therefore be scored as centrists.[6] Another possibility is reverse engineering of opinions. Suppose instead of having fixed policy opinions and then choosing a party based on them, voters picked their party first and then matched their opinions to the party. If conservatives do this more than liberals, than we would observe more policy discipline among conservatives.

Of course, it is also possible that the phenomenon is an artifact of the particular questions asked. Perhaps with different questions the results would change. There is also the possibility that the pattern of non-responses and the imputation technique is creating the effect, as discussed in the last part of section 4.6. This chapter has discussed a measure of voter ideology.

Another important finding in this chapter is the number of centrists does indeed vary state from state. If this were not the case, many of the methods described in chapter 3 would be useless. Or, at the very least, the theory of chapter 3 would imply that methods such as uniform swing would be perfectly adequate to find seats/votes curves and distribution of district ideologies.

This chapter has also considered the appropriateness of using principal factor rank scores. The scores give similar results to non-parametric scores. The distribution of the scores is consistent with our intuition regarding sub-populations, such as self-identified liberals, partisans, and within states. Furthermore, an alternative approach, self-reported ideology, has been shown to have problems. People round off when reporting on a 100 point scale, have attractions and repulsions to focal scores (0, 1, 50, 100), and as discussed above, self-identified moderates may in fact be left of center.

---

[6]Special thanks to Gabe Lenz for alerting me to this possibility.

# 4.7 Appendix: Kernel Density Graphs

In this chapter kernel density graphs are used to estimate the distribution of ideologies within sub populations of interest. This section is meant as a quick and elementary reminder of kernel density estimation and as documentation of the specific methods used to aid in replication of the results. More thorough treatments of kernel density estimation can be found in Silverman (1986).

To estimate the density at a point X non-parametrically, we need a measure of the fraction of observations "near" X. What counts as near? One simple solution is to use the fraction of points within a fixed bandwidth, say d, of each point of interest. This amounts to giving all points within d a weight of 1, and all further points a weight of zero. Alternatively, we could weight points on a sliding scale, assigning nearby points with more weight and further points less weight. The weighting scheme corresponds to the "kernel" in kernel density estimation.

For the graphs in this section, I used an Epanechnikov kernel, which is designed to minimize the mean-squared deviation from the "true" underlying density. In addition to the functional form, a bandwidth must be chosen for the kernel, which influences how much weight the kernel gives to far-off points. Let $\sigma_x$ be the standard deviation of $x$. For the graphs that follow, I use a standard rule of thumb bandwidth $0.9 * \frac{\sigma_x}{n^{\frac{1}{5}}}$.[7]

Another issue for kernel density estimation is determining a confidence interval for the estimate. To find a confidence interval, I use the following bootstrapping method. Suppose I am looking at a subpopulation with N people. First, I generate 1000 random draws, each of size N, drawn with replacement from the population of interest. This generates a set of weights. Some members are drawn zero times and

---

[7]Hall (1991) argues that one should use a narrower bandwidth when presenting bootstrapped confidence intervals. This is because the bandwidth was chosen to minimize mean squared error, while conventional bootstrapping assumes zero bias. Reducing the bandwidth will sacrifice some mean squared error, but have the advantage of reducing bias, making bootstrapping more appropriate. On the other hand, Chen (1996) argues and provides Monte Carlo evidence that perhaps one should use a wider interval. It is quite possible we should use a wider or smaller bandwidth. Other bandwidths were tried and generally seemed to smooth away subtle ripples in distribution, or provide overly spiky distributions. It is also worth noting that the particular bandwidth rule of thumb chosen is optimized assuming normally distributed deviations from the true distribution. To the extent this assumption is false, the rule of thumb bandwidth will not be optimal anyway. The best bandwidth choice could very well be larger or smaller.

are assigned a weight of zero; some are drawn once and are assigned a weight of one; others are drawn multiple times and are assigned a weight equal to the number of times drawn. I then multiply these weights by the sampling weights of the survey to generate a combined weighting value. I then construct a kernel density estimator for each of the thousand samplings, using the corresponding new weighting values for each application.

The above process generates 1000 density estimates for each point. At each point I sort the estimates from highest to lowest. The density graphs that follow consist of the 25th, 500th, and 975th points on the list. The top and bottom graphs then represent the bounds of a 95% confidence interval for the density function.

There is one other technical complication worth mentioning here. Standard kernel density estimation assumes that the density being estimated is continuous. Suppose the true density jumps in value from a zero to a two. The kernel density method will try to "smooth" over the jump, giving a density of around one on both sides of the jump.

Many of the measures this paper considers are bounded above and below. The density of ideologies outside the bounds is, by construction, exactly zero. The density just inside the bounds will be underestimated by standard kernel density estimation. Kernal density estimation will also estimate positive density outside the bounds - a nonsense result.

There are several approaches taken in the literature to the correct the problem (Silverman 1986, Cowling and Hall 1996).[8] The solution taken here is to reflect the data around the boundary points. The kernel estimator will then use each nearby point twice to estimate the density. Naturally, this makes the estimator more sensitive to points near the boundary. Because of the bootstrapping method used to find confidence intervals, this will be properly accounted for. Typically the estimated confidence interval, appropriately, expands near the boundary values. Another possible problem is the method constrains the derivative of the estimated density function to

---

[8]Other solutions include shrinking the bandwidth as one approaches the discontinuity, or using asymmetric kernel density functions.

equal zero at the boundary. However, we will not focus on the derivative of density functions. The density functions are usually measured with enough noise that the derivatives would be meaningless. Nevertheless, one should be extra wary of interpreting the derivative of the estimated density functions near the boundary points.

# Chapter 5

# Distribution of Districts

## 5.1 Introduction

In this chapter we will use the principal factor rank scores from chapter 4 to find hypothetical seats/votes curves and the distribution of districts. We will use the theory discussed in chapters two and three to interpret the curves and their meaning for representation and competitiveness.

Specifically, the theory we will apply is as follows. If we observe seats/votes bias (that is, seat share does not equal 0.5 when vote share equals 0.5) then this implies that the median voter of the median district is not the median voter in the general population (chapter 3). Furthermore, since party platforms only depend on the distribution of districts (chapter 2) party platforms will not correspond to the interests of the population as a whole. Seats/votes curves will be derived assuming perfect ideological voting in section 5.2 below.

We will also apply the Quartile Voter Theorem of chapter 2. To do so, we will have to transform the scores so that utility functions are symmetric. This is discussed in section 5.3 and 5.4 below. The Quartile Voter Theorem states that if utility functions and the distribution of districts are symmetric, then the long-run equilibrium platforms will be at the first and third quartiles. We will therefore check the quartiles, and see if they are indeed equidistant from the median district (a necessary condition for symmetry.) As discussed in chapter 2, if one quartile is closer to the

median than the other quartile, this implies that the corresponding party is likely to have a platform closer to the median voter of the median district, and therefore will have an electoral advantage.

Section 5.5 adds party ID and incumbency to the seats/votes curves. The advantage of this model is it should be a more realistic portrayal of the actual seats/votes curve. The drawback is that it can not be interpreted in terms of the distribution of districts. However, the formulation considered here does have one advantage over traditional seats/votes curves. That is, in addition to the seats/votes relation, we can also construct the seats/valence and votes/valence relationships. One can therefore see how seats will vary with valence directly, instead of using vote share as a proxy for valence. One can also check whether vote share is indeed a good proxy for valence.

## 5.2   Hypothetical Seats/Votes Curves

According to the perfectly ideological voter model, in any given election there will be a cut-point (See chapter 2). All voters left of that cut-point will vote for the left party; all voters right of the cut-point will vote for the conservative party. The rank-order PF score makes it easy to evaluate hypothetical seats/votes curves. Suppose vote share for the left party is V. Under the perfect ideological voter model, everyone with rank from 0 to V will vote for the Democratic party.

In theory to find the seat share one can simply find the median rank ideology in each district. At a national Democratic vote share of V, all states with median rank ideology less than V will vote Democratic. We can then finds pairs of V and S accordingly.

