Measuring Legislative Preferences

Nolan McCarty

February 26, 2010

1 Introduction

Innovation in the estimation of spatial models of roll call voting has been one of the most important developments in the study of Congress and other legislative and judicial institutions. The seminal contributions of Poole and Rosenthal (1991, 1997) launched a massive literature marked by sustained methodological innovation and new applications. Alternative estimators of ideal points have been developed by Heckman and Snyder (1997), Londregan (2000a), Martin and Quinn (2002), Clinton, Jackman, and Rivers (2004), and Poole (2000). The scope of application has expanded greatly from the original work on the U.S. Congress. Spatial mappings and ideal points have now been estimated for all fifty state legislatures (Wright and Schaffner 2002, Shor and McCarty 2010), the Supreme Court (e.g. Martin and Quinn; Bailey and Chang 2001, Bailey 2007), U.S. presidents (e.g. McCarty and Poole 1995; Bailey and Chang 2001, Bailey 2007), a large number of non-U.S. legislatures (e.g. Londregan 2000b, Morgenstern 2004), the European Parliament (e.g. Hix, Noury, and Roland 2006), and the U.N. General Assembly (Voeten 2000).

The popularity of ideal point estimation results in large part from its very close link to theoretical work on legislative politics and collective decision making. Many of the models and paradigms of contemporary legislative decision making are based on spatial representations of preferences.[1] Con-

[1] A non-exhaustive sampling of a vast literature includes Gilligan and Krehbiel (1987), Krehbiel (1998), Cameron (2000), and Cox and McCubbins (2005) .

sequently, ideal point estimates are key ingredients for much of the empirical work on legislatures, and increasingly on courts and executives.[2] This has contributed to a much tighter link between theory and empirics in these subfields of political science.[3]

The goal of this essay is to provide a general, less technical overview of the literature on ideal point estimation. So attention is paid to the concerns of the end-user; the empirical researcher who wishes to use ideal point estimates in applications. In order to highlight the advantages and disadvantages of using ideal point estimates in applied work, I make explicit comparisons to the primary alternative: interest group ratings. My main argument is that the choice of legislative preference measures often involves substantial trade-offs that hinge on seemingly subtle modeling choices and identification strategies. While these variations may or may not affect the results of particular applications, it is very important for the applied researcher to understand the nature of these trade-offs to avoid incorrect inferences and interpretations.

2 The Spatial Model

Although there is a longer tradition of using factor or cluster analysis to extract ideological or position scales from roll call voting data, I concentrate

[2]A sample of such work includes Cox and McCubbins (1993), McCarty and Poole (1995), Clinton (2007), Cameron (2000), Clinton and Meirowitz (2003a,b).

[3]This is not to say that there is no slippage between statistical and theoretical spatial models. I return to the issue of the congruence between empirical and theoretical work below.

exclusively on those models that are generated explicitly from the spatial model of voting.[4] The spatial model assumes that policy alternatives can be represented as points in a geometric space – a line, plane, or hyperplane.[5] Legislators have preferences defined over these alternatives.[6] In almost all of the statistical applications of the spatial model, researchers assume these preferences satisfy two properties:

- Single-peakedness: When alternatives are arranged spatially, the legislator cannot have two policies that they rank higher than all adjacent alternatives. In other words, for all policies but one, there is a nearby policy that is better. Consequently, the legislator's most preferred outcome is a single point. We call this point the legislator's *ideal point.*

- Symmetry: If x and y are alternatives represented by two points equal distance from a legislator's ideal point, the legislator is indifferent between the two.

To make these ideas concrete, consider Figure 1 that introduces a simple motivating example that I use throughout the essay. The example assumes

[4] My scope is limited both for space reasons and a desire to focus on the link between empirical and theoretical work. See Poole (2005 p. 8-11) for a discussion of this earlier work.

[5] For a slightly more technical introduction to the spatial model, see McCarty and Meirowitz (2006; pp. 21-24).

[6] I refer to the voter throughout the essay as a legislator even though ideal points of executives, judges, and regulators have also been estimated using these techniques.

that policies can be represented as points on a single line and that Senators Russell Feingold, Olympia Snowe, and Tom Coburn have ideal points.

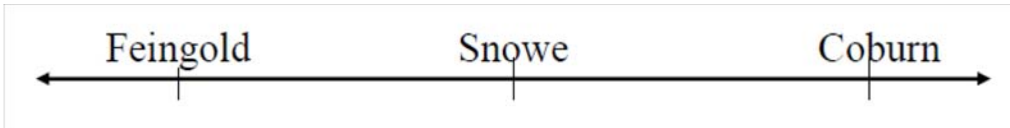

Figure 1: Ideal Points of Three Senators

Figure 2 places two voting alternatives, *yea* and *nay,* on the line along with the ideal points of the senators. Under the assumption that preferences are symmetric, the model predicts that in any binary comparison each senator prefers the policy closest to his or her ideal point. Given the simple pairwise comparison of *yea* and *nay,* it seems natural to assume that each senator would vote for the closest outcome. This assumption is known as *sincere voting.*[7] Clearly, Feingold is closer to *yea* than to *nay* and so is predicted to vote for it. Alternatively, Snowe and Coburn would support the *nay* alternative. More generally, knowing the spatial positions of the alternatives allows us to distinguish precisely between the ideal points of supporters and opponents.

A second useful fact is that given our assumption of symmetric preferences, each roll call can be characterized by a *cut point* or *cut line* that divides the ideal points of supporters from those of opponents. When the space of ideal points and alternatives is unidimensional as is the case in Figures 1 and 2, the cut line is simply a point. This point falls exactly half-way between

[7]In more complex settings where legislators vote over a sequence of proposals to reach a final outcome, sincere voting may not be a reasonable assumption.

the position of the *yea* and *nay* outcomes. This cut point is represented in Figure 2 where clearly all senators with ideal points to the left support the motion and all those to the right oppose it.

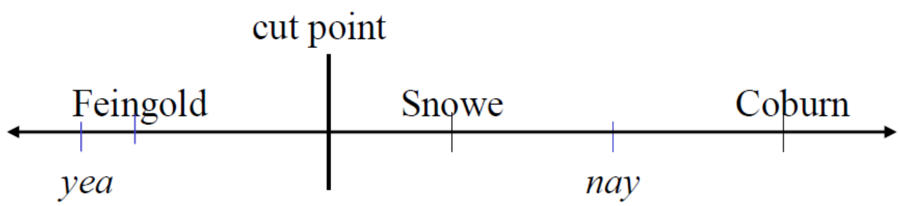

Figure 2: Cut Point of a Roll Call Vote

Consequently, if voting is based solely on the spatial preferences of legislators and there is no random component to vote choice, we can represent all voting coalitions in terms of ideal points and cut lines. This property turns out to be a crucial one for models of ideal point estimation. But it is important to remember that this convenience comes at the cost of the somewhat restrictive assumptions of symmetry and single-peakedness. To see why the assumption of symmetry is important, assume that Coburn's preferences in Figure 2 are asymmetric in that he prefers alternatives d units to the left of his ideal point to those d units to the right. This would make it possible to identify combinations of *yea* and *nay* outcomes for which Feingold and Coburn vote together against Snowe. Such a coalition structure cannot be represented by a single cut point. Similarly, if Coburn's preferences had two peaks the cut point condition could be violated. If he had second preference peak between Feingold and Snowe, it is easy to generate a roll call with a Feingold-Coburn versus Snowe outcome.

Before turning to the statistical models that have been developed to estimate legislative ideal points (and cut lines), it is instructive to consider the primary alternative to ideal point estimates: interest group ratings. The properties of these measures help clarify the potentials and the pitfalls of ideal point measures.

3 Interest Group Ratings

Interest group ratings of legislators have been compiled by a very diverse set of organizations, most notably the Americans for Democratic Action, the American Conservative Union, and the League of Conservation Voters. Many of the ratings go back a long time. Though precise details differ, interest group ratings are generally constructed in the following way:

1. An interest group identifies a set of roll calls that are important to the group's legislative agenda.[8]
2. The group identifies the position on the roll call that supports the group's agenda.
3. A rating is computed by dividing the number of votes in support of the group's agenda by the total number of votes identified by the group.[9]

[8] Usually the roll calls are selected after the votes have taken place, but on some occasions a group will announce that an upcoming vote will be included in their rating.

[9] Some groups treat abstention and absences as votes against the group's position.

For example, suppose a group chooses 20 votes. A legislator who votes favorably 18 times gets a 90% rating and one who supports the group 5 times gets a 25% rating.

It is easy to see how an interest group rating might be used as an estimate of a legislator's ideal point on the dimension defined by the group's agenda.[10] Assume that a group chooses p roll calls and that the cut points are $c_1 \leq c_2 \leq ... \leq c_{p-1} \leq c_p$.[11] Further, assume that the group has an ideal point greater than c_p. If all legislators vote in perfect accordance with their spatial preferences, all legislators with ideal points greater than c_p vote with the group p out of p times and get a 100% rating. Conversely, legislators with ideal points less than c_1 never support the group. In general, we can infer (under the assumption of perfect voting) that a legislator with a rating of $\frac{j}{p}$ has an ideal point between c_j and c_{j+1}. Unfortunately, we know only that $c_{j+1} > c_j$ and cannot observe c_j directly. Thus, interest group ratings provide only the ordinal ranking of ideal points. The upshot of this is that we have no way of knowing whether the distance between a 40% and a 50% rating is the same as the distance between a 50% and a 60% rating. This is a point ignored by almost all empirical work that uses interest group based measures.

