(HOW) CAN WE ESTIMATE THE IDEOLOGY OF CITIZENS AND POLITICAL ELITES ON THE SAME SCALE?

STEPHEN A. JESSEE

THE UNIVERSITY OF TEXAS AT AUSTIN

Abstract. The estimation of the ideology of political elites such as candidates and elected officials on the same scale as that of ordinary citizens has been shown to have great potential to provide new understandings of voting behavior, representation and other political phenomena. There has been limited attention, however, to the fundamental practical and conceptual issues involved in these scalings or to the sensitivity of these estimates to modeling assumptions and data choices. I show that the standard strategy of estimating ideal point models for preference data on citizens and elites can suffer from potentially problematic pathologies. This paper addresses these issues and presents a modeling approach that can be used to investigate the effects of modeling assumptions on resulting estimates and also to impose restrictions on the ideological dimension being estimated in a straightforward way.

Prepared for the Ideal Point Modeling Conference, MIT May 2015. Please send all comments and suggestions to sjessee@utexas.edu.

1. Introduction

Many theories and hypotheses in political science deal with the ideological positions of citizens in relation to those of candidates, elected representatives, or other political elites. In recent decades, a large body of research has been devoted to the measurement of the ideology of legislators based on their roll call voting in Congress. Furthermore, several measures of citizen ideology have a long history in the political behavior literature, often as a result of being based on identical survey questions that have been asked in similar or identical form for decades. Until recently, though, it has been difficult to directly compare the ideologies of candidates or elected officials with those of ordinary citizens on the same scale.[1] Recent work has relied on new combinations of survey and statistical techniques in order to estimate the ideology of survey respondents and political elites on the same scale. While these new measures have shown great potential to provide new insights in areas such as voting behavior and representation, it is important to verify their properties and ensure that their results are reasonable.

New approaches have leveraged survey questions that can, in one way or another, be matched to positions taken by legislators or candidates for office in order to estimate the ideologies of these two groups on the same scale (see e.g. Jessee [2009], Bafumi and Herron [2010], Shor [2009]). Applications of joint scaling have also estimated ideological positions for other types of actors based on many different types of data. These include Groseclose and Milyo [2005], who estimate the positions of media outlets alongside members of Congress, Bailey [2007], who estimates a single ideological scale for courts, Congress and the president across time, Bonica [2014], who estimates the positions of candidates and donors from campaign contribution data, and Tausanovitch and Warshaw [2014] who estimate the ideological positions of municipal governments and their constituencies. These data are often analyzed using the same ideal point models which have been used in the past to estimate the ideology of members of Congress based on their roll call voting behavior.

[1]Some older work has produced estimates of candidate or legislator ideology on the same scale as citizens using data on citizens' perceptions of candidate or legislator ideology (Aldrich and McKelvey 1977, Brady and Sniderman 1985, Alvarez and Nagler 1995, Adams et al. 2005) or by scaling citizen preferences for or ratings of candidates (Weisberg and Rusk 1970, Cahoon and Ordeshook 1978). But bridging techniques that rely directly on policy positions without making strong assumptions about perceptions or preference formation, have received attention relatively recently.

Under the assumptions of these models, the estimates produced allow researchers to go beyond associational measures that have been cited as problematic in testing hypotheses about representation and other issues (see Achen 1977) to examine the actual correspondence between the ideological positions of citizens and elites. The results of these studies have provided important insights into areas such as representation, candidate positioning and spatial voting theory.

But the methods used to obtain these measures rely on several important assumptions. In particular, it is usually assumed that there is a single ideological dimension that structures both citizen and legislator views across different policies. Furthermore, it is assumed that citizens and legislators having the same ideological position will have the same probability of supporting a given proposal. While verifying these assumptions is potentially difficult, it is important for establishing the validity of ideology measures that have been produced. At the least, it would be desirable to show that any violations of these assumptions tend to have a minimal impact on our inferences about the ideologies of legislators and citizens. The literature to date has paid little attention to these concerns. In fact, the standard approach to joint scaling involves applying some sort of scaling procedure to a dataset that includes including bridge observations, assuming (usually tacitly) that the model will estimate the correct dimension—i.e. the dimension relevant for the theory or hypothesis under study.