In practice, however, the procedure used to construct the following graphs utilized bootstrapping.

1. Sample with replacement from every subject in the survey. The total sample size equals the number of survey respondents.

2. Pick a value of V

3. For each state, find the weighted number of subjects with scores equal to or

greater than V.

4. Count the number of states where the weighted Democratic support is greater than 0.5. This is the estimated seat share at vote share V for a given sampling iteration

5. Repeat steps 2-4 for each value of V of interest.

6. Repeat steps 1-5 1000 times. This generates 1000 seat shares at each value of V. Graph the 25, 500, and 950th seat share for each V.

First consider the House.
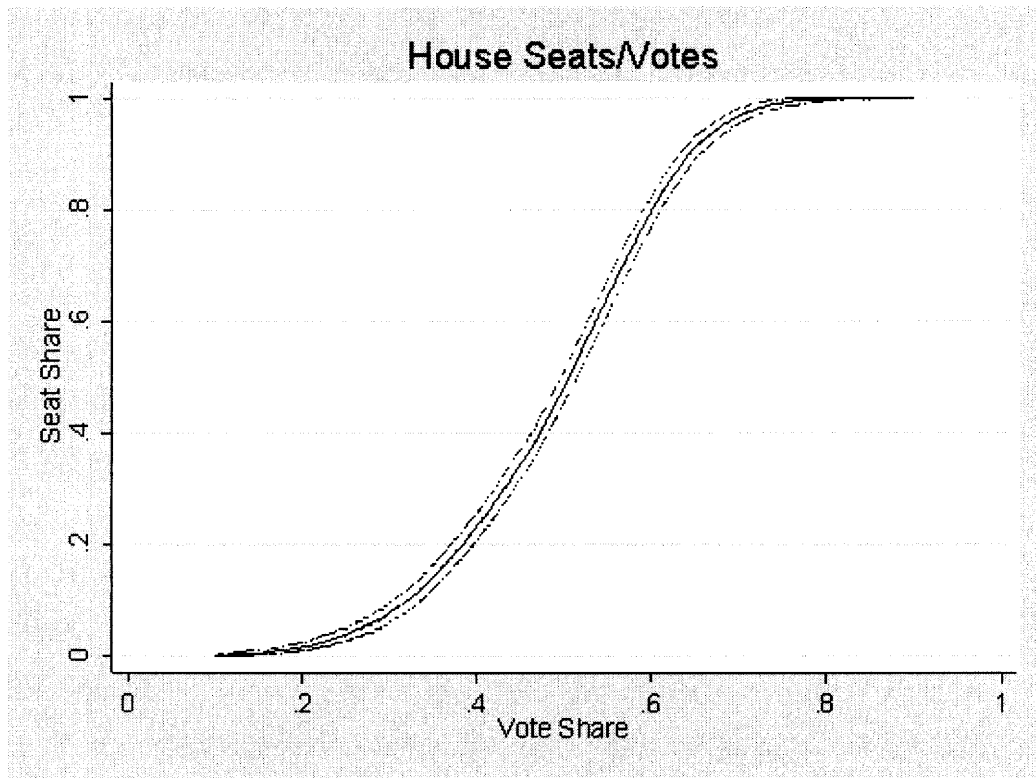


The seat share at vote share of 50 is 0.48 (95% 1000-iteration bootstrapped confidence interval 0.45,0.51). There is slight, but not statistically significant, evidence of a conservative "bias." The presence of a Republican bias is also found in (Kastellec, Gelman, and Chandler 2007).

An alternative way to look at bias is to consider the vote share required for a seat share of 50%. By the point estimates of the seats/votes relationship, this happens between 50.5 and 50.6 vote share. Stated in this fashion the bias seems much less

severe. It is impossible to get exactly 50% seat share in the model, because there are an odd number of seats. Democrats receive half plus one seats at 49.7%; half minus one seats at 49.6%. By the upper bound curve this happens between 49.6 and 49.7%. By the lower bound curve the median district lies between 51.4 and 51.5%. More compactly, I will write this estimate and measure of uncertainty as 50.55 (49.6, 51.5). I will write future descriptions of the district distribution in a similar manner.

According to the party position model in chapter 2, parties should position themselves approximately at the quartiles, if the distribution of districts is approximately symmetric. The first quartile will be the vote share corresponding to a 25% seat share; the third quartile is described by the vote share corresponding to a 75% seat share. The first quartile is at 0.4085 (0.395,0.421) and the third quartile is at 0.5845 (0.573,0.594). Note that one can NOT make any claims at this point about which party is further from the center in absolute terms. These are ordinal, rank scores only. Cardinal versions will be considered in section 4.4 below. Descriptive statements such as "approximately 41% of the public is more liberal than the Democratic party platform" might be valid. However, having the parties at the quartiles assumes the quartiles are equally distant from each other. Whether this is the case is considered in section 5.4 below.

Now consider the Senate. The following graph shows point estimates of state median ideologies.

The following graph is the hypothetical seats/votes curve for the Senate. The curve also assumes that all Senate seats are up for reelection simultaneously.

The Democratic seat share at a vote share of 0.5 is 0.36 (0.3, 0.44) - an apparently massive "bias" in favor of Republicans. However, the range of vote shares that yield a seat share of 0.5 (the median district) is 0.519-0.521 (0.507, 0.531). This is a case where ambiguity about how to define bias can cause confusion. By one measure bias looks amazingly large; by another, not so large.

The senate first quartile district is at 0.4695(0.451,0.488). The third quartile district is at 0.5565(0.543,0.570). According to the party position model, the Senate will be overall more conservative (median is further right) than both the House and