[10]For expositional purposes, I assume throughout this section that legislators engage in *perfect spatial voting* in that behavior is determined solely by spatial preferences and without any random component. All of the issues would continue to arise with probabilistic spatial voting.

[11]Note that the indexing is arbitrary so this string of inequalities is without any loss of generality. Ruling out $c_i = c_{i+1}$ is for simplicity, but I shall return to it shortly.

Clearly, interest group ratings have many advantages. First, the scores directly relate to the policy concerns of the groups that compile them. League of Conservation Voters scores are based on environmental votes; the National Right To Life committee chooses votes on abortion, euthanasia, and stem cells. Second, groups often focus on important votes, whereas many of the statistical estimators discussed below use all or almost all votes. The expertise of the interest group in identifying key amendment or procedural votes adds value to their measures.[12] Finally, interest group ratings are easy to understand: Senator x supported group y's position p percent of the time.

But there are many ways in which interest group ratings perform poorly as estimates of legislator ideal points. I discuss each not to criticize interest group ratings, but because some of the issues reappear in ideal point estimation (albeit in a less transparent way).

3.1 Lumpiness

The first concern with interest group ratings as measures of preferences is that they are "lumpy" in that they take on only a small number of distinct values. If p votes are used to construct a rating, then the rating takes on only $p + 1$ different values: $0, 100/p, 200/p, \ldots,$ and 100. In many cases, this entails a significant loss of information about legislative preferences. If two members vote identically on the 20 votes selected by a group, they

[12] The concept of "importance" may not be clear in some cases, however. Groups may often include votes that represent purely symbolic support of their position.

receive the same interest group score regardless of how consistent their voting behavior is on all of the other votes. So legislators with very different true positions may achieve the same score. Lumpiness also exacerbates problems of measurement error (beyond those caused by the small sample of votes used). Because scores can only take on a small set of values, small deviations from pure spatial voting can lead to large changes in voting score. Suppose an interest group has chosen 10 votes that generate the cut points in Figure 3 below. The figure illustrates the interest group rating for each legislator located between adjacent cut points. The interest group rating for legislator A is 60% and it is 70% for legislators B and C. But suppose there was some small idiosyncratic factor that caused B to vote against the group on vote 7. Then B would have a 60% rating which is the same as A despite the fact that legislator B is located much closer to C who still scores 70%. Obviously, part of the problem is that the interest group has selected too few votes. If the group selected enough votes such that there were cut lines between A and B and between B and C the problem would be ameliorated somewhat. But no interest group chooses enough roll calls to distinguish 435 House members and 100 senators. But even if one did select enough votes, the interest group rating would still only reflect the order of the ideal points.

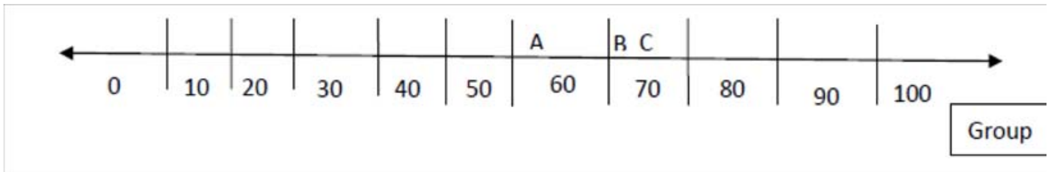

Figure 3: Measurement Error in Interest Group Ratings

3.2 Artificial Extremism

A second problem with interest group ratings concerns the relationship between the distributions of interest group ratings and ideal points. This problem was first identified by Snyder (1992). He provides a much more formal analysis of the problem, but it can be illustrated easily with a couple of figures. In Figure 4, there are five legislators and roll call cut points separate each of them. Consequently, each legislator gets a distinct score so that the distribution of ratings more or less matches the distribution of ideal points. But consider Figure 5. The difference is that now the cut points are concentrated in the middle of the spectrum.

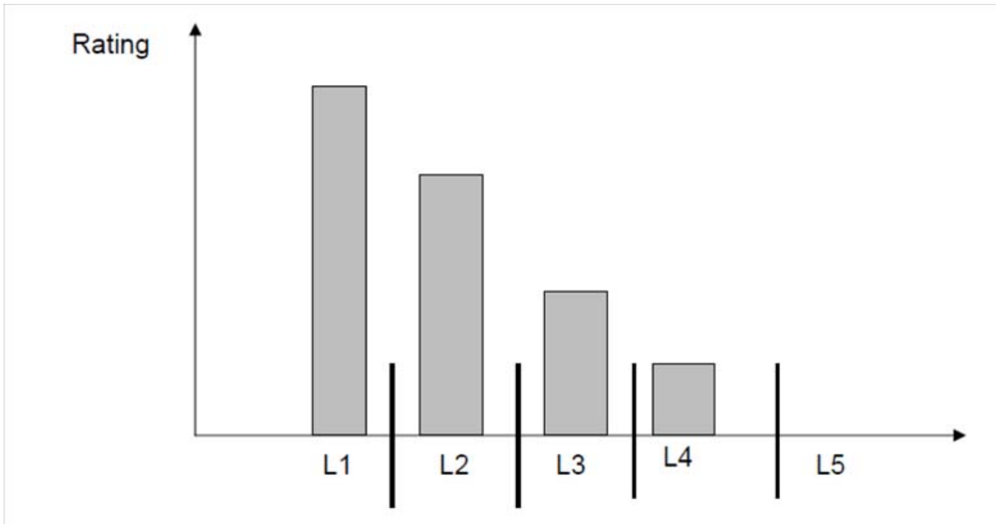

Figure 4: Interest Group Ratings with Uniform Cut Points

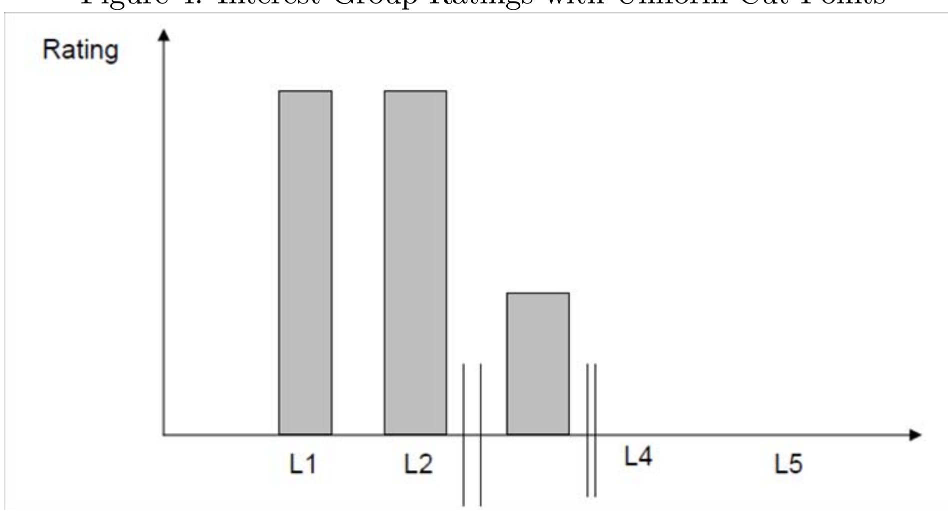

Figure 5: Artificial Extremism in Interest Group Ratings

Now legislators 1 and 2 have perfect 100% ratings and legislators 4 and 5 have 0% ratings. So it appears that the legislature is extremely polarized. But this is simply an artifact of the group having selected votes where the cutting lines are concentrated in the middle.

In generalizing this argument, Snyder proves that if the variance of the distribution of cut points is smaller than the variance of ideal points, the distribution of ideal points will be bimodal even if preferences are unimodal. Ultimately, the severity of this problem depends on the selection criteria that interest groups use. But it seems entirely plausible that groups are more interested in a rough division of the legislature into friends and enemies than in creating fine-grained measures of preferences for political science research.[13]

3.3 Comparisons over Time and Across Legislatures

Often researchers would like to compare the voting records of two legislators serving at different points in time or in different legislative bodies. Interest group ratings have been used for this purpose under the supposition that groups will maintain consistent standards for evaluation. Unfortunately, the supposition is invalid. The key to comparing ideal points of different legislators is the ability to observe how they vote on a common set of roll calls. If legislator A is voting over apples (Granny Smith versus McIntosh) and legislator B is voting over oranges (Navel versus Clementine), there is no way to compare their positions. This problem has both temporal and cross-sectional dimensions.

[13] In 2008, 20% of senators and 17% of House members recieved either a 100% or a 0% rating from the Americans for Democratic Action. The total number of 100% ratings would have been larger but for the practice of counting abstensions and missed votes as votes against the group.