This paper addresses these issues, beginning by asking whether joint scalings of citizens and legislators are robust to seemingly innocuous factors such as the number of respondent included in the data. I analyze two different datasets in which the positions of ordinary citizens are measured on the same issues as those of elected officials, identifying potential problems with the standard approach to joint scaling. These findings suggest some general guidance as to how researchers should design surveys with the aim of jointly scaling citizens and legislators. More broadly, I consider what should be done in the face of discrepancies between the structure of ideology in different groups—does this render the entire enterprise of joint scaling futile or does there remain a useful way forward? I introduce an approach for estimating the ideology of members of multiple groups on the same scale under the constraint that the ideological dimension is structured based on the data from one particular group. This approach can be used to assess the similarity between the ideological dimensions underlying the preferences of citizens and legislators as well as to impose desired structure on estimates. I conclude by arguing that while it is centrally important for researchers to explore the validity of joint scaling assumptions and results, the question of what dimension is studied should ultimately be driven by substantive and theoretical concerns.

2. Jointly Scaling Groups with "Bridging" Ideal Point Analyses

The basic idea underlying ideal point analyses is that an ideological space, typically consisting of one or a small number of dimensions, underlies the revealed preferences or actions of political actors. These data are seen as indicators, which are generated stochastically based on each actor's underlying ideal point and the characteristics of the policies being voted on. Ideal point models thus provide a way to uncover a latent space that structures the preferences of the actors under study, reducing a large number of variables into a single-dimensional or low-dimensional representation of preferences.

Once a set of indicators has been chosen that is thought to tap the latent trait of interest, researchers must choose a model and method for estimating these underlying values. Most recent work scaling respondents and legislators together has utilized the ideal point model found in Clinton et al. [2004] (CJR). In practice, the specific form of the ideal point model tends to have only a minor impact on the resulting estimates. Other alternatives include NOMINATE [Poole and Rosenthal, 1985] and factor analytic techniques [Heckman and Snyder, 1997]. Because it has been most commonly used in recent studies involving joint scalings, I focus on the CJR model here.

The CJR ideal point model assumes that each actor, indexed by i, casts votes on a series of proposals, indexed by j based on quadratic utility functions over alternatives subject to independent normal disturbances (errors) so that $U_i(\zeta_j) = -\|x_i - \zeta_j\|^2 + \eta_{ij}$ and $U_i(\psi_j) = -\|x_i - \psi_j\|^2 + \nu_{ij}$ are respondent i's utilities for the yea and nay (support and oppose) alternatives, respectively, on policy j. Here, x_i is respondent i's most preferred policy position, often called her ideal point, ζ_j and ψ_j are the positions of the yea and nay alternatives (often proposal and status quo locations) on policy j, and η_{ij} and ν_{ij} are error terms assumed to be normally distributed with the same mean and independent of each other and across all i and j. It is straightforward to show that under these assumptions the likelihood for any given vote can be written as

$$P(y_{ij} = 1) = \Phi(x_i \beta_j - \alpha_j) \tag{1}$$

where y_{ij} equals 1 if respondent i votes yes on proposal j and equals 0 if he votes against it, with all votes being independent conditional on x_i, β_j and α_j. This correspondence is established by letting $\beta_j = 2\left(\zeta_j - \psi_j\right)/\sigma_j$ and $\alpha_j = \left(\zeta_j^2 - \psi_j^2\right)/\sigma_j$ where σ_j is the variance of $\eta_{ij} - \nu_{ij}$.[2] In the model's standard form, it is assumed that $\sigma_j = 1$ for all policies. Here the model is estimated in a Bayesian framework using a Gibbs sampler with independent standard normal priors on the actor ideal points x_i and independent normal priors with mean zero and variance 25 on proposal discrimination and difficulty parameters β_j and α_j.[3]

The interpretation of the item parameters in equation 1 can be simplified by transforming to obtain the model

$$P(y_{ij} = 1) = \Phi\left[\beta_j \left(x_i - \gamma_j\right)\right] \tag{2}$$

where $\gamma_j = {}^{\alpha_j}/_{\beta_j}$ now represents the cutpoint for vote j, which lies halfway between the "yea" and "nay" alternatives (or "support" and "oppose" positions) for a given policy. This is therefore the ideological position at which an actor would be indifferent between supporting and opposing policy j and thus would have a 50 percent chance of voting either way. The discrimination parameters β_j have an identical meanings under the specifications in equations 1 and 2, indicating the strength and direction of the relationship between an actor's ideal point x_i and their likelihood of supporting the policy.

[2]These expressions are for the unidimensional case. In multiple dimensions, β_j remains the same but $\alpha_j = (\zeta_j'\zeta_j - \psi_j'\psi_j)$.
[3]The model is estimated using the ideal function from the pscl library in R Jackman [2009].