Senate Seats/Votes
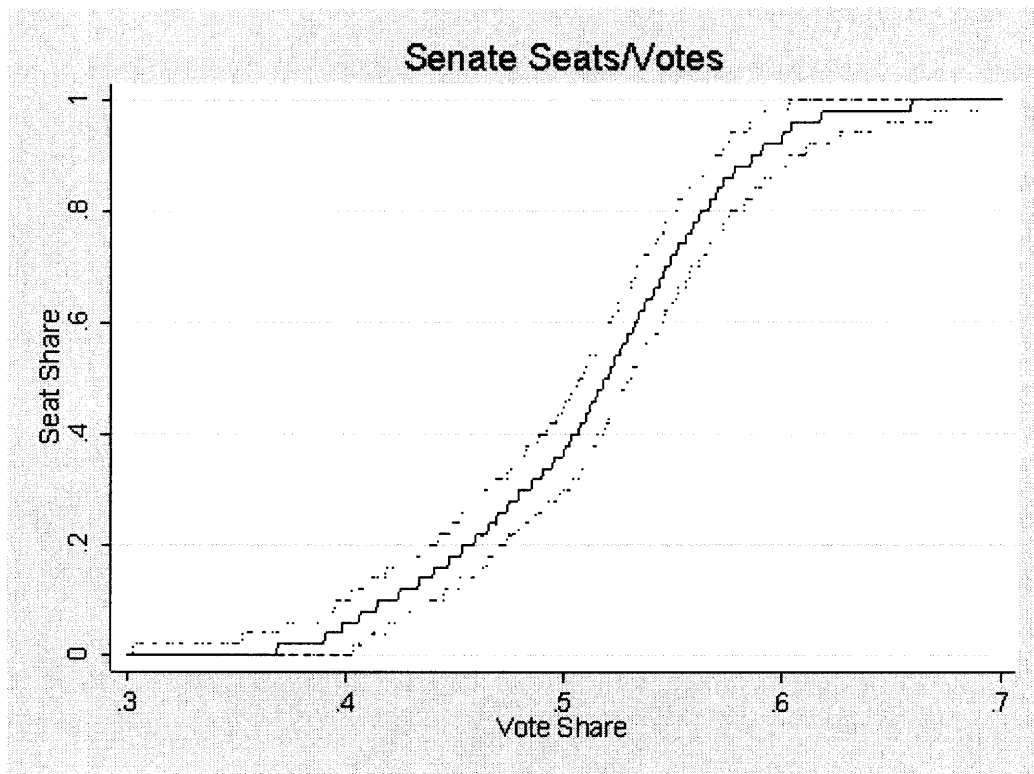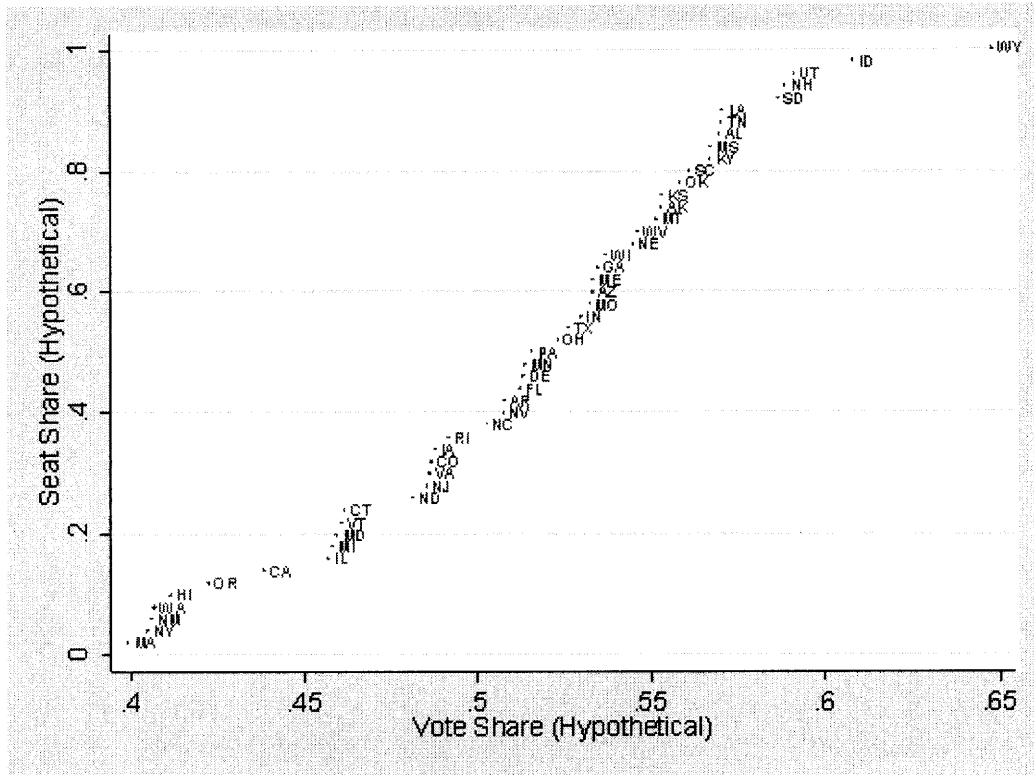
the general population. Furthermore, the parties should be closer together (quartiles are closer together). Overall according to the point estimates, House Democrats should be furthest left, followed by Senate Democrats, followed by the median voter, followed by Senate Republicans, followed by House Republicans.

## 5.3   Ideology and Vote Choice

If ideology is correlated with vote choice, and principal factor rank scores are correlated with ideology, then principal factor scores should be correlated with vote choice. And indeed they are. Let the variable rep_over_dem be equal to one if someone reported as voting Republican, zero if they voted Democratic in the 2006 house election. Then rep_over_dem and principal factor score rank have a Pearson correlation coefficient of 0.767.

Another approach to studying the relationship between ideology and voting is to consider a particular voting model. The following model is a modified special case of the model used in chapter 2.

Let party platforms be $x$ and $y$, and a voter's ideal point be $z$. Let voter utility then be $-(x-z_i)^2+(y-z_i)^2+v+v_i$, where $v$ is a global valence shift, and $v_i$ is an individual, idiosyncratic valence term. Expanding, we have $(y^2 - x^2 + v) + (2x - 2y) * z_i + v_i$. That is, utility is linear in ideal point. If $v_i$ is drawn from a standard normal, then voter behavior will relate to ideology according to a probit model.

The estimated probit coefficient of rep_over_dem on principal factor rank is 4.76 (0.47), with a pseudo R2 of 0.536.

There are a few problems with the above model, however. First rank scores only contain ordinal information – there is no a priori reason why they should have a simple functional relationship to vote choice. Even if voter utility was quadratic in some ideology metric, the rank scores need not be that metric.

A quick and dirty nonparametric way to estimate the relationship between votes and rank scores is to relate a moving average of vote choices (0=democratic, 1=republican) with respect to principal factor rank. The graph below was generated by

116

averaging over the 401 nearest values of vote (200 below, 200 above, and the value itself).
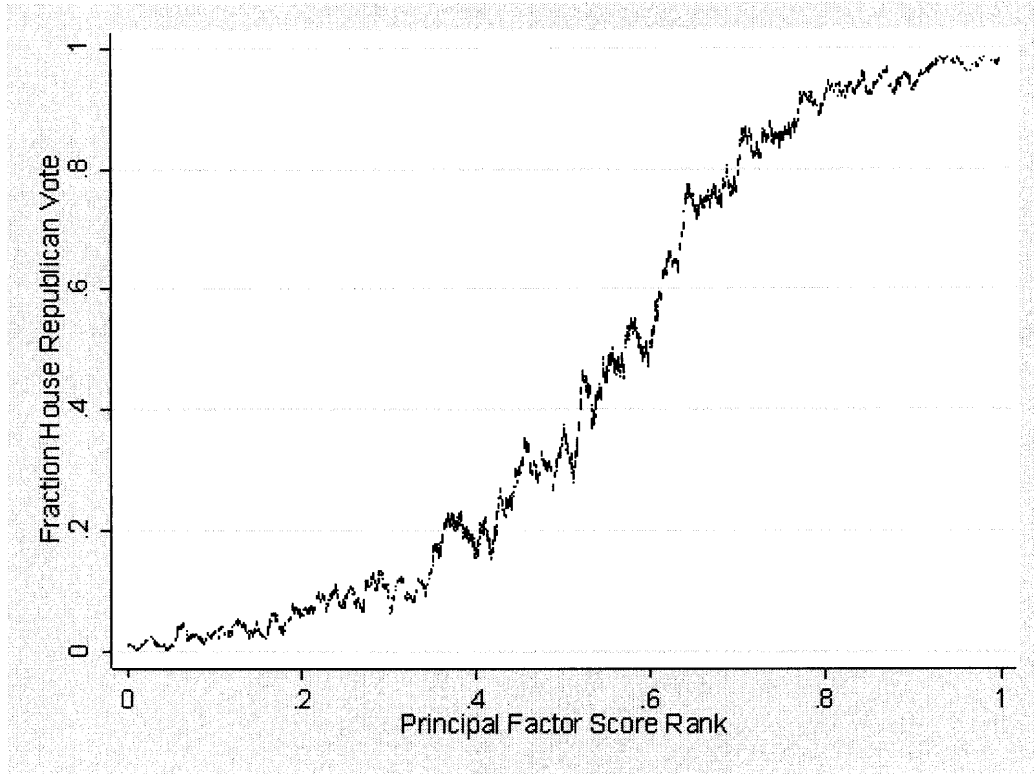


Here we see two deviations from the probit model – there is a small jump at around median rank, and the slope of the underlying xb seems to change across 0.5.

As far as the jump is concerned, up to this point I have been assuming a pure ideological voting model. In particular, incumbency and party ID have been omitted from the analysis. It is quite plausible this omission could cause the jump at 0.5. Consider an extreme example. Suppose everyone left of center was a Democrat and everyone right of center was a Republican. And suppose further that there was an independent fixed utility to voting with one's party. Then if one related ideology to vote choice, without controlling for party, then one would observe a sudden jump at 0.5, even if the direct relationship between ideology and vote choice is smooth.