It should be clear from the discussion above that comparability of interest group ratings requires that the distribution of cut points be the same across time or across legislatures. Of course, this is an impossibly stringent condition likely never to be satisfied. Consequently, obtaining a rating of 60% in time t may be quite different from obtaining a rating of 60% at time $t+1$. A score of 75% in the House is not the same as a score of 75% in the Senate.

Because variation in the distribution of cut points is inevitable, temporal and longitudinal comparisons of interest group ratings require strong assumptions to adjust scores into a common metric. For example, Groseclose, Levitt, and Snyder (1999) assume that each legislator's average latent Americans for Democratic Action score remains constant over time and upon moving from the House to the Senate. Similar problems persist in models of ideal point estimation. But as I discuss below, ideal point models provide additional leverage for dealing with these problems.

3.4 Folding and Dimensionality

Properly interpreting interest group ratings as (ordinal) measures of ideal points requires two additional assumptions. The first is that the interest group's ideal point is not "interior" to the set of legislator ideal points.[14] The second is that the interest group's agenda covers only a single dimension.

The importance of the requirement that the interest group occupy an

[14] See Poole and Rosenthal (1997; chapter 8) for a more extended discussion.

extreme position on its issue agenda is straightforward. If a moderate interest group compiles a rating, legislators occupying distinct positions to the group's left and its right will obtain the same score. Thus, rankings will not correlate with ideal points. A related problem concerns ratings from a group concerned with multiple policy areas where legislative preferences are not perfectly correlated. Suppose a group is concerned with liberalism on both social and economic issues. If the number of selected votes is the same across dimensions, a 50% rating would be obtained by a legislator supporting the group 50% on each issue and one supporting 100% on one issue and 0% on the other. Clearly, interest group ratings would not accurately reflect ideal points on either dimension.

4 Ideal Point Estimation

The preceding discussion highlights many of the difficulties in using interest group ratings as measures of legislative preferences. Most of these problems, however, are not unique and resurface in the statistical models discussed below. Because there are no free lunches, the improvements afforded always come at some cost. Either we must make strong assumptions about behavior or we must allow the models to perform less well in some other aspect. Because it ultimately falls to the end-user to decide which measures to use, understanding the underlying assumptions and trade-offs is crucial.

4.1 The Basic Logic

The underlying assumption of the spatial model is that each legislator votes *yea* or *nay* depending on which outcome location is closer to his or her ideal point. Of course the legislator may make mistakes and depart from what would usually be expected, as a result of pressures from campaign contributors, constituents, courage of conviction, or just plain randomness. But if we assume that legislators generally vote on the basis of their spatial preferences and that errors are infrequent, we can estimate the ideal points of the members of Congress directly from the hundreds or thousands of roll call choices made by each legislator.

To understand better how this is done, consider the following three senator example. Suppose we observed only the following roll call voting patterns from Senators Feingold, Snowe, and Coburn.

Vote	Feingold	Snowe	Coburn
1	YEA	NAY	NAY
2	YEA	YEA	NAY
3	NAY	YEA	YEA
4	NAY	NAY	YEA
5	YEA	YEA	YEA
6	NAY	NAY	NAY

Notice that all of these voting patterns can be explained by a simple model where all senators are assigned an ideal point on a left-right scale and every roll call is given a cut point that divides the senators who vote *yea* from those who vote *nay*. For example, if we assign ideal points such that Feingold<Snowe<Coburn, vote 1 can be perfectly explained by a cut point between Feingold and Snowe, and vote 2 can be explained by a cut point between Snowe and Coburn. In fact, all six votes can be explained in this way. Note that a scale with Coburn<Snowe<Feingold works just as well. But, a single cut point cannot rationalize votes 1-4 if the ideal points are ordered Snowe<Feingold<Coburn, Snowe<Coburn<Feingold, Coburn<Feingold<Snowe, or Feingold<Coburn<Snowe. Therefore none of these orderings is consistent with a one-dimensional spatial model. It is worth emphasizing that the data contained in the table is incapable (without further modeling assumptions) of producing a cardinal preference scale. Just like interest group ratings, it is impossible to know whether Coburn is closer to Snowe than Snowe is to Feingold.

As the two orderings of ideal points work equally well, which one should we choose? Given that Feingold espouses liberal (left wing) views and Coburn is known for his conservative (right wing) views, Feingold<Snowe<Coburn seems like a logical choice. Alternatively, one may look at the substance of the votes. If votes 1, 3, and 5 are liberal initiatives and 2, 4, and 6 are conservative proposals, the Feingold<Snowe<Coburn ordering also seems natural. But it is important to remember that there is no information contained in the matrix of roll calls itself to make this determination. It is purely an interpretive exercise conducted by the researcher.

An issue that recurs throughout the literature on ideal point estimation concerns unanimous votes like 5 and 6. Clearly, any ordering of legislators and any designation of cut points exterior to the range of ideal points can rationalize these votes. So in the sense of classical statistics, they are uninformative and would therefore play no role in the estimation of a spatial model.[15]

4.2 Probabilistic Voting

The real world is rarely so well behaved as to generate the nice patterns of votes 1-6. What if we observe that Coburn and Feingold occasionally vote together against Snowe, as in votes 7 and 8 below? Clearly, such votes cannot be explained by the ordering Feingold<Snowe<Coburn.

Vote	Feingold	Snowe	Coburn
7	YEA	NAY	YEA
8	NAY	YEA	NAY

If there are only a few votes like 7 and 8, it is reasonable to conclude that they may be generated by more or less random factors outside the model. To account for such random or *stochastic* behavior, estimators for spatial models assume that voting is probabilistic. There are many ways to generate probabilistic voting in a spatial context. One might assume for example

[15] Such votes may be informative in the Bayesian models I discuss below if one assumes informative priors about the distributions of ideal points and roll call outcomes.

that legislator ideal points are stochastic: Coburn might vote with Feingold against Snowe if his ideal point receives a sufficiently larger liberal shock than Snowe's does. Alternatively, one might assume that the voting alternatives are perceived differently by different legislators: Coburn might vote *yea* with Feingold against Snowe if he believes that the *yea* is more conservative than Snowe perceives it to be. Despite these logically coherent alternatives, the literature on ideal point estimation has converged on the random utility model. In the random utility model, a legislator i with ideal point x_i is assumed to evaluate alternative z_j according to some utility function $U(x_i, z_j)$ plus some error tern ε_{ij}. In such a framework, we might observe vote 7 if the underlying preferences predict vote 1 but Senator Coburn receives a large positive shock in favor of the *yea* outcome. Of course, such an outcome can be rationalized in many other ways. A shock to Snowe's utility could lead a vote 5 to be observed as vote 7. So identification of the ideal points and bill locations is sensitive to both the specification of the utility function U and the distribution of the error terms.[16] Within the range of modeling assumptions found in in the literature, the differences in estimates are usually small.

One of the payoffs to a probabilistic specification is that cardinal ideal point measures can be obtained whereas the deterministic analysis above produced only an ordinal ranking of ideal points. In the random utility framework, the frequency of the deviant votes provides additional informa-

[16]See Kalandrakis (forthcoming) for a disussion of the importance of various parametric assumptions for obtaining ideal points from roll call data.

tion about cardinal values of the ideal points. Suppose we assume that small shocks to the utility functions are more frequent than large shocks. Then, if there are few votes pitting Coburn and Feingold against Snowe, the random utility models place Coburn and Feingold far apart, to mimic the improbability that random events lead them to vote together. Alternatively, if the Coburn-Feingold coalition were common, the models place them closer together, consistent with the idea that small random events can lead to such a voting pattern. But clearly, if we assume that large shocks are more common than small shocks, the logic would be reversed. So estimates of nominal ideal points are somewhat sensitive to the specification of random process.[17]

4.3 Multiple Dimensions

Sometimes there are so many votes like 7 and 8 that it becomes unreasonable to maintain that they are simply the result of random utility shocks. An alternative is to assume that a Coburn-Feingold coalition forms because there exists some other policy dimension on which they are closer together than they are to Snowe. We can accommodate such behavior by estimating ideal points on a second dimension. In this example, a second dimension in which Coburn and Feingold share a position distinct from Snowe's, ex-

[17] This problem is not unique to ideal point estimation. It is generic to the estimation of discrete choice models. For example, in a probit or logit model, the predicted probabilities are identified purely from the form of the distribution function. Post-estimation analysis tends to support the assumption that the error process is unimodal around zero as most deviations from the prediction of the spatial model cluster around the cut point. See Poole and Rosenthal (1997; p. 33).

plains votes 7 and 8. In fact, both dimensions combined explain all of the votes. Obviously, in a richer example with 100 senators rather than 3, two dimensions cannot explain all the votes, but adding a second dimension adds explanatory power. So the primary question about whether to estimate a one, two, or more dimensional model is one of whether the higher dimensions can both explain substantially more behavior and can be interpreted substantively. Otherwise, the higher dimensions may simply be fitting noise.

5 Estimation

How exactly are ideal points estimated? For a clearer presentation, I focus on the case of a one dimensional model. The generalization to multiple dimensions is fairly straightforward, but I will indicate where it is not.