In studies attempting to estimate the ideological positions of actors of multiple types (e.g. legislators and ordinary citizens) on the same scale, it is typically necessary to observe common items between the two groups. Without this requirement, one can only compare estimated ideologies within groups, but not between them. To take an education testing example, if two classrooms of students take different tests, with no overlap between the questions, the grades on these exams can tell us which students are the smartest in each classroom, but not whether a student in the first classroom is smarter than a student in the second classroom. Modifying this testing example, if the exams taken in the two classrooms were the same, or even if there was simply overlap between the questions on the two different exams, we gain the leverage needed to compare our estimates of student intelligence not just within the two classrooms, but now also between them.

In the context of ideal point estimation, this means that in order to compare the ideological positions of actors from two different groups, we must find examples in which the members of the two groups state their preferences on the same policy proposals. It is common to call these exercises bridging ideal point analyses, with the policies asked of both groups called bridge items. Under this setup ζ_j and ψ_j, and therefore β_j and α_j or γ_j, are the same for the two groups on each of the bridging items. This allows researchers to pool members of the two groups together and estimate their ideology on the same scale. A key assumption here is that ideological space underlying the preferences of the two groups is structured in the same way.

Following equation 3, we can consider two (possibly identical) ideological spaces with each defined by the relationship between actors' ideological position x_i and the likelihood of supporting each policy proposal.[4] In other words, these spaces can be defined by β_j and γ_j, which are now allowed to vary not just across policy items j, but also across groups so that we obtain

[4]It is also possible to have more than two classes of actors, each with their own ideological space.

$$P(y_{ij} = 1) = \Phi\left[\beta_{g(i),j}\left(x_i - \gamma_{g(i),j}\right)\right] \tag{3}$$

where $g(i)$ is the group of actor i.

If the item parameters are different for the two groups, it is not clear how we can compare the ideal points between the two groups. There are multiple reasons why we may worry about this. For example, if a given application uses survey questions about specific votes in Congress, we may worry that the questions do not correspond perfectly with the policies being voted on. This could be because of the wording of the questions or the context in which the decisions are being made by survey respondents as opposed to legislators. Alternatively, it could be that even though the actual items are identical (or nearly so) between the two groups, the structure of the ideological dimension is simply different. For example, support for a certain policy could be strongly related to ideological position for members of Congress, but not for ordinary citizens. One could imagine a vote on a relatively obscure piece of legislation that political elites would have sharply divided, views about, but which the vast majority of ordinary voters would not understand, let alone view in ideological terms.

One way to think of the assumptions underlying these bridging estimations is very rigidly—either the item parameters are all exactly equal between the two groups or they are not. Testing this point null hypothesis is one way to answer the question of whether the ideologies of two types of actors can meaningfully be compared on the same scale. If this hypothesis is rejected, one might conclude that we simply cannot compare estimate a the two groups' ideologies on a comparable scale. For example, Lewis and Tausanovitch [2013] analyze the data from two previously published studies in which the ideologies of ordinary citizens are jointly scaled alongside those of elected officials, rejecting the hypothesis that the item parameters are equal between the two groups for each of the two datasets analyzed.

But this sharp null hypothesis testing approach is quite a strong one. In particular, it takes very literally a choice model that is intended as an approximation of the process by which people take positions on various policies. In this case, the idea that every actor will have exactly the same structure to his or her ideological preferences seems *ex ante* implausible, even within, not just between groups. For example, in Congress, it is clear that different types of members (e.g. Tea Party Republicans or Blue Dog Democrats) have different structures underlying their preferences. Applying such a test to legislators grouped by party or by region of the country (especially south versus northeast for example) would also be expected to easily reject the hypothesis of strict equality of item parameters between groups.[5]

The political science literature on latent traits estimation of political ideology includes many examples of choosing parsimony over complexity, even when a literal interpretation of the model relying on formalized hypothesis tests might suggest a different strategy. For example ideal point models of congressional voting typically assume only one or two dimensions [Poole and Rosenthal, 1985, Clinton et al., 2004], citing this as a useful balance of explanatory power and parsimony. But one could also consider the dimensionality of congressional roll call voting very literally, with the number of dimensions as a parameter to be estimated. For example, Tahk [2005] uses a cross-validation approach to estimate the number of dimensions across a range of congressional sessions, finding between three and eight dimensions for most years, with some sessions of Congress estimated to have as many as fifteen dimensions that add meaningful explanatory power.