Next, consider the asymmetric relationship between ideology and vote choice. By construction, rank scores are uniform (and thus symmetric) across the country as a whole. This is an artificial restriction. One of the central arguments of the previous

chapter was the national distribution of ideology - in terms of cut points - need not be uniform or symmetric. This is the "F" function in chapter 3. If we constrain utility functions to be symmetric, then the underlying distribution of voters could in principal be anything.

A solution to these two problems is to explicitly control for missing variables, and allow an asymmetric relationship. The following table allows for asymmetric effect of ideology, but does not include controls. The score variable is broken up into two parts. One part is equal to the score if the score is less than 0.5, and is equal to zero otherwise. The other part is equal to the score if the score is greater than 0.5, and is equal to zero otherwise. A dummy variable is also included that is equal to zero for scores less than 0.5 and one otherwise.

No controls: (24349 observations, pseudo R2 = 0.54)

| Variable | Coefficient | Standard Error |
|----------|-------------|----------------|
| Score, if under 0.5 | 3.65 | 0.13 |
| Score, if over 0.5 | 5.25 | 0.12 |
| Dummy, over 0.5 | -0.65 | 0.10 |
| Constant | -2.28 | 0.04 |

If the probability curve is smooth, then one half times the first coefficient should equal one half times the second coefficient plus the dummy coefficient. Otherwise, there will be a jump in probability of voting as the ideology score passes from just under 0.5 to just over 0.5. Without controls, the test for this linear restriction fails; the probability of seeing such a large deviation under the null of no jump is 0.0004.

The next table lists probit regression coefficients, including controls for party ID and incumbency. The omitted party ID category is independents, with no party lean. The incumbency variable is coded 1 if Democratic incumbent, -1 if Republican incumbent, 0 otherwise.

With controls: (24084 observations, pseudo R2 = 0.63)

With the controls, the linear restriction of half score under 0.5 = half score over 0.5 + dummy can not be rejected at 95% confidence level. (Probability of at least a large deviation under the null is 0.14). The data is at least somewhat consistent with

118

| Variable | Coefficient | Standard Error |
|---|---|---|
| Score, if under 0.5 | 3.01 | 0.14 |
| Score, if over 0.5 | 4.09 | 0.14 |
| Dummy, over 0.5 | -0.61 | 0.11 |
| Constant | -1.97 | 0.06 |
| Incumbency | -0.58 | 0.01 |
| Strong Democrat | -0.69 | 0.05 |
| Weak Democrat | -0.31 | 0.05 |
| Lean Democrat | -0.45 | 0.05 |
| Lean Republican | 0.56 | 0.05 |
| Weak Republican | 0.53 | 0.05 |
| Strong Republican | 0.88 | 0.05 |

no discontinuous jump at 0.5.

The above probit regressions show that a given shift in rank score has more utility impact (under the symmetric, quadratic utility with normally distributed idiosyncratic error) on vote choice if the subject is right of center than left of center. One way to deal with this result would be to generate new ideology scores via the formulas new score = 3.01*old score if old score less than 0.5 or new score = 4.09*old score -0.61 if the old score is greater than 0.5. These new scores would then be directly interpretable in utility terms. The national distribution of these scores would be piecewise uniform, with a high density from 0 to around 1.5 and a low density from around 1.5 to 4.1. (The precise break value is ambiguous because of the slight discontinuous jump at rank score 0.5. If we take the coefficients as given the probability of voting Republican is not increasing monotonically.)

One quirk of the above method, however, is that there is a different distribution of ideology scores for each office. For Senate races, the probit coefficients are as follows:

With controls: (20713 observations, pseudo R2 = 0.69)

While the qualitative results remain the same, the exact impact of low rank scores and high scores is slightly different. In particular, conservatives tend to be more sensitive to ideology in their Senate choices instead of House choices.

| Variable | Coefficient | Standard Error |
|---|---|---|
| Score, if under 0.5 | 2.98 | 0.17 |
| Score, if over 0.5 | 4.69 | 0.17 |
| Dummy, over 0.5 | -0.86 | 0.13 |
| Constant | -1.78 | 0.07 |
| Incumbency | -0.286 | 0.01 |
| Strong Democrat | -0.93 | 0.06 |
| Weak Democrat | -0.52 | 0.05 |
| Lean Democrat | -0.80 | 0.06 |
| Lean Republican | 0.64 | 0.06 |
| Weak Republican | 0.45 | 0.05 |
| Strong Republican | 1.01 | 0.06 |

## 5.4   Distribution of Districts

In section 5.3, we showed that in a quadratic utility world, with normal idiosyncratic errors, ideology should influence probit scores linearly, when ideology is in the correct units. Principal factor rank scores however influence probit scores non-linearly. In particular, scores right of center have more impact on vote choice than scores left of center. One way to view this is utility is asymmetric, and the distribution of scores is (by construction) symmetric. However, the exact same voting behavior can be described by symmetric utility and an asymmetric distribution of ideology. Since the quartile voter theorem from chapter 2 requires symmetric utility, the natural solution is to transform ideology scores. The coefficient on scores right of center is 4.09; the coefficient of scores left of center is 3.01. Therefore, if we multiply all scores right of center by 4.09/3.01, the new transformed scores will correspond to symmetric, quadratic utility.

The third quartile house district is at rank score 0.5845. Therefore, its transformed score is 0.5 + 0.0845(4.09/3.01) = 0.6148. The first quartile score (and transformed score) remains 0.4085. The median district is near 0.5055, so its transformed score is 0.5073. The Democratic party is therefore 0.5073 - 0.4085 = 0.988 utility units from the median voter of the median district. The Republican party is 0.6148 - 0.5073= 0.1075 utility units away. The point estimates indicate that the Republican party is

120

slightly DISADVANTAGED in terms of seats in the long run[1], when we consider the party position logic, even though the seats/votes curve is slightly "biased" in "favor" of Republicans. Because of the distribution of districts, Republicans represent overly conservative districts, which should drag their platform to the right. As discussed in chapter 3, seats/votes bias at most implies that the parties aren't representative - the median voter in the median district is more conservative than the median voter in the general population. If parties compete for the former and not the latter, then parties will not represent the general population. Furthermore, the left quartile is closer to the median voter in the general population than the right quartile. Therefore, according to the party position model, Democrats should in the long run tend to get more than 50% of the votes and more than 50% of the seats, even though the seats/votes curve will seem biased against Democrats. In other words, when Democrats get 50% of the vote, they'll get less than 50% of the seats... but in the long run Democrats will be expected to get BOTH more than 50% of the vote AND more than 50% of the seats.

Naturally, these estimates are all extremely approximate, due to sampling error, model approximation error, and so forth. One way to interpret the estimates is as a defense of a global null hypothesis: the system is not biased toward either party. There is some evidence of seats/votes bias in favor of Republicans - although the large confidence intervals do not reject the null hypothesis of no bias. But the estimates combined with the party position theory point to a structural long-run bias in favor of Democrats. By one criterion Republicans might have an advantage; by another criterion Democrats might have an advantage. There is reason to be wary of rejecting the natural null that things are overall fair and balanced. That is, perhaps neither party has an advantage.

For the Senate the ratio of high to low score coefficients is 4.69/2.98 = 1.57. Therefore, the first quartile remains at 0.4695, the median, using an initial value of 0.52, is now 0.5314, and the third quartile is at 0.5887. Democrats are therefore 0.5314 -0.4695 = 0.0619 units from the median district, while Republicans are 0.5887-

---

[1]In the short run, Republicans might have had an advantage due to incumbency and the short-run moderating impact of being a majority on party platforms. See chapter 2.

0.5314 = 0.0573. Republicans therefore may have a (very) slight advantage in terms of predicted party platforms. However, the approximate long-run Democratic platform (disregarding the prediction that Democratic platform should be slightly left of the first quartile) should be closer to the median voter in the general population. In other words, according to the party position model, Democrats should over the long-run consistently get decidedly more than 50% of the aggregate vote share, and slightly less than 50% of the seat share.

## 5.5   Corrected Seats/Votes

The seats votes curves given in section 5.2 are appropriate (with the modifications in section 5.4) for application to the party position model. However, they are inappropriate for application to actual seats/votes curves. They fail to account for party ID, incumbency, and idiosyncratic vote choice.