As discussed above, the common framework is a random utility model where the utilities of voting for a particular outcome are based on a deterministic utility function over the location of the outcome and a random component. Formally, let x_i be legislator i's ideal point, y_j be the spatial location associated with the *yea* outcome on vote j and n_j be the location of the *nay* outcome. Moreover, let ε_j^y and ε_j^n be random shocks to the utilities of *yea* and *nay*, respectively. Therefore, the utilities for voting *yea* and *nay* can be written as

$$U(x_i, y_j) + \varepsilon_j^y$$

$$U(x_i, n_j) + \varepsilon_j^n$$

where it is assumed that U is decreasing in the distance between the ideal point and the location of the alternative. It is further assumed that the utility functions are Bernoulli functions that satisfy the axioms of the von Neuman-Morgernstern theorem.[18] A consequence of that assumption is that we can rescale the x_i, y_j,and n_j without affecting voting behavior. Specifically, estimates of ideal points and bill locations are identified only up to a linear transformation.[19] This issue generates problems similar to those associated with comparing interest group ratings across chambers or years. Without common legislators or common votes, the ideal point estimates of different chambers differ by unobserved scale factors. I discuss below several attempts to work around this problem.

Given a specification of utility functions, the behavioral assumption is that each legislator votes for the outcome that generates the highest utility.[20] Specifying a functional form for the random shocks allows the derivation of choice probabilities and the likelihood function of the observed votes which

[18] See McCarty and Meirowitz (2006; p36-37).

[19] Formally, $x_i' = \alpha + \beta x_i$, $y_j' = \alpha + \beta y_j$ and $n_j' = \alpha + \beta n_j$ produce identical behavior as x_i, y_j, and n_j.

[20] This assumption is not innocuous. It rules out some forms of strategic voting. But if legislators vote on a binary agenda, we can reinterpret y_j and n_j as the sophisticated equivalents of a *yea* and *nay* vote (see Ordeshook 1986).

can be used for maximum likelihood or Bayesian estimation.[21]

A complication arises in that except under fairly restrictive modeling choices, the likelihood function will be extremely non-linear in its parameters. So typically estimating ideal point models will either involve alternating procedures (e.g. Poole and Rosenthal 1997) or Bayesian simulation (e.g. Clinton, Jackman, Rivers 2004; Martin and Quinn 2002).

6 NOMINATE

The seminal contribution to estimating legislator ideal points from a probabilistic spatial voting model is Poole and Rosenthal's (1985) NOMINATE model.[22]. The earliest static version of the model implements a probabilistic voting model by assuming that the utility of alternative z for a legislator with ideal point x is

$$U(x,z) = \beta \exp\left[-\frac{(x-z)^2}{2}\right]$$

[21] Formally, the model predicts that legislator i votes *yea* on roll call j if and only if

$$\begin{aligned} U(x_i, y_j) + \varepsilon_j^y &\geq U(x_i, n_j) + \varepsilon_j^n \\ U(x_i, y_j) - U(x_i, n_j) &\geq \varepsilon_j^n - \varepsilon_j^y \end{aligned}$$

Let F be the cumulative distribution function of $\varepsilon_j^n - \varepsilon_j^y$, then the probabilities of voting *yea* and *nay* are simply

$$\begin{aligned} \Pr\{yea\} &= F(U(x_i, y_j) - U(x_i, n_j)) \\ \Pr\{nay\} &= 1 - F(U(x_i, y_j) - U(x_i, n_j)) \end{aligned}$$

[22] The term NOMINATE is derived from *NOMINA*l *T*hree-step *E*stimation.

and that the random shocks are distributed according to the Type I extreme value distribution. The parameter β represents the "signal-to-noise" ratio or weight on the deterministic portion of the utility function.[23]

The utility function employed by NOMINATE has the same shape as the density of the normal distribution and is therefore bell-shaped. For convenience in estimation, Poole and Rosenthal transform the model so that the roll call parameters y and n are replaced by a cut point parameter $m = \frac{y+n}{2}$ and a distance parameter $d = \frac{y-n}{2}$.

Although Poole and Rosenthal selected this functional form to facilitate the estimation of the *yea* and *nay* outcome positions[24], it has important substantive consequences. This exponential form implies that a legislator will be roughly indifferent between two alternatives that are located very far from her ideal point (in the tails, the utilities converge to zero). This is quite different from the implications of the quadratic utility function $U(x,z) = -(x-z)^2$ used in much of the applied theoretical literature and later models of ideal point estimation. With quadratic utility functions, the difference in utilities between two alternatives grows at an increasing rate as the alternatives move away from the ideal point.[25] As a substantive conjecture about behavior, the

[23] Under these assumptions, $\varepsilon_j^n - \varepsilon_j^y$ is distributed logistically and

$$\Pr\{yea\} = \frac{\exp\left[U(x_i, y_j)\right]}{\exp\left[U(x_i, y_j)\right] + \exp\left[U(x_i, n_j)\right]}$$

[24] See Poole and Rosenthal (1991, fn 6).

[25] Carroll et al (forthcoming) show that within the empirically relevant range of roll call locations, the difference in choice probabilities generated by these two utility functions are quite small.

exponential assumption seems more reasonable. Who would perceive bigger differences between Fabian socialism and communism? A free-market conservative or a communist? The communist seems the better bet. Clearly, however, it is unsettling that the identification of y and n depend on the choice of function. But while estimates of d are less than robust, the cut point m is estimated precisely.

Poole and Rosenthal (1997) extend this static model to a dynamic one (D-NOMINATE) and estimate the ideal points of almost all legislators serving between 1789 and 1986 and the parameters associated with almost every roll call.[26] In estimating the dynamic model, Poole and Rosenthal confront the same comparability problem that I discussed above in the context of interest group ratings. Their main leverage for establishing comparability is that many members of Congress serve multiple terms and that Congress never turns over all at once. So there are many overlapping cohorts of legislators. These overlapping cohorts can be used to facilitate comparability. For example the fact that Kay Bailey Hutchison served with both Phil Gramm and John Cornyn as Senators from Texas allows us to compare Gramm and Cornyn even though they never served together. This would be accomplished most directly if we assume that Hutchison's ideal point was fixed throughout

[26] Obviously, estimating a legislator's ideal point requires a reasonable sample of roll calls. Poole and Rosenthal decided only to include those legislators who voted at least 25 times. Recall from the discussion above, unanimous votes are not informative in that they are consistent with an infinite number of cut points (any that are exterior to the set of ideal points). When voting is probabilistic, near unanimous roll calls are not very informative either. So Poole and Rosenthal include only roll calls where at least 2.5% of legislators vote on the minority side.

her career. But that assumption is much stronger than what is required. Instead, Poole and Rosenthal assume that each legislator's ideal point moves as a polynomial function of time served, though they find that a linear trend for each legislator is sufficient.

Despite the fact that D-NOMINATE produces a scale on which Ted Kennedy can be compared to John Kennedy and to Harry Truman, some caution is obviously warranted in making too much of those comparisons. Although the model can constrain the movements of legislators over time, the substance of the policy agenda is free to move. Being liberal in 1939 means something different than liberal in 1959 or in 2009. So one has to interpret NOMINATE scores in different eras relative to the policy agendas and debates of each.[27]

Perhaps the most important substantive finding of their dynamic analysis is that legislative voting is very well explained by low dimension spatial models. With the exception of two eras (the so-called "Era of Good Feeling" and the period leading up to the Civil War) a single dimension explains the bulk of legislative voting decisions. Across all congresses the single

[27] In an attempt to overcome this problem, Bailey (2007) exploits the fact that Supreme Court justices, presidents and legislators often opine about old Supreme Court decisions. If one assumes that these statements are good predictors of how these actors would have voted on those cases, justices, presidents, and legislators can be estimated on a common scale with a fixed policy context. For example, if Justice Scalia says he supports the decision in Brown, we are to infer that he would have voted for it and we can use that information to rank his preferences along with those who were on the court in 1953. But this is a very strong assumption. Perhaps Scalia supports Brown because it is settled law or the social costs of reversal are high, or it is just bad politics now to say otherwise. Thus, it would be difficult to infer from his contemporary statements how he would have voted.

dimension spatial model correctly predicts 83% of the vote choices. Of course, unlike the case of interest group ratings, labeling that dimension is somewhat subjective. Poole and Rosenthal argue that the first dimension primarily reflects disagreements about the role of the federal government especially in economic matters. But of course the content of this debate changed dramatically over time from internal improvements, to bimetallism, to the income tax, and so on.

Overall, a two-dimensional version of the D-NOMINATE model explains 87% of voting choices, just 4% more than the one-dimensional model. But there are periods in which a second dimension increases explanatory power substantially. The most sustained appearance of a second dimension runs from the end of WWII through the 1960s where racial and civil rights issues formed cleavages within the Democratic party that differed from conflicts on the economic dimension.