This raises the question of what to take away from such an exercise. If we test whether a model is literally and exactly true in a situation where we know it is only an approximation to reality and in which we have a large amount of data, it is essentially a foregone conclusion that we will reject the null hypothesis. In one sense, the model itself can be thought of as a data reduction tool, condensing the many different policy positions of each actor into a single ideal point that summarizes their ideological position. The question

[5] Another way to put this is that if we were to define groups so that the structure of ideology is exactly the same for members in any given group, we would need many groups even within Congress, perhaps going so far as to put each legislator in his or her own group, rendering the whole exercise of estimating ideological positions essentially pointless.

here, following the classic quote, isn't whether the model and its resulting ideal point estimates spanning multiple types of actors are wrong, but whether they are useful.

3. Data: Senate Representation Study

The Senate Representation Study presents a particularly good test case for bridging ideal point analyses. Fielded between December 2005 and January 2006, the survey includes policy questions written to correspond to specific roll call votes that had taken place in the Senate during 2004 and 2005.[6] The survey was administered online to 5,871 respondents from the Polimetrix (now YouGov/Polimetrix) online panel. The sample was not constructed to be representative at the national level. In particular, because one of the aims of the study was to analyze respondent perceptions of their senators, at least 100 respondents from each state were included in the sample. Therefore, the survey is not representative at the national level. The sample also includes higher levels of political information on average than nationally representative surveys such as the 2004 ANES and includes fewer weak partisans and minorities. Further information about the survey is available from the author upon request. A list of the 27 policy questions analyzed here is shown in Table 1.

The Senate Representation Survey is particularly well suited for examining the assumptions of bridging ideal point modeling for several reasons. First, it contains a large number of questions on a wide range of policies that were vote on in the Senate. These items provide direct bridges between the positions taken by senators and survey respondents. The policies, furthermore, include more mainstream issues such as gun control, environmental protection and raising the minimum wage, which respondents might be expected to have opinions on, as well as more obscure policies such as bankruptcy reform and overtime regulations, on which respondents might not have well thought out views. In this way, the Senate Representation Survey might be thought to represent a "hard test" for joint scaling. The variety of different policy types included

[6]In order to focus on only policy items, two questions on Supreme Court nominees were dropped. One policy item restricting ammunition sales was also dropped because it was a Republican substitute to a stronger Democratic measure and therefore was likely to be perceived very differently by respondents and legislators.

Table 1. List of Senate Votes Used in Senate Representation Survey

	Bill Number	Title	Senators Yea-Nay Votes	Respondents Y-N-DK %
–	HR 4250	Jumpstart Our Business Strength Act	78-15	44-32-23
	S. Amdt. 1085 to HR 2419	Remove Funding for "Bunker Buster" Nuclear Warhead	43-53	52-41-8
–	S 1307	Central American Free Trade Agreement	61-34	45-39-15
–	S 256	Bankruptcy Abuse Prevention and Consumer Protection Act	74-25	54-30-16
	S. Amdt. 367 to HR 1268	Remove Funding for Guantanamo Bay Detention Center	27-71	46-45-9
+	HR 1308	Working Families Tax Relief Act	92-3	79-10-12
	S. Amdt. 2937 to HR 4	Child Care Funding for Welfare Recipients	78-20	50-38-13
	S. Amdt. 1026 to HR 2161	Prohibiting Roads in Tongass National Forest	39-59	56-31-13
	S. Amdt. 1626 to S 397	Child Safety Locks Amendment	70-30	75-21-4
–	S. Amdt. 3584 to HR 4567	Stopping Privatization of Federal Jobs	49-47	50-35-16
–	S. Amdt. 3158 to S 2400	Military Base Closure Delays	47-49	48-36-16
+	S. Amdt. 44 to S. 256	Minimum Wage Increase	46-49	67-29-4
–	S 397	Protection of Lawful Commerce in Arms Act	65-31	74-19-6
	S. Amdt. 2799 to S. Con. Res. 95	Cigarette Tax Increase	32-64	59-37-4
	S. J. Res. 20	Disapproval of Mercury Emissions Rule	47-51	71-12-17
	S. Amdt. 278 to S. 600	Family Planning Aid Policy (Mexico City Policy)	52-46	50-44-6
+	S. Amdt. 2807 to S. 600	Raise Tax Rate on Income over One Million Dollars	40-57	62-32-6
+	S. Amdt. 3379 to S. 2400	Raise Tax Rate on Highest Income Bracket	44-53	49-44-6
+	HR 1997	Unborn Victims of Violence Act	90-9	68-24-9
+	S. Amdt. 3183 to S. 2400	Federal Hate Crimes Amendment	65-33	49-42-9
	S. Amdt. 902 to HR 6	Fuel Economy Standards	28-67	70-22-8
	S. Amdt. 826 to HR 6	Greenhouse Gas Reduction and Credit Trading System	38-60	48-36-16
+	S. Amdt. 1977 to HR 2863	Banning Torture by U.S. Military Interrogators	90-9	57-38-5
	S. Amdt. 1615 to S. 397	Broaden Definition of Armor Piercing Ammunition	31-64	70-22-8
+	S. Amdt. 168 to S. Con. Res. 18	Prohibit Drilling in Arctic National Wildlife Refuge	49-51	48-48-4
–	S. Amdt. 3107 to S. 1637	Overtime Pay Regulations	52-47	44-44-12
–	S. 5	Class Action Fairness Act	72-26	53-22-24