To correct for this we need to use transformed scores, and account for the fact that someone left of an election cut-point does not vote for the left party 100% of the time. The procedure used is as follows:

1. Take the linear predicted scores from the probit regression with controls described in section 5.3, $s_i$. The dependent variable differs when estimating Senate and House elections; therefore the scores will be different.

2. Pick a global valence shock,$dU$.

3. For each person, find the probability of voting Republican, $p_i$, which is equal to the cumulative normal of $s_i + dU$

4. For the country as a whole, find average $p_i$. (Weighted by sample weights) This is the vote share at $dU$.

5. For each district, find average $p_i$ within the district. (Again, weighted by sample weights) Each district with share over 0.5 adds one to the Republican seat total

6. Graph vote versus valence, seats versus valence, and seats versus votes.

The following graph represents predicted Republican House seats (solid line) and

Republican vote share (dotted line) as a function of valence. A zero valence indicates a zero valence shift relative to the 2006 election. A zero valence does NOT represent a hypothetical election where voters had no collective attraction to either party.
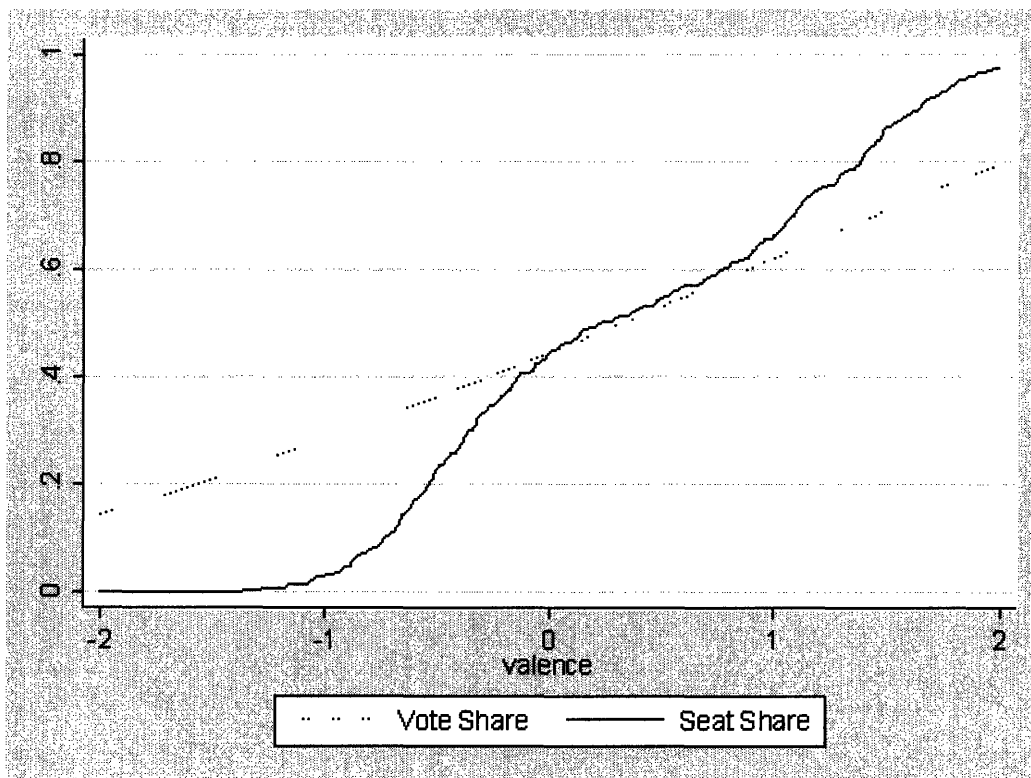


One immediate observation is the relationship between votes and valance is roughly linear. This lends support for the view that vote share is a good proxy for aggregate valence. Such a view is implicit in some interpretations of seats/votes responsiveness. For example, Tufte (1973) interprets a high seats/votes responsiveness as implying a high responsiveness of electoral outcomes to economic performance. However, the linearity is partly an artifact of the probit model, which is approximately linear for probabilities near 50%. To fully establish this linearity other voting models would have to be checked. Doing so is left to future work.

Nevertheless, in principal despite the use of probit model there could of been nonlinearities for certain distributions of voters. For example, consider voters who leaned fairly strongly (but not too strongly) toward the Democratic Party. Then for those voters positive aggregate valence shocks for the Democratic Party would have

little impact (from nearly 100% support to at most 100% support) while negative shocks would have more impact (from nearly 100% support to less than 100%). If there were significantly more of these types of voters than corresponding conservative Republican voters, than valence shocks toward Democrats would have less impact on vote share than valence shocks toward Republicans.

Another observation from the graph is that the seats/valence curve comes roughly in three regions - a flat region in the middle, with sharp highly responsive regions at the ends. This can also be seen in the seats/votes relation:
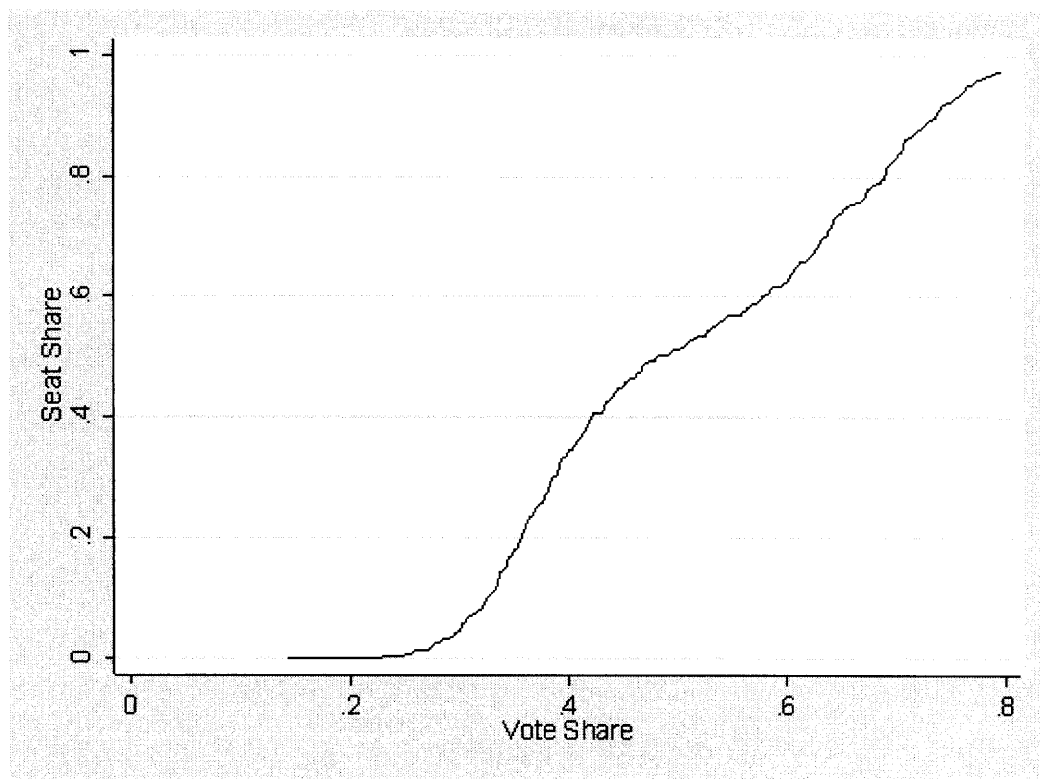


This three-region responsiveness makes sense if we consider the impact of incumbency. For normal election years, incumbency will act as a buffer, reducing seats responsiveness to vote shares. But in a sufficiently severe landslide, the incumbency effect would be overwhelmed, causing a sudden and drastic loss of seats.

The non-convexity of the seats/votes curve is particularly striking if we consider that if the seats/votes curve were logistic, the responsiveness should be highest in the center, and lowest in the tails. Instead, responsiveness is low, than high, than low,

than high, than low again.