6.1 Newer Flavors

Subsequent to their work using D-NOMINATE, Poole and Rosenthal have refined their models in a variety of directions. D-NOMINATE assumes that legislators place equal weight on each policy dimension. Consequently, the importance of a dimension is reflected in the variation of ideal points and bill locations along that dimension. The variation of ideal points increases with the salience of the dimension. An alternative approach is to fix the variation of ideal points and bill locations and allow the weight that legislators place on each dimension to vary. W-NOMINATE implements just such an alternative. Additionally, W-NOMINATE contains several technical innovations that optimize it for use on desktop computers (D-NOMINATE was originally estimated on a supercomputer).

Subsequently, McCarty, Poole, and Rosenthal (1997) develop a dynamic version of W-NOMINATE. In addition to distinct weights for each dimension, DW-NOMINATE differs from D-NOMINATE in that the stochastic component of the utility function is based on the normal distribution rather than the Type II extreme value.

While D- and DW-NOMINATE address the intertemporal comparability problem by restricting the movement of legislators over time, the sets of scores for the House and Senate are not comparable. In order to address this issue, Poole (1998) develops a model that uses members who serve in both chambers to transform DW-NOMINATE scores into a common scaling for both chambers. He has dubbed these results "common space NOMINATE." Finally, Poole (2001) develops a related model based on quadratic utilities and normal error distributions. This is often referred to as the QN model.

7 Estimation Issues

All of the standard ideal point models have to confront a number of practical issues that emerge in estimation. Although some of these issues may seem a little subtle or arcane, it is in how these issues are handled that distinguish the primary approaches to ideal point estimation. Consequently, the applied researcher should be familiar with these issues and the consequences of different means of addressing them.

7.1 Scale Choice

As I discussed above, the scale of ideal points is latent and identified only up to a linear transformation. Consequently, any estimation procedure needs to make some assumptions to pin down the scale. For example, in one dimension, NOMINATE assumes that the leftmost legislator is located at -1 and the rightmost is located at 1. Not only does this assumption help pin down the scale, but it alleviates the following problem. Suppose a legislator was so conservative that she voted in the conservative direction on every single roll call. Independent of any other ideal point location, her ideal point could be 1, 10, or 100 with very little impact on the likelihood of the estimate. Constraining her ideal point to be no higher than 1 and constraining the gap between her and the nearest legislator alleviates what Poole and Rosenthal dub the "sag" problem – an appeal to the image of extreme legislators' positions spreading out like an old waistband.

The estimates of some roll call parameters must also be constrained for identification reasons. Consider the cut point of the roll call $m = \frac{y+n}{2}$. Suppose that there is a near unanimous roll call in favor of a liberal proposal. Then any $m > 1$ might be a reasonable estimate of this parameter. Consequently, m is constrained a location between -1 and 1. Problems also arise with the distance parameter $d = \frac{y-n}{2}$. Suppose that on some roll call every legislator flips a fair coin. Very different values of d can produce the appropriate likelihood function. When $d = 0$, the alternatives are the same so that legislators flip coins. When $d = \infty$ (and m is between -1 and 1), both alternatives are so bad that a legislator is indifferent and flips a coin. Given this problem, d is constrained so that at least one of the bill locations (y or n) lies on the unit interval.[28]

A final issue in the selection of the scale concerns the variance of the random utility shocks. Whether NOMINATE is estimated with a logit function (as in D-) or a probit function (as in DW-), the assumed variance of the shocks is fixed – one roll call has just as much randomness as another. The parameter β, however, controls for the weight placed on the deterministic part of the utility function so that the effects of the variance are scaled by $\frac{1}{\beta}$. Without β to control the effects of variance, the estimates of the distance parameter d would be distorted in trying to account for it. To see this, compare two roll calls that differ only in the variance of the error terms. In the noisier roll call, the choice probabilities should all be closer to .5. One way to achieve this is to move the estimate of d closer to zero (i.e. make y and n more similar). Consequently, our confidence in estimates of d (and therefore y and z) depends on β capturing all of the effects of the variance of the stochastic term. Since d is imprecisely estimated, the *yea* and *nay* outcome coordinates will be as well. Therefore, use of the outcome

[28] These constraints together imply that $|\min(m + d, m - d)| < 1$.

coordinates is not recommended without adjusting for the level of noise in the roll call (see McCarty and Poole 1995). This problem has limited the applicability of ideal point models for studying policy change.

7.2 Sample Size

The number of parameters per dimension for the NOMINATE models is $p+2q$ where p is the number of legislators and q is the number of roll calls. Of course, for any typical legislature this will be a very large number of parameters. Fortunately, the sample of vote choices is pq and is consequently larger than the number of parameters so long as $p > \frac{2q}{q-1}$. However, because one cannot increase the sample size without increasing the number of parameters, it is impossible to guarantee that the parameter estimates converge to their true values as the sample size goes to infinity i.e. the estimates are inconsistent.[29] Therefore, Poole and Rosenthal conducted numerous Monte Carlo studies to establish that NOMINATE does a reasonable job at recovering the underlying parameters in finite samples.

Heckman and Snyder (1997) propose an alternative model that does consistent estimates of ideal points, but not bill locations. In addition to the assumption of quadratic preferences, the Heckman-Snyder estimator requires that $\varepsilon_y - \varepsilon_n$ be distributed uniformly. They demonstrate that under these assumptions ideal points can be estimated using factor analysis.[30] When

[29] This is known as the incidental parameters problem.

[30] While Heckman and Snyder's estimates of bill locations are inconsistent, the linearity of the model prevents this inconsistency from feeding back into the estimates of ideal

implementing the model, they find that their results for one or two dimensions are almost identical to NOMINATE apart from some differences in the extremes of the ideal point distribution. This suggests that the consequences of the inconsistency of NOMINATE are small.[31]

Both the asymptotic results of Heckman and Snyder and the Monte Carlo work of Poole and Rosenthal suggest that it is important for both p and q to be large. The following example from Londregan (2000 a,b) helps illustrate why. Consider a situation with only three legislators 1, 2 and 3. On a particular roll call, they vote as shown in Figure 6. Note that both cut points m' and m'' are consistent with the observed voting pattern. The precise estimate of m (and therefor y and n) will depend entirely on the functional form of the random component of the utilities. Consequently, m is also likely to be estimated with large amounts of error. Of course, if we are only interested in the ideal points this may be tolerable. But remember that the quality of the estimates of the ideal points will depend on the quality of the estimates of m. So the ideal points will be estimated poorly as well.

Unfortunately, many of the institutions for which we would like ideal point estimates, such as courts and regulatory boards, are quite small. So how should researchers approach such applications?

---

points.

[31] The primary differences between Poole-Rosenthal and Heckman-Snyder concern the dimensionality of the policy space. I take this issue up below.

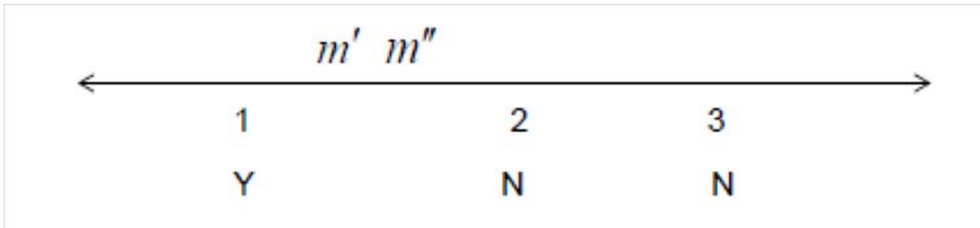

Figure 6: The Granularity Problem

An obvious choice is to simply accept that the problem exists and go ahead and run NOMINATE or Heckman-Snyder. The downside, of course, is that the estimates will not be precise.[32] Doing better than that requires an accurate diagnosis of the problem. At the root of the problem is that roll call voting data contains precious little information necessary to generate cardinal estimates. As I discussed above, cardinality requires making assumptions about the random process that generates voting errors. When there are few legislators, the reliance on parametric assumptions rises disproportionately. The real problem is that roll call data by itself is inadequate. More data about legislative preferences or proposals can help ameliorate this problem.

First consider observable covariates about preferences. Let's say we have an observed variable w. Something like region, value-added from manufacturing, or district partisanship that we believe is plausibly related to legislative policy preferences. Then we could model each ideal point as

[32] In the case of the earlier versions of NOMINATE, this problem is confounded by the fact that its iterative maximum likelihood procedure underestimates the uncertainty associated with its estimates. Estimation of the covariance matrix in the Heckman-Snyder model is computationally prohibitive.

More recently, Lewis and Poole (2004) have implemented bootstrapping procedures to better recover the uncertainty in parameter estimates. The Bayesian procedures described below deal with estimation uncertainty directly.

$x_i = \alpha_{1i} + \alpha_{2i} w_i$. The inclusion of w_i helps pin down the scale and locate ideal points. This in turn improves the estimation of roll call cut points, which improves ideal point estimation, and so on.

Information about proposals can also be useful. The best application of this insight is Krehbiel and River's (1988) work on the minimum wage. Because minimum wage proposals are denominated in dollars, y and n (and therefore m) are observed directly. Given the observed cut points, one only has to estimate the ideal points on the scale defined by dollars.