Table shows Senate vote totals and percentages of 2004 survey respondents supporting, opposing, and saying "don't know" to each surveyed policy. Leftmost column shows coding of "easy" and "hard" issues issues, represented by + and -, respectively.

in the survey also provide a useful test case for which type(s) of items may be the most useful and which may be the most problematic in bridging applications.

4. Assessing the Performance of Joint Scaling

One way to compare the estimated ideal points of respondents and senators under different scalings is to estimate the ideal point model in equation 2 separately for respondents and for senators, then compare the estimated ideal points (x's) from these separate scalings to those from a full joint scaling of these two groups together. This exercise, produces extremely high correlations between separate and joint ideal point estimates from the Senate Representation Study data: .98 for the estimated ideal points of senators and well over .99 for respondents.[7] At first glance, these high correlations might be thought to indicate that these bridging estimates are well behaved. But these correlations do not tell us how close to being equal these sets of estimates actually are for two reasons. First, correlation is only a measure of linear association, not equality. Second, and more fundamentally for our purposes, the values on the estimated scales from the joint and separate estimations are not directly comparable.[8] Therefore, we need other techniques in order to assess the viability of this joint scaling exercise.

Another way to think about the estimated dimension in joint scaling applications is to view it as a compromise, loosely speaking, between the dimension structured by each of the groups being analyzed, here respondents and senators. The degree of compromise in a given joint scaling application—how close the jointly estimated dimension is to the separately estimated dimensions for each group—is dictated by the model's fit to the data under different parameter values. When the dimensions underlying the views of the two different groups differ, the fit of the pooled model can be dramatically affected by factors that are not central to the phenomena under study, but are instead external or arbitrary.

A useful thought experiment is to consider what would happen to the estimated ideological dimension if the ratio of the number of respondents to senators in our dataset were different. Because this ratio is dictated mostly by factors apart from the underlying political dynamics we seek to study, we should hope that it is not strongly impacted by it. For example, the number of respondents may depend on things like the amount of money a scholar has in her research account, and while more data are generally better because they tend to provide more precise estimates, we should be concerned if these things directly impact the characteristics of our estimates.

[7]All estimates based on the standard CJR model are produced using the ideal function in the pscl library in R Jackman [2009]. Each set of estimates here was based on 250,000 iterations of the sampler, discarding the first 50,000 as burn-in and recording every 50th iterations thereafter. Evidence of convergence was strong after several thousand iterations, but the model was run for much longer as a conservative strategy.

[8]The scales would be comparable if, for example, we fixed some item parameters to the same specific values across these three scalings. But doing this would assume that the ideology scales were the same for the individual groups, at least on those items, which is undesirable given that this is what is being tested.