What is especially interesting about the model's prediction, is that the Democrats came very close to, but did not quite reach, the highly responsive region. If it had been a slightly more Democratic year, according to the model and survey, Democrats would have received significantly more seats per additional vote (or valence) shift.

Another important thing to check is whether the graph corresponds to the actual seats/votes value of the 2006 election. At zero valence, the Republican seat share should be 44.6% and the vote share 44.5%. In fact the seat share was 46.4% and the vote share 46.7% (Kastellec, Gelman, and Chandler 2007).

In terms of bias, at 50% vote, the model predicts Republicans would win 223 seats, or 51.3%. For the Republicans to win 50% of the seats, they would need to win approximately 48.1% of the vote. Kastellec, Gelman, and Chandler (2007) using district level data also find a 48% vote required for an expected 50% seat share, although their measure of vote is average district vote, not average vote.

The graphs that follow refer to the Senate. The first presents the fraction of contested senate seats (Connecticut omitted) won and fraction of vote share won as a function of global ideology shift. Again, a valence of 0 represents the aggregate valence of the 2006 election. It does NOT represent a hypothetical election where each party was equally attractive in valence terms. The second displays the relation between seats and votes.

valence

Vote Share

State Share

Seat Share

Vote Share

126
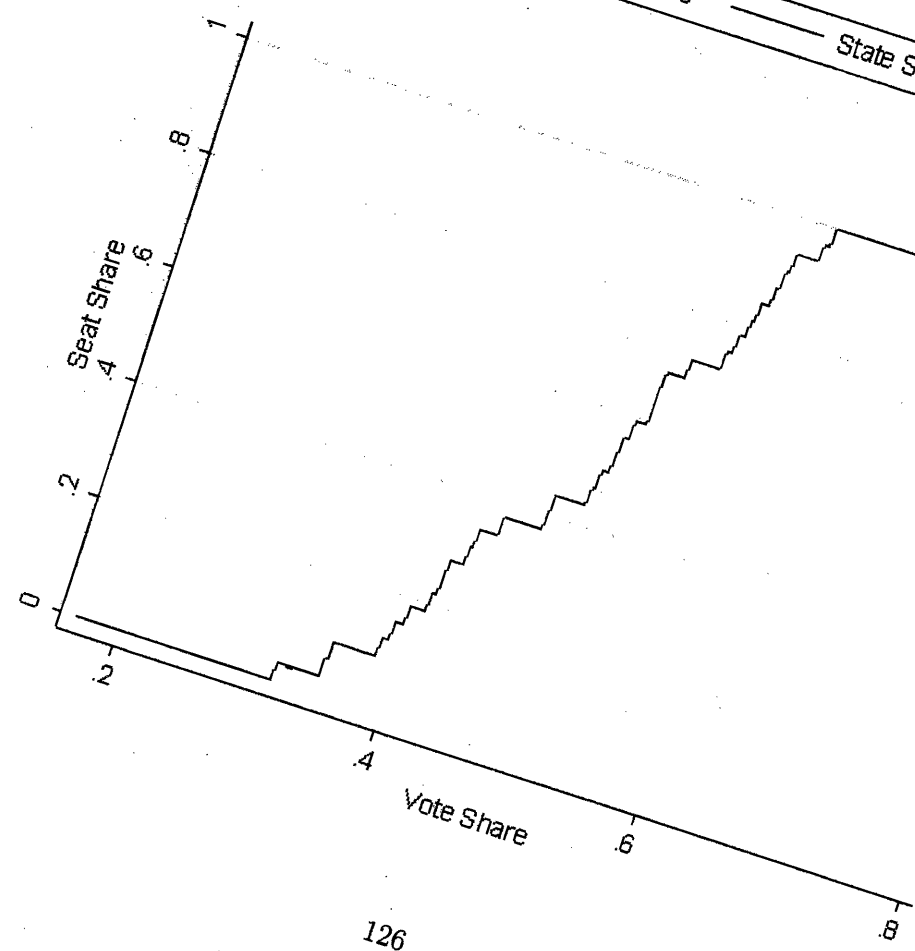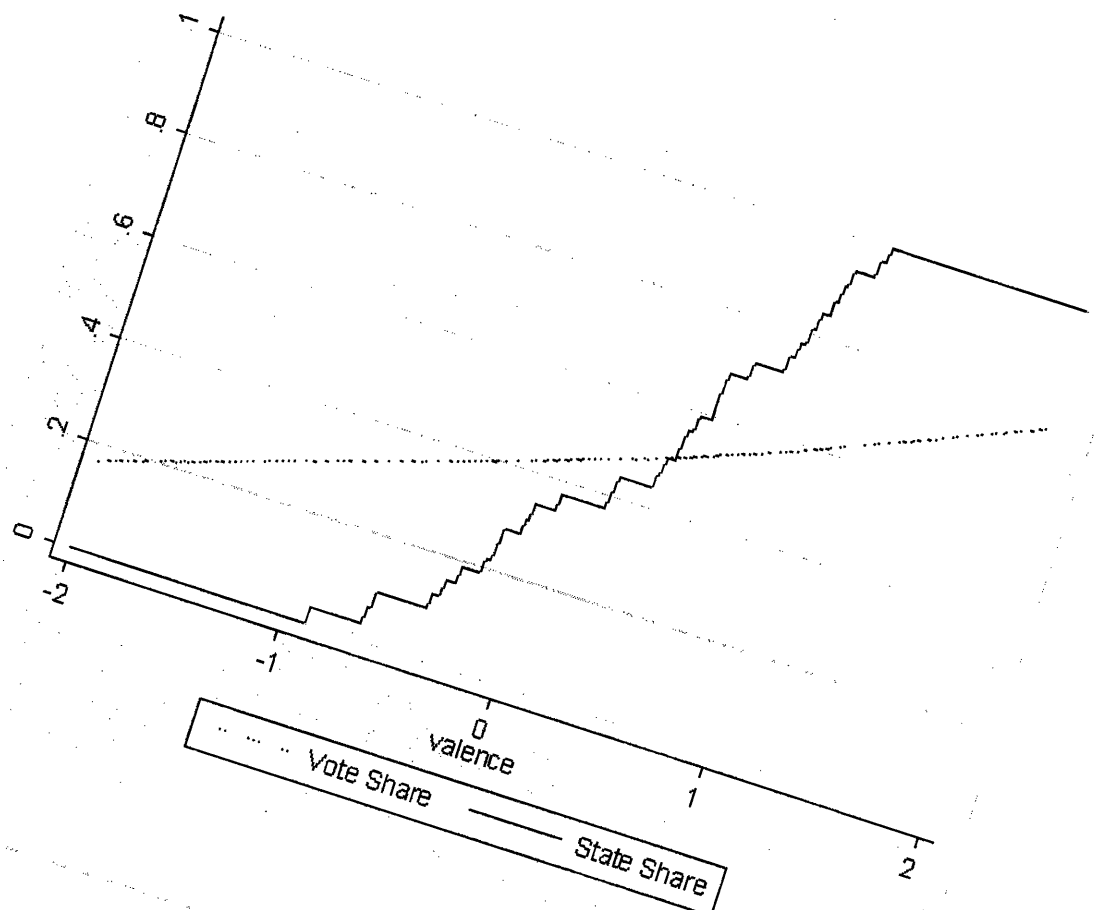
Determining bias for the Senate is a bit tricky. One approach is to directly examine point estimates corresponding to vote shares near 50%:

| Valence | Vote | Seats |
|---------|-------|-------|
| 0.41 | 0.498 | 15 |
| 0.42 | 0.499 | 15 |
| 0.43 | 0.501 | 16 |
| 0.44 | 0.502 | 16 |
| 0.45 | 0.504 | 16 |
| 0.46 | 0.505 | 16 |

Since there are 32 seats in the analysis, 16 seats corresponds to a 50% share, 15 seats corresponds to a 47% share. There is either a 3% bias toward the Democrats, or no bias.