The difficulty of both approaches relates to the availability of auxiliary data. Lots of potential covariates exist for preferences. The trick is generating a parsimonious specification. Moreover, many scholars are interested in an unobserved component (ideology?) of legislative preferences, so preference covariates can never eliminate the problem. One encounters the opposite problem when it comes to modeling proposals with observable variables. Many legislative proposals cannot be quantified like budgets and wage floors can be. Londregan (2000a-b) takes an approach that makes fewer demands in terms of observable data. Rather than attempt to measure preferences and proposals, he models the proposal making process. In general, such an approach would involve assuming that legislator i's optimal proposal can be related to the other parameters of the model. Such assumptions can be used to pin down some of the model's parameters. Londregan assumes that legislators always propose their own ideal point.[33] Of course, the accuracy of

[33] Londregan's model departs from the standard model by assuming that some legislators

the estimates depends on the validity of the proposal function.

A similar approach is employed by Clinton and Meirowitz (2003, 2004). They leverage the fact that along an agenda sequence, one of two things must be true. If a new proposal is adopted at time $t-1$, it becomes the status quo at time t. If the new proposal fails at time $t-1$, then the status quo from $t-1$ becomes the status quo at time t. Imposing these constraints helps pin down the proposal parameters.

It is important to note, however, that all of these approaches simply shift the weight of one set of parametric assumptions – the stochastic process – to another set – modeling choices about covariates, proposal making, or agendas. The only alternative to this trade-off is to give up on the ability to generate cardinal ideal points and settle for extracting the ordinal information from the roll call data. Such is the approach of Poole's (2000) Optimal Classification (OC) algorithm. As I demonstrated above, when legislative voting is in perfect accord with spatial preferences, it is possible to rank order the ideal points of legislators on the issue dimension. But of course, the distances between any two legislators is unidentified without voting errors and assumptions about the distribution of the shocks that generate those errors. In the presence of voting error, Poole's algorithm makes no assumptions about the process generating those errors. It simply tries to order the legislators in such a way as to minimize the number of errors. For large legislatures, the

make "better" proposals than others. This valence effect is equivalent to assuming that the distribution of ε_y varies across legislators.

OC estimates correlate very highly with the ranking of NOMINATE ideal points. For smaller bodies the correlations are much lower, reflecting the importance of parametric assumptions in pinning down cardinal estimates. But even in large legislatures, OC and NOMINATE sometimes produce substantively important differences. Suppose a legislator is a "maverick" like John McCain or Russ Feingold and makes large voting errors. Parametric models like NOMINATE penalize such errors harshly. Consequently, the model will move legislators to more moderate positions in order to minimize the large errors. In OC, an error is an error, no matter how large. Without the extra penalty, OC will weigh Feingold's predictably liberal votes and McCain's predictably conservative votes relatively more than their "mavericky" votes. So Feingold is more likely to be identified as a liberal and McCain as a conservative in OC than in NOMINATE.

8 Bayesian Estimation

Over the past decade, there has been tremendous progress in applying Bayesian Item Response Theory (IRT) to the task of estimating spatial models of roll call voting. IRT was originally developed in the context of educational testing to facilitate the estimation of test takers' ability when the quality and difficulty of an examination is unknown. Beginning in the 1990s, scholars began applying Bayesian estimation techniques such as Markov Chain Monte Carlo (MCMC) and Gibbs sampling (Albert 1992, Albert and Chibb

1993).[34]

To justify the use of the Bayesian IRT, scholars (Jackman 2001, Jackman, Clinton, Rivers 2004, Martin and Quinn 2002) assume that legislator preferences are quadratic. Together with the assumption that voting errors are normally distributed, this implies that the probability of a *yea* outcome is simply $\Phi(\alpha_{j0}+\alpha_{j1}x_i)$ where where $\alpha_{j0} = n_j^2 - y_j^2$, $\alpha_{j1} = 2(y_j - n_j)$, and Φ is the normal distribution function. This probability is analogous to the probability of a correct answer in educational testing. In IRT, x_i represents the student's ability. Assuming $\alpha_1 > 0$, higher ability students are more likely to answer questions correctly. The parameter α_{j0} represents difficulty of the test item; easy questions have higher values of α_{j0}. Finally, α_{j1} is known as the discrimination parameter as it determines the marginal impact of student ability on the probability of a correct answer. If α_{j1} is close to zero, good and bad students get the correct answer at roughly equal probabilities as the item fails to discriminate on ability.

These interpretations extend (albeit imperfectly) to the ideal point context. The difficulty parameter α_{j0} is the difference between the distances from 0 for the two alternatives. When $y_j^2 < n_j^2$, the likelihood of voting *yea* is higher, holding the ideal point constant. The discrimination parameter α_{j1} plays two roles. First, it defines the polarity of the roll call so that when $\alpha_{j1} > 0$, the *yea* outcome is the conservative outcome and legislators with

[34] Carroll et al (forthcoming) provide a very comprehensive survey of the differences between W-NOMINATE and the Bayesian model in the case of a single dimension.

higher values of x_i are more likely to support it. Second, by reflecting the difference between the two alternatives, the absolute value of α_{j1} controls how well the roll call discriminates between liberal and conservative legislators.[35]

Perhaps the most direct benefit of the Bayesian approach is that uncertainty in the estimated parameters is easily measured and summarized. The magnitude of this advantage, however, may have narrowed with the advent of the parametric bootstrap for computing the standard errors of NOMINATE parameters.

The Bayesian model provides some distinct computational advantages and disadvantages. A major advantage is that it is much more straightforward to use covariates and other information to model ideal points and proposal parameters. For example, Clinton, Jackman, and Rivers (2004) incorporate party membership directly into the model to estimate distinct cut points for each party.[36] Because the likelihood function of NOMINATE is already highly non-linear and non-concave, estimating more complex models is often infeasible. The primary disadvantage, however, is that even the simplest IRT model can be very time consuming to estimate when compared to NOMINATE or Heckman-Snyder.

Identification of the Bayesian model can be achieved either by the imposition of prior distributions on the parameters or through constraints like those used in NOMINATE. Early implementations such as Jackman (2001)

[35] The cut point and distance parameters can be easily recovered from the IRT parameters as $m = -\frac{\alpha_0}{\alpha_1}$ and $d = \frac{\alpha_1}{2}$.

[36] See also McCarty, Poole, and Rosenthal (2000).

assumes a prior distribution that is normal with mean 0 and unit variance. But because this is an assumption about the population from which ideal points are drawn, rather than the sample of legislators, it does not fully anchor the scale and leads to an overstatement of estimation uncertainty (see Lewis and Poole 2004). So later models such as Clinton, Jackman, Rivers (2004) identify the scale by setting one legislator to -1 and another to $+1$ a priori. In estimating their dynamic model of Supreme Court voting, Martin and Quinn (2002) assume informative priors for several justices (in their first term) to pin down the scale.

In the end, it appears that differences in estimates between NOMINATE and the Bayesian models are quite small. As the extensive Monte Carlo analysis of Carroll et al (forthcoming) concludes, subtle differences are created by the different functional form assumptions as well as procedural choice in the handling of identification restrictions. But with the development of the Poole's QN model and MCMC versions of NOMINATE (Carroll et al 2009), the differences are likely to get smaller and the choice between estimators will become one of convenience and taste.

9 Model Evaluation

While in principle one could apply any measure of fit appropriate for a maximum likelihood or Bayesian estimator, much of the literature, beginning with Poole and Rosenthal, focuses on measures of predictive or "classification" success. Such measures are straightforward. They reflect the percentage of cases where the vote choice with the higher estimated likelihood corresponds to the observed vote. The primary virtue of classification measures is that they are intuitive and easy to interpret. A second advantage of classification success is that it is straightforward to compare very different models. Not only does it allow comparisons between different parametric models like NOMINATE, Heckman-Snyder, and the Bayesian model, it can also compare them against non-parametric alternatives like Optimal Classification.[37]

Classification success has a number of drawbacks as well. The first is that it can be artificially inflated if there are a large number of lopsided votes. Suppose for example that the average margin on a roll call vote is 65%. Then a 65% classification success can be generated by a naive model that predicts that all legislators vote for the winning alternative. If for some reason winning margins increase, the baseline classification rate will rise even if the ideal point model has no additional explanatory power. For this reason, scholars beginning with Poole and Rosenthal have often focused on a measure called proportional reduction in error (PRE). Intuitively, the PRE specifies how much better the spatial model performs than a "majority" model that assumes that every legislator votes for the more popular

[37] Of course, because OC maximizes classification success, it will generally outperform on this criterion the alternatives based on maximum likelihood.

alternative. Formally,

$$PRE = \frac{Majority\ Errors\ -\ Model\ Errors}{Majority\ Errors}$$

where *Majority Errors* are the number of votes against the majority position and *Model Errors* are the number of incorrect predictions of the spatial model. A second concern is that both classification success and the PRE weigh mistakes close to the cut line just as heavily as big mistakes. To penalize mistakes according to their magnitude generally requires the use of likelihood-based measures. Poole and Rosenthal propose the use of the geometric mean probability (GMP) which is computed as the anti-log of the average log-likelihood. Unfortunately, such measures are more difficult to compare across estimators.