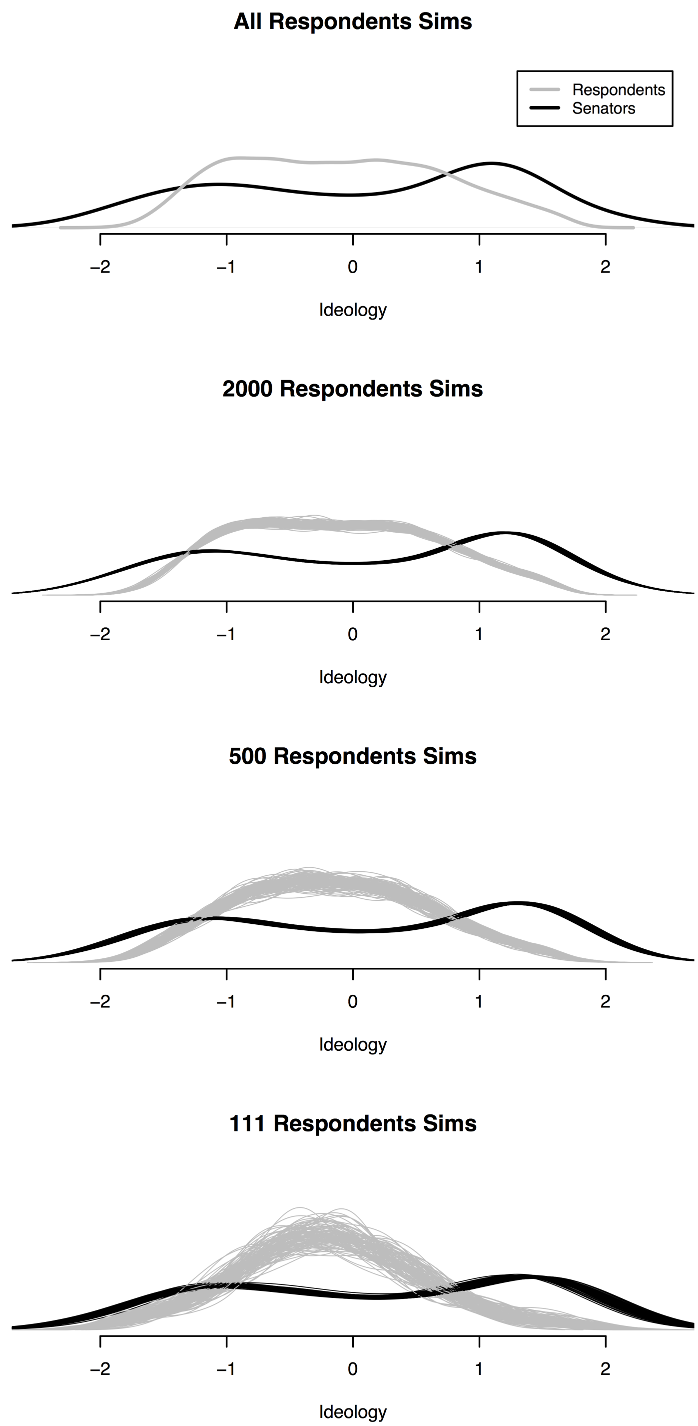

FIGURE 1. *Characteristics of Respondent and Senator Ideology Estimates Differ Sharply Based on Respondent Sample Size Used.* Panes show densities of estimated ideal points for senators and respondents from joint ideal point models estimated for random samples of given size from respondents along with all 111 senators in the dataset. Top pane shows estimates using all 5,871 respondents, while lower panes show densities for 100 different estimates, each using a different random sample of respondents of the specified size.

Although we cannot create new respondents for the already-fielded survey, we can drop respondents to create a smaller dataset consisting of a different balance of respondents to senators. The lower three panes of Figure 1 do just this, comparing the estimates from the full joint model using all 5,871 respondents to those using only 1,000, 500 and 111 respondents, respectively, the last of these being equal to the number of senators used in the scaling. For each simulation, a given number of respondents is randomly sampled without replacement from the full survey sample. The ideal point model is then estimated using these respondents along with all 111 senators, pooling them together and assuming a single common ideological dimension. For each of these sample sizes, one hundred such simulations are run, each using a new random sample of respondents of the specified size to ensure that the results are not driven by the particular subset of respondents chosen for a given sample size.[9]

Looking at Figure 1 it is clear that the overall character of the estimated ideal points changes systematically with the number of respondents. As the number of respondents gets smaller, thus creating a more equal ratio of respondents to senators, the estimated ideal points of respondents appear more moderate relative to those of senators. The logic behind this is that the estimated dimension in these models is a compromise between the senator and respondent dimensions. When respondents constitute the overwhelming majority of the data as in the top pane of Figure 1, the item parameters are estimated largely based on respondent choices. When the number of respondents and senators in the data become more equal as in the lower panes of Figure 1, the item parameters are estimated based on a more equal compromise between the structure of the two groups' ideological dimensions. In order for an actor to have an ideological extreme ideal point estimate, he must take either consistently liberal or consistently conservative positions. This becomes less likely as the ideological dimension is estimated more based on the structure of another group's views. Therefore all else equal, the distribution of ideology in a given group tends to look more moderate as the ideological dimension is structured more based on the preferences of a different group of actors.

[9]Because of the large number of estimations, the sampler for each is run for 25,000 iterations, with the first 5,000 discarded as burn-in and every iteration thereafter recorded. The sampler appeared to converge rapidly and the recorded samples appear to provide a reasonable amount of information, particularly since only posterior means (not variances of HPDs) are examined.