Another possibility is to try a linear fit through the seats/votes curve. One issue with this approach is that the linear model clearly fails outside of certain bounds, as seat share eventually becomes capped at 0% or 100%. The least squares minimum fit of seat shares on vote shares, given vote share between 0.4 and 0.6 yields a responsiveness coefficient of 3.72 with a constant of -1.315. Therefore, at vote share of 0.5, the seat share would be 3.72*0.5-1.315 = 0.54; a substantial pro-Republican bias. However, this result is sensitive to the chosen bounds. If one restricts attention to vote shares between 0.48 and 0.52, the coefficients become 5.24 and -2.12, with a seat share at vote 0.5 of 0.498. There is either no bias, or at most a slight Democratic bias. To make matters worse, there is another issue - the Senate seats/votes curve is not symmetric around 0.5. If we consider vote shares between 0.4 and 0.5, the responsiveness coefficient is 2.62; for vote shares from 0.5 to 0.6, the coefficient is 4.28.

At the end of the day there simply are not enough seats and too much non-linearity to make definite statements about seats/votes bias in the Senate.

There is some evidence for the high-low-high responsiveness pattern in the Senate, but not nearly as dramatic as the House. The responsiveness between 0.4 and 0.45 is 4.38; from 0.45 to 0.5 the responsiveness is 2.09; from 0.5 to 0.55 it is 5.10; from 0.55

to 0.60 it is 5.69. As is the case in the House, responsiveness is low for vote shares near 0.5.

## 5.6 Conclusion

Let us review the findings of this chapter. The predicted seats/votes curve of both the House and Senate are consistently biased "against" Democrats. The model implies therefore implies that the median voter of the median district will be right of the median voter in the general population. However, for the House at least, the bias does not pass conventional tests for statistical significance. Furthermore Democrats should not only be electorally competitive in the House, but may even have an advantage. To win 50% of the seats Democrats need more than 50% of the vote, but they should in the long run receive more than 50% seats anyway.

Overall, there is little evidence of systematic distortions, with the possible exception of the Senate districts being overly conservative relative to the country as a whole. However, this contrasts with the hypothetical case of party-list system in the US. According to the distribution of voters and the party position model, there would be a massive advantage to the left of center party if districts were eliminated. Of course, this hypothetical assumes that only two parties would continue to exist. Nevertheless, the hypothetical illustrates how balanced the status quo is.

This does not imply that the shape of districts is having no impact, however. In particular, the model predicts based on the distribution of districts that the parties in the Senate should be closer together than the house (and MUCH more converged than the hypothetical party list system.) Besides the direct implications for representation, the model also implies that incumbents should be more vulnerable to national tides in the Senate. When platforms are close together, voters can not differentiate between the parties as well on ideology, and thus put more effective weight on valence shocks.

# Chapter 6

# Concluding Remarks

This thesis has explored some implications of a one dimensional voting model for parties (with the added assumption of self-interested incumbents) and seats/votes curves. Chapters 2 and 3 set up a broad research agenda, with chapters 4 and 5 only exploring a small slice of that agenda.

There are many implications from the models of chapters 2 and 3 that have not yet been explored. Past good performance should result in moderation of party platforms and, consequently, improved future performance. The well-known phenomenon of mid-term seat-loss by presidents is consistent with this predicted outcome. That is, suppose a party wins the presidency, and a large number of seats in congress in one election. According to the model, they should lose seats in the next election, but NOT lose all the seats they won with the presidential coattails. According to the model this should happen, on average, even with no incumbency effect. This is a prediction that would be interesting to check. Even better would be to find a measure of party platforms, and check whether a winning party moderates, and a losing party moves to the extreme.

Chapter 3 considers a variety of ways to find the distribution of districts using the seats/votes relationship. Chapter 4 took the approach of directly measuring the distribution of voter ideology. However, the applicability of that approach depended on large part of the availability of the large dataset of the CCES. To examine seats/votes curves in other venues may require the use of cruder measures of ideology. Step one

is to replicate chapter 5 using weighted-uniform-swing technique from chapter 3. If the results are reasonably similar, one can then try weighted uniform swing technique for past years using rough ideology data.

Another approach suggested in chapter 3 is to use demographic data to approximate the distribution of ideology. Again, step one of such a procedure would be to use, for example, 2000 census demographic data on the 2006 election and see if the ideology results closely correspond to full information case. If they do, then one can proceed with using other demographic data on other elections.

One issue that could use additional exploration is the consequence of using voter-level models versus district-level models of seats/votes curve and incumbency. The method used in chapter 5 is a voter-level model. This contrasts with using district level information, such as Gelman and King (1994), to generate seats/votes curves. Do the two approaches yield similar results? Or are there aggregation issues that corrupt district level analysis? Or district-level effects that can not be measured by looking at individuals that corrupt individual-level analysis?

Just as the number of moderates can influence how national tides impact each individual district, the number of moderates can also influence the incumbency effect. As discussed in chapter 3, if incumbency does not directly impact votes, but instead influences voters utility preferences, then the more marginal voters in a district, the larger the incumbency effect. Measuring the incumbency effect taking this into account may help generate more efficient estimates.

There are uses of detailed information on voter ideologies that are not directly tied to the theory of chapters 2 and 3. For example, one could check whether incumbent voting behavior is correlated with district ideology. Preliminary results indicate that congressional voting behavior from 2000-2006 is indeed correlated with district median scores calculated from the 2006 CCES. Furthermore, with highly detailed descriptions of district ideology we can test more subtle relationships. For example, consider the variance of ideologies within a state. In a highly polarized state, one might expect a higher difference in voting behavior between senators. Preliminary results show this is indeed the case using the 2006 data. Another possible thing to explore is incumbency

and variance. In addition to the number of moderates impact on the incumbency effect, having a highly polarized district might impact incumbents' ability to build a long-run constituency (Fenno 1978).

In addition to exploring the research agenda implied by the theory of chapters 2 and 3, there is substantial room for improvement to the theory itself. For example, the model of chapter 2 assumes myopic incumbents. A natural extension tot he model would be to consider the strategic implications of a longer time horizon. Another issue is voter behavior. While the strong assumption of pure ideological voting helps clarify the argument and the model, it is quite arguably unrealistic. It would be useful to check how robust the results remain if we allow different voter behaviors. For example, I have assumed that voter ideology is exogenous. What if instead some voters to some extent pick parties first, and then match their ideology to the party platforms. To what extent would reverse causation invalidate the predictions of the model? Another possible approach is to allow voters to have a pool of possible responses when asked a policy question, instead of a unique response, ala Zaller (1992). Should a voter with both conservative and liberal opinions be coded as a centrist? I suspect the model and predictions would be robust to some of these different paradigms of voter behavior, but further work is required to establish this.

# References

Aldrich, John H. 1983. "A Downsian Spatial Model with Party Activism." *American Political Science Review* 77: 974-990.

Aldrich, John H. 1995. *Why Parties?*. Chicago: University of Chicago Press.

Aldrich, John H., and Michael D. McGinnis. 1989. "A Model of Party Constraints on Optimal Candidate Positions." *Mathematical and Computer Modeling* 12: 437-450.

Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart, III. 2000. "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage." *American Journal of Political Science* 44: 17-34.

Ansolabehere, Stephen, James M. Snyder, Jr. and Charles Stewart, III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45: 136-159.

Aranson, Peter, and Peter Ordeshook. 1972. "Spatial Strategies for Sequential Elections." In *Probability Models of Collective Decision Making,* Edited by Richard Niemi and Herbert Weisberg. Columbus, OH: Charles E. Merrill.

Austen-Smith, David. 1984. "Two-Party Competition with Many Constituencies." *Mathematical Social Sciences* 7: 177-198.

Austen-Smith, David. 1986. "Legislative Coalitions and Electoral Equilibrium." *Public Choice* 50: 185-210.

Baron, David P. 1994. "Electoral Competition with Informed and Uninformed Voters." *American Political Science Review* 88: 33-47.