Finally, classification measures do not easily lend themselves to formal statistical inference the way that other methods such as the Wald statistic and likelihood-ratio test do. This issue has been the most prominent in the debate over the dimensionality of roll call voting. On the basis of classification analysis, Poole and Rosenthal conclude that congressional voting in the United States is largely one dimensional. With the exception of a few definable eras, the classification success of a two dimensional model is generally no more that two percentage points higher than a one dimensional model. But efforts to apply formal statistical inference to the question of dimensionality generally reject the unidimensional model in favor of higher dimensional alternatives (see Heckman and Snyder 1997 and Jackman 2001). These different findings present something of a conundrum to applied researchers. Clearly formal statistical tests are preferable to ad hoc rules of thumb.[38] But substantive significance and parsimony are also important considerations. Given these competing considerations, this debate will likely continue.

10 Interpretation

In the end, the key problem for legislative scholars is figuring out exactly what the ideal point estimates mean. Technically, ideal points are simply a low dimension representation of all the considerations that go into roll call voting. But what are the political, psychological, and strategic factors that create the types of coalition structures that are so easily represented in one or two dimensions? Drawing on the ideas of Phillip Converse (1964), Poole and Rosenthal have explained low dimensionality in terms of *belief constraint*. This is the notion that political elites share beliefs about which issue opinions go together. Support for tax cuts is correlated with support for deregulation which is correlated with higher defense spending because

[38]There are reasons to be wary of the statistical tests that have thus far been proposed. First, testing a two-dimensional model versus a one-dimensional model involves one restricted parameter per legislator and two per roll call. So tests can be sensitive to the penalties imposed for additional parameters (see Jackman 2001). Second, the tests have been carried out under the assumption that unobserved errors are independent across legislators. Higher dimensional models may simply be fitting cross-legislator correlations that do not constitute substantively important dimensions.

elites believe that those issues are related. But constraint is really little more than a description of the phenomena, not an explanation. There is no logical underpinning for the issue configurations that define left and right. Why exactly should a pro-life position predict support for tax cuts? Moreover, these configurations have changed over time. Protectionism used to be the province of the conservative. Ideal point estimation and dimensional analysis can identify when such changes occur, but it has more limited value in explaining why.

One of the more contentious debates surrounding the interpretation of ideal points is the argument about the role of ideology. The argument that ideal points reflect ideological preferences is based on several empirical findings. The first is that ideal point estimates are quite stable for politicians throughout their career. Of course, there are a few prominent examples of politicians whose positions did change; such as the right to left movements of Wayne Morse, or the left to right movements of Richard Schweiker or the right to left to right movements of John McCain. A very small number of politicians like William Proxmire have been nothing if not erratic. But for the most part, legislators' ideal points only move significantly if they switch parties (and of course party switching is quite rare).[39] Even a member whose constituency changes quite dramatically, either by elevation to the Senate or through major redistricting, rarely changes positions in a significant way. In a very careful study, the assumption that legislators maintain the same

[39] See McCarty, Poole, and Rosenthal (2001).

ideological position throughout their careers performs just as well statistically as the assumption that legislators are able to change positions in each biennial term.[40]

The second piece of evidence in favor of an ideological interpretation is that the behavior of legislators deviates in large and systematic ways from the preferences of their average or median constituent. This finding persists even when we do not worry about the mismeasurement of constituency interests or preferences. For example, senators from the same state do not vote identically. Most obviously, senators from the same state but different parties, such as Bill Nelson and Mel Martinez of Florida, pursue very different policy goals. The difference is picked up in their polarized ideal point estimates. If the two senators are from the same party, they are, of course, more similar. Even here, however, there are differences. Consider California Democrats Diane Feinstein and Barbara Boxer. They not only represent the same state but were first elected by exactly the same electorate on the same day in 1992. In the most recent Senate term, Boxer has a DW-NOMINATE score of -0.601 making her the third most liberal member of the U.S. Senate. Conversely, Dianne Feinstein's ideal point is just -0.384 making her the 37th most liberal. Moreover, there is nothing unusual about this California duo. Four other states have pairs of senators from the same party whose NOMINATE scores differ at least as much.

House districts, being single-member, do not allow the same natural ex-

[40] See Poole (2007).

periment that is possible for the Senate. It is possible, however, to compare the voting behavior of a member to his or her successor. The same-party replacements of House members can have ideal points that are very different from those of their predecessors. True, a relatively liberal Democrat is likely to be replaced by another liberal Democrat, but the variation in the scores of the same-party replacements is very large. It is about half as large as the total variation of positions within the party.[41] In other words, the ideal point of the outgoing incumbent is at best a crude predictor of the position of the new member even if they are in the same party.

While the evidence is strong that ideal points reflect ideology to some degree, clearly ideal point estimates may also reflect any number of other influences such as constituent interests, interest group pressure, or partisan strategies. A large body of literature trys to parse these components but there is currently little methodological or substantive agreement.[42]

Another important issue for the end-user concerns how to interpret ideal points in multiple dimensions. Perhaps because ideal point producers like to name their dimensions – economic, social, race, etc. – end-users often mistakenly infer that the ideal point coordinate from the social dimension is the best predictor of abortion votes and the coordinate from the race dimension is the best predictor of civil rights votes. This confusion arises because

[41] See Poole and Romer (1993).

[42] A sampling of some of this work includes Glazer and Robbins (1985), Kalt and Zupan (1990), Levitt (1996), Rothenberg and Sanders (2000), Snyder and Ting (2003), Lee, Moretti, and Butler (2004), and McCarty, Poole, and Rosenthal (2006).

the estimated dimensions are orthogonal by design. But few substantive issues are completely uncorrelated from all others. Consequently, it is not the coordinates from specific dimensions that are best predictors, but rather the projection of the ideal point vector (the coordinates from all dimensions) onto a line that is perpendicular to the cut line. To illustrate, consider Figure 7 which represents Senate ideal points and voting on the 1964 Civil Rights Act. Clearly, the positions on the first dimension are inadequate to explain this vote, as they predict that many Southern Democrats would actually support the bill. But the second dimension alone does not do so well either. It would considerably over predict the number of Republicans who voted for the bill. Instead, multi-dimensionality is reflected in the fact that the cut line for the bill deviates considerably from the vertical line one would expect for a first dimension vote (yet it also is not horizontal as would be expected by a purely second dimension vote). But we can generate a single measure reflecting civil rights voting. First, we choose a line perpendicular to the cut line such as the dotted line in the figure. Then we projecte the ideal points onto the perpendicular line. In other words, we move each ideal point to its closest point on the perpendicular line. The projection is illustrated by the arrows in the figure. Note that the new positions do as good a job representing the vote as the two dimensional configuration. Any senator with a projection below the intersection is predicted to support the measure and any one with a point above it is predicted to oppose it. Consequently, the projection represents the substantive civil rights dimension.[43]

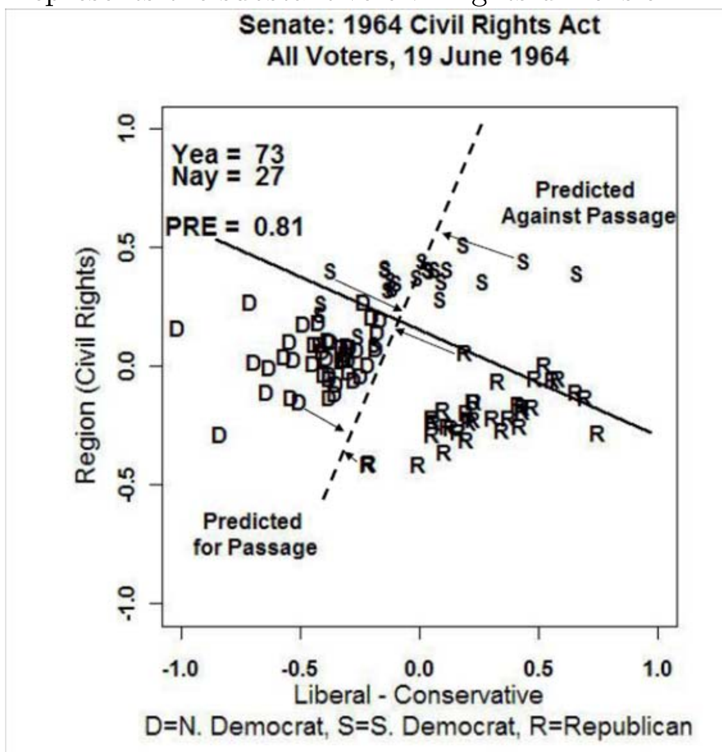

Figure 7: The Spatial Mapping of the Senate Vote on the 1964 Civil Rights Act

In practice, however, it will usually be unnecessary to compute this projection. If one wants simply to estimate a regression model of a vote on NOMINATE scores, including the coordinate for each dimension is usually sufficient. The coefficient on the second dimension coordinate would be properly interpreted as the effect of the second dimension holding positions on the first dimension constant. If one, however, wants a single summary measure of positions on a substantive issue, computing the projection is the appropriate way to proceed.