The systematic variation in the overall character of the estimates shown in Figure 1 is obviously not desirable. The way in which the model produces the ideological space is affected by the seemingly irrelevant factor of the number of respondents in the survey. A researcher who runs a survey with a large number of respondents would learn something different about, for example, the relative polarization of respondents and senators, than someone who ran a smaller survey. This is not just due to the extra uncertainty that comes from a smaller dataset. It is clear from these simulations that the sampling distributions for these two polarization estimates—the "large dataset" and "small dataset" examples—are quite different.

5. A Method for Group-Based Ideology Estimation

If the ideological dimensions underlying the preferences of two groups are structured differently, the appropriate response may in fact be dictated by theoretical, rather than statistical, concerns. Depending on the theory being tested or the question being asked, the dimension of interest may be the one used by one group in particular. For example, if one is seeking to test theories of spatial voting in elections (e.g. Hotelling 1929, Downs 1957, Enelow and Hinich 1984), the relevant ideological dimension may be the one underlying the views of ordinary citizens (or voters) rather than the dimension structured around the positions taken by candidates or elected officials or the dimension that results from simply pooling candidates and citizens together and estimating a single compromise dimension between them. This is because spatial voting theory posits that voters choose candidates who are ideologically closer to their own positions. Because the voter is evaluating the ideological proximity of the candidate, we might look at ideological proximity based on the ideological dimension understood by voters.

Researchers, then, may want a technique that can impose such structure on the ideological dimension in a given ideal point estimation, rather than simply pooling all of the data together and using whichever dimension is estimated by the model. This motivation is all the more important given that ideal point estimation bridging two groups can suffer from the pathologies illustrated above. We may also want to compare the estimated ideal points and item parameters structured by each group separately as a diagnostic exercise, seeking to understand whether the dimensions underlying the preferences of each group differ meaningfully. This section describes a technique for estimating the ideal points of a set of actors from multiple groups while restricting the estimated ideological dimension to be structured by the positions of a specific subgroup of actors.

The CJR ideal point model discussed above is estimated using a Gibbs sampler that produces draws from the posterior distribution over the model's unknown parameters given the observed data. This is accomplished by cycling through samples from the conditional posterior distributions for each set of parameters setting all other parameters at their most recently sampled values. In order to conduct a restricted ideal point estimation where the ideological dimension is structured only based on the preferences of one group, this process can be modified so that the item parameters β and γ are sampled at each iteration from the conditional posterior given the ideal points x and latent utility differences y^* of one particular group. In other words, the sampling procedure is identical to the one used in CJR except that inferences about the item parameters, which structure the underlying ideological dimension, are affected only by the policy positions of the chosen subgroup of actors. This procedure is equivalent to running the standard model on only the data from the group structuring the ideological dimension, then mapping the ideological positions of the out of group members into this ideological space by sampling from the conditional posterior of their ideal points given the item parameter values at each iteration of the sampler. One could also imagine analogous procedure for maximum-likelihood-based estimators such as NOMINATE, where the conditional maximization of the item parameters is based only on the ideal points and positions of the chosen group. Online Appendix, Section 1 describes this process in more detail.

Using this technique, it is possible to compare the ideological spaces, including ideal point and item parameter estimates, that structure the preferences underlying each group in a given dataset while still estimating the ideology of all actors in the data and allowing for direct comparisons between the two sets group-based estimates. Figure 2 shows the results of this exercise for the Senate Representation Study, comparing the estimates (posterior means) for senator and respondent ideal points that result from restricting the sampler to let senator or respondent preferences, respectively, structure the underlying ideological dimension.[10] In contrast to the separate estimation strategy discussed above, this group-based scaling produces estimates that can be meaningfully compared on the same scale. This is achieved by imposing the same identifying restriction on the ideal points across the two scalings: at each iteration the estimates, including ideal points (x_i) and item parameters (β_j and γ_j) are rescaled such that the average of the mean respondent ideal point and the mean legislator ideal point is zero and the average of the variance of respondent ideal points and the variance of senator ideal points is 1, and the space is oriented such that higher ideal point values represent more conservative ideological positions.[11] This means that we can assess how close to equal these two sets of estimates are to being equal, not just how strong the relationship is between them.

Because correlation does not speak directly to how close to equal two variables are, this would suggest using a statistic such as the mean squared difference (MSD) between the two sets of estimates. This measure, however, does not have an easily interpretable scale. Here I standardize the measure by dividing by the standard deviations of each variable, and rescale the measure to be bounded between zero and one where one, calling the resulting statistic standardized mean squared difference (sMSD) between the two group-based ideal point scalings which will be defined as

$$\text{sMSD} = \frac{1}{1 + \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_{i,(1)} - x_{i,(2)}\right)^2}{\sigma_{(1)}\sigma_{(2)}}} \tag{4}$$

[10]The group-based ideal point model is estimated using a modified version of the ideal function in the pscl library in R. All estimates are based on runs of 250,000 iterations with the first 50,000 iterations discarded as burn-in, recording every 50th iteration thereafter. All estimations appeared to converge rapidly.