Bernaards, CA. and K Sijtsma 2000. "Influence of Imputation and EM Methods on Factor Analysis when Item Nonresponse in Questionaire." *Multivariate Behavioral Research* 35: 321-364.

Bernhardt, Daniel M., and Daniel E. Ingberman. 1985. "Candidate Reputations and the 'Incumbency Effect'." *Journal of Public Economics* 27: 47-67.

Besley, Timothy and Stephen Coate. 1997. "An Economic Model of Representative Democracy." *Quarterly Journal of Economics* 112: 85-114.

Brady, David. 1988. *Critical Elections and Congressional Policy Making.* Stanford: Standford University Press.

Butler, David and Donald Stokes. 1969 *Political Change in Britain: Forces Shaping Electoral Choice.* London: Macmillan.

Callander, Steven. 2005. "Electoral Competition in Heterogenous Districts." Journal of Political Economy 113: 1116-1145.

Calvert, Randall L. 1985. "Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence." *American Journal of Political Science* 29:69-95.

Calvert, Randall L., and R. Mark Isaac. 1981. "The Inherent Disadvantage of the Presidential Party in Midterm Congressional Elections." *Public Choice* 36: 141-146.

Chen, Song Xi. 1996. "Empirical likelihood confidence intervals for nonparametric density estimation." *Biometrika* 83: 329-341.

Cox, Gary W. 1990. "Centripetal and Centrifugal Incentives in Electoral Systems." *American Journal of Political Science* 34: 903-35.

Cox, Gary W. and Matthew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House.* Berkeley, CA: University of California Press.

Coate, Stephen and Brian Knight. 2005. "Socially Optimal Districting." *NBER Working Paper Series* Working Paper 11462.

Cowling, Ann and Peter Hall. 1996. "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation." *Journal of the Royal Statistical Society. Series B (Methodological)* 58: 551-563

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Edgeworth, Frances Y. 1898. "Miscellaneous Applications of the Calculus of Probabilities." *Journal of the Royal Statistical Society* 51: 534-44.

Epstein, David and Sharyn O'Halloran. 1999. "A Social Science Approach to Race, Redistricting, and Representation." *American Political Science Review* 93: 187-191.

Erikson, Robert S. and Gerald C. Wright. 2000 "Representation of Constituency Ideology in Congress," in David Brady, John F. Cogan and Morris P. Fiorina (eds.), Continuity and Change in U.S. House Elections. Stanford University Press, pp. 148-177.

Enelow, James M., and Melvin J. Hinich. 1981. "A New Approach to Voter Uncertainty in the Downsian Spatial Model." *American Journal of Political Science* 25: 483-493.

Fenno, Richard F., Jr. 1978. *Home style.* Boston: Little, Brown.

Fiorina, Morris. 1980. "The Decline of Collective Responsibility in American Politics." *Daedalus* 109: 25-45.

Gelman, Andrew and Gary King. 1994 "Enhancing Democracy Through Legislative Redistricting." *The American Political Science Review* 88: 541-559.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2005 "Rich state, poor state, red state, blue state: What's the matter with Connecticut?" Unpublished.

Gomberg, Andrei M., Francisco Marhuenda, and Ignacio Ortuno Ortin, 2004. "Equilibrium in a Model of Endogenous Political Party Formation." *Economic Theory* 24: 373-394.

Green, Donald, Eric Schickler, and Brad Palmquist. 2003. *Partisan Hearts and Minds.* New Haven, CT: Yale University Press.

Gudgin, Graham and Peter J. Taylor. 1974. "Electoral Bias and the Distribution of Party Voters." *Transactions of the Institute of British Geographers* 63: 53-74.

Guttman, Louis. 1944. "A Basis for Scaling Qualitative Data." *American Sociological Review* 9: 139-150.

Hall, P. 1991. "Egeworth Expansions for Nonparametric density Estimators, with Applications." *Statistics* 22: 215-32

Hotelling, Harold, 1929. "Stability in Competition." *Economic Journal* 29: 41-57.

Ingberman, Daniel E., and John Villani. 1993. "An Institutional Theory of Divided Government and Party Polarization." *American Journal of Political Science* 37:429-471.

Jacobson, Gary. 1987. "The Marginals Never Vanished: Incumbency and Competition in Elections to the House of Representatives, 1952-82." *American Journal of Political Science* 31: 126-141.

Kendall, M.G. and A. Stuart. 1950. "The Law of Cubic Proportion in Election Results." *British Journal of Sociology* 1: 183-197.

Keith, B., D. Magleby, C. Nelson, E. Orr, M. Westyle, and R. Wolfinger. 1992. *The Myth of the Independent Voter.* Berkeley: University of California Press.

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

Krehbiel, Keith. 2000. "Party Discipline and Measures of Partisanship." *American Journal of Political Science* 44: 212-227.

Londregan, John, and Thomas Romer. 1993. "Polarization, Incumbency, and the Personal Vote." In *Political Economy: Institutions, Information, Competition, and Representation*, edited by W.A. Barnett, M.J. Hinich, and N.J. Schofield. Cambridge: Cambridge University Press.

March, James G. 1957-1958. "Party Legislative Representation as a Function of Election Results." *Public Opinion Quarterly* 21: 521-542.

Mayhew, David. 1974. *Congress: The Electoral Connection* New Haven: Yale University Press.

Mokken, RJ. 1971. *A Theory and Procedure of Scale Analysis: With Applications in Political Research.* Mouton: The Hague

Mokken, RJ and C Lewis. 1982. "A nonparametric approach to the analysis of dichotomous item responses." *Applied Psychological Measurement* 6: 417-430.

Palfrey, Thomas R. 1984. "Spatial Equilibrium with Entry." *Review of Economic Studies* 51: 139-156.

Poutvaara, Panu. 2003. "Party Platforms with Endogenous Party Membership." *Public Choice* 117: 79-98.

Robertson, David. 1976. *A Theory of Party Competition* London: Wiley and Sons.

Roemer, John E. 1998. "Why the Poor Do Not Expropriate the Rich: An Old Argument in New Garb." *Journal of Public Economics* 70: 399-424.

Roemer, John E. 2002. *Political Competition: Theory and Applications.* Harvard University Press.

Schofield, Norman. 2004. "Local Political Equilibria." In *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks*, edited by David Austen-Smith and John Duggan. Heidelberg: Springer.

Serra, Gilles. 2005. "Primary Elections in Two-Party Systems: The Effects of Party Nominations on Public Policy." Unpublished manuscript, Harvard University.

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

Shotts, Kenneth W. 2002. "Gerrymandering, Legislative Composition, and National Policy Outcomes." *American Journal of Political Science* 456:398-414.

Sijtsma, Klaas. 2002. *Introduction to nonparametric item response theory.* Thousand Oaks, CA: SAGE Publications.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* New York: Chapman and Hall.

Snyder, James M., Jr. 1994. "Safe Seats, Marginal Seats, and Party Platforms: The Logic of Party Platform Differentiation." *Economics and Politics* 6: 201-213.

Snyder, James M., Jr., and Michael M. Ting. 2002. "An Informational Rationale for Political Parties." *American Journal of Political Science* 46: 90-110.

Tufte, Edward R. 1973. "The Relationship between Seats and Votes in Two-Party Systems." *The American Political Science Review* 67: 540-554

Ward, Hugh. 2000. "'If the Party Won't Go to the Median Voter, Then the Median Voter Must Come to the Party': A Spatial Model of Two-Party Competition with Endogenous Voter Preferences." Paper for the Political Studies Association-UK 50th Annual Conference.

Wittman, Donald. 1977. "Candidates with Policy Preferences: A Dynamic Model." *Journal of Economic Theory* 15: 86-103.

Wittman, Donald. 1983. "Candidate Motivation: A Synthesis of Alternatives." *American Political Science Review* 77: 142-157.

*The Yale Law Journal.* 1963. "Baker V. Carr and Legislative Apportionments: A Problem of Standards" (in Notes and Comments) 72: 968-1040.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion.* Cambridge University Press.