[43] If one wants to use more than one vote in constructing the measure, one would project to a line perpendicular to the average cut line of the votes.

11 Directions for the Future

Clearly ideal point estimation has become a large and influential literature with a broad impact on the study of institutions in the United States and the world. Certainly there remain open questions about the implementation and interpretation of the basic spatial model. But I suspect much of the efforts moving forward will involve estimating more complex and theoretically-driven models of legislative behavior. In particular, work on strategic voting and logrolling has been limited. Although the basic model can accommodate simple forms of strategic voting, richer models of strategic voting, such as those under incomplete information, are just now being considered. Similarly, logrolling and vote trading are inconsistent with the independence assumptions underlying the basic model. How vote trading affects ideal point estimation and confounds the evaluation of dimensionality clearly needs further work. The use of scaling models to detect vote trading and other forms of legislator coordination may also be another important avenue for work. Finally, it seems that the real potential of Bayesian MCMC has yet to be reached in this area. Its ability to estimate hierarchical models of ideal points and alternatives with additional sources of data is far from exhausted.

While such technical advances will be welcome, even more work is yet to be done in terms of interpreting the implications of ideal point estimates for our understanding of politics. Obviously, issues as to whether ideal points are reflections of member ideology, party pressure, or constituency interest are far from resolved. Moreover, very little is known about the dynamics of how issues map (or not) into the major dimensions of conflict over time. Progress on this front will entail significantly more attention to methodological challenges of estimating policy outcomes in addition to ideal points.

References

[1] Albert, James. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling."*Journal of Educational Statistics* 17: 251-69.

[2] Albert, James H., and Siddartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88: 669–79.

[3] Bailey, Michael and Kelly Chang. 2001. "Comparing Presidents, Senators, and Publications Justices: Inter-institutional Preference Estimation." *Journal of Law, Economics and Organization* 17(2): 477-506.

[4] Bailey, Michael. 2007. "Comparable Preference Estimates across Time and Institutions for the Court, Congress and Presidency." *American Journal of Political Science* 51(3): 433-448.

[5] Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2009. "The Structure of Utility in Spatial Models of Voting." Presented at the Fourth Annual Conference on Empirical Legal Studies, Los Angeles.

[6] Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. n.d. "Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Tests." forthcoming *Legislative Studies Quarterly.*

[7] Clinton, Joshua D. 2007. "Lawmaking and Roll Calls," *Journal of Politics* 69:457-469.

[8] Clinton, Joshua D., Simon D. Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data: A Unified Approach." *American Political Science Review* 98:355-370.

[9] Clinton, Joshua D. and Adam H. Meirowitz. 2003. "Integrating Voting Theory and Roll Call Analysis: A Framework." *Political Analysis* 11:381-396.

[10] Clinton, Joshua D. and Adam H. Meirowitz. 2004. "Testing Accounts of Legislative Strategic Voting." *American Journal of Political Science* 48:675-689.

[11] Converse, Phillip. 1964. "The Nature of Belief Systems in the Mass Public." In David E. Apter, ed. *Ideology and Discontent* New York: Free Press, 206-61.

[12] Cox, Gary W. and Mathew D. McCubbins. 2005. *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives.* New York: Cambridge University Press.

[13] Gilligan, Thomas W. and Keith Krehbiel. 1987. "Collective Decision-making and Standing Committees: An Informational Rationale for Retrictive Amendment Procedures." *Journal of Law, Economics, and Organizations* 3(2):287-335.

[14] Glazer, Amihai and Marc Robbins. 1985. "Congressional Responsiveness to Constituency Change." *American Journal of Political Science* 29:259-272.

[15] Groseclose, Tim, Steven D. Levitt and James M. Snyder Jr. 1999. "Comparing Interest Group Scores across Time and Chambers: Adjusted ADA Scores for the U.S. Congress." *American Political Science Review* 93(1):33-50.

[16] Heckman, James and James Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics* 28: S142-89.

[17] Hix, Simon, Abdul Noury and Gerard Roland. 2006. *Democratic Politics in the European Parliament.* Cambridge: Cambridge University Press.

[18] Jackman, Simon. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking." *Political Analysis* 9: 227–41.

[19] Kalandrakis, Tasos. n.d. "Rationalizable Voting." forthcoming, *Theoretical Economics.*

[20] Krehbiel, Keith and Douglas Rivers. 1988. "The Analysis of Committee Power: An Application to Senate Voting on the Minimum Wage." *American Journal of Political Science.* 32(4):1151-1174.

[21] Krehbiel, Keith. 1989. *Informational and Legislative Organization.* Ann Arbor: University of Michigan Press.

[22] Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking.* Chicago: University of Chicago Press.

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

[24] Levitt, Steven. 1996. "How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology." *American Political Science Review* 83(3):425-441.

[25] Londregan, John. 2000a. "Estimating Legislators' Preferred Points." *Political Analysis* 8(1):35-56.

[26] Londregan, John. 2000b. *Legislative Institutions and Ideology in Chile.* New York: Cambridge University Press.

[27] Martin, Andrew and Kevin Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2): 134-153.

[28] McCarty, Nolan and Adam Meirowitz. 2006. *Political Game Theory: An Introduction.* New York: Cambridge University Press.

[29] McCarty, Nolan and Keith T. Poole. 1995. "Veto Power and Legislation: An Empirical Analysis of Executive-Legislative Bargaining from 1961-1986." *Journal of Law, Economics, & Organization* 11(2):282-312.

[30] McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 1997. *Income Redistribution and the Realignment of American Politics* Washington D.C.: AEI Press.

[31] McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2000. "The Hunt for Party Discipline in Congress." *American Political Science Review* 95:673-687.

[32] McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, MA: MIT Press.

[33] Morgenstern, Scott. 2004. *Patterns of Legislative Politics: Roll Call Voting in the Latin America and the United States.* Cambridge: Cambridge University Press.

[34] Kalt, Joesph P. and Mark A Zupan. 1990 "The Apparent Ideological Behavior of Legislators: Testing for Principal-Agent Slack in Political Institutions." *Journal of Law and Economics* 33:103-131.

[35] Londregan, John B. 2000a. "Estimating Legislators' Preferred Points." *Political Analysis* 8(1):35-56.

[36] Londregan, John B. 2000b. *Legislative Institutions and Ideology in Chile.* New York: Cambridge University Press.

[37] Lott, John R., Jr., and Stephen G. Bronars. 1993. "Time Series Evidence on Shirking in the U.S. House of Representatives." *Public Choice* 74:461–484.

[38] Lewis, Jeffrey B. and Keith T. Poole. 2004. "Measuring Bias and Uncertainty in Ideal Point Estimates via the Parametric Bootstrap." *Political Analysis* 12:105-127.

[39] Ordeshook, Peter C. 1986. *Game Theory and Political Theory.* New York: Cambridge University Press.

[40] Poole, Keith T. and Howard Rosenthal. 1985. "A Spatial Model for Legisaltive Roll Call Analysis." *American Journal of Political Scinece* 29:357-384.

[41] Poole, Keith T. 1998. "Recovering a Basic Space From a Set of Issue Scales." *American Journal of Political Science*, 42(3): 954-993.

[42] Poole, Keith T. 2000. "Non-parametric Unfolding of Binary Choice Data." *Political Analysis* 8(3):211-237.

[43] Poole, Keith T. 2001. "The Geometry of Multidimensional Quadratic Utility in Models of Parliamentary Roll Call Voting." *Political Analysis* 9, 211–226.

[44] Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting.* New York: Cambridge University Press.

[45] Poole, Keith T. 2007. "Changing Minds? Not in Congress." *Public Choice*, 131:435-451.

[46] Poole, Keith T. and R. Steven Daniels. 1985. "Ideology, Party, and Voting in the U.S. Congress, 1959-1980." *American Political Science Review* 79(2): 373-399.

[47] Poole, Keith T. and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35:228-278.

[48] Poole, Keith T. and Thomas Romer. 1993. "Ideology, Shirking, and Representation." *Public Choice* 77:185-196.

[49] Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political Economic History of Rollcall Voting.* New York: Oxford University Press.

[50] Rothenberg, Lawrence S., and Mitchell S. Sanders. 2000. "Severing the Electoral Connection: Shirking in the Contemporary Congress." *American Journal of Political Science* 44:316–325.

[51] Voeten, Erik. 2000. "Clashes in the Assembly," *International Organization* 54(2): 185-215.

[52] Shor, Boris and Nolan McCarty. 2010. "The Ideological Mapping of the American States." Typescript, University of Chicago.

[53] Snyder, James M. Jr. 1992. "Artificial Extremism in Interest Groups Ratings." *Legislative Studies Quarterly* 17:319-345.

[54] Snyder, James and Michael Ting, "Party Labels, Roll-Call Votes, and Elections." *Political Analysis*, 11(4):419-444.

[55] Wright, Gerald C. and Brian F. Schaffner. 2002. "The Influence of Party: Evidence from the State Legislatures." *American Political Science Review* 96(2): 367-379.