[11]If the ideal points were identified by, for example restricting the mean and variance of one particular group's ideal points to be 0 and 1, respectively, this would artificially make this group's ideal point estimates look more similar between the two scalings. If, by contrast, the identifying restriction simply made the mean and variance of all ideal points (legislators and respondents together) have mean 0 and variance 1, this would be similar to making a respondent-based restriction due to the much larger number of respondents than senators. The particular identifying restriction chosen here allows respondents as a whole and senators as a whole to have the same influence, loosely speaking, on identifying the ideal point space. Although different identifying restrictions do change many of the values calculated below (including, notably, sMSDs) the overall pattern of findings remains the same.

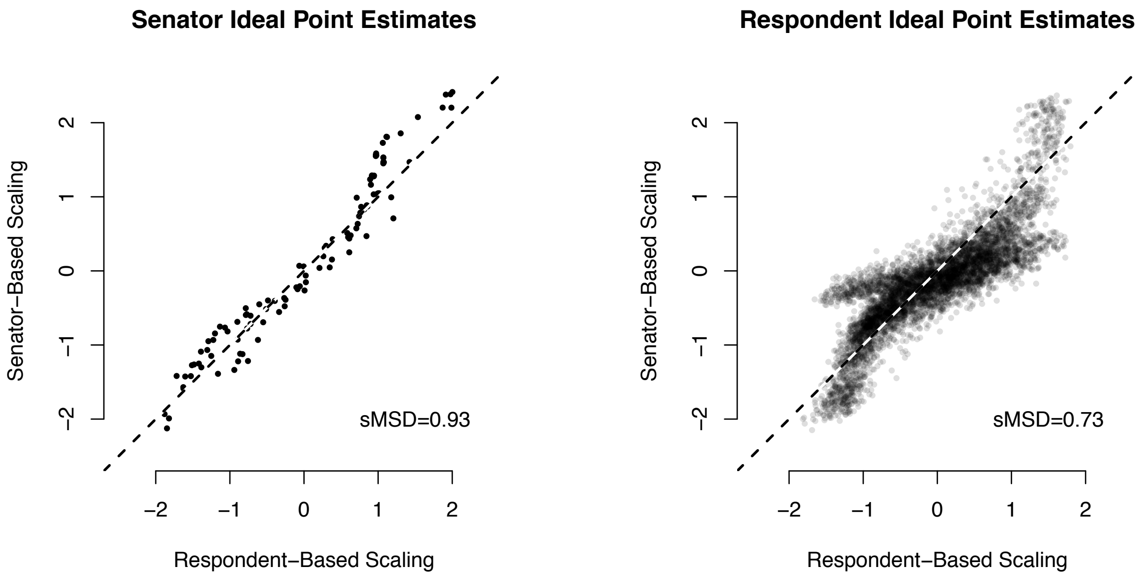

FIGURE 2. *Respondent-Based and Senator-Based Ideal Point Estimates from Senate Representation Study Show Small Differences for Senators, Large Differences for Respondents.* Plot compares ideal point estimates (posterior means) from respondent- and senator-based scalings separately for senators (right pane) and respondents (left pane). Respondent estimates are plotted with transparency to better show overlapping points.

where $x_{i,(1)}$ and $x_{i,(2)}$ are the estimated ideal points for actor i from scalings based on groups 1 and 2, respectively and $\sigma_{(1)}$ and $\sigma_{(2)}$ are the standard deviations of the two sets of estimates. This measure has several useful properties. First, it approaches 0 as the two sets of estimates become farther apart and approaches 1 when $x_{i,(1)}$ and $x_{i,(2)}$ approach equality for each actor i. The measure is also invariant to linear transformations applied to the two scales together, meaning that it is not sensitive to the identifying restriction chosen to fix the two sets of estimates on the same scale.

Looking at Figure 2 it is obvious that while the estimated ideal points for senators are nearly identical whether the dimension is structured based on the preferences of senators or respondents, the estimates for respondents, are less similar. The sMSD for senators is .93, but for respondents it is .73. There are many respondents whose ideal points are quite different depending on which group structures the estimates, being much more moderate in the senator-based scaling, but more ideologically extreme under the respondent-based scaling. The densities of the estimated ideal points under the two group-based scalings, shown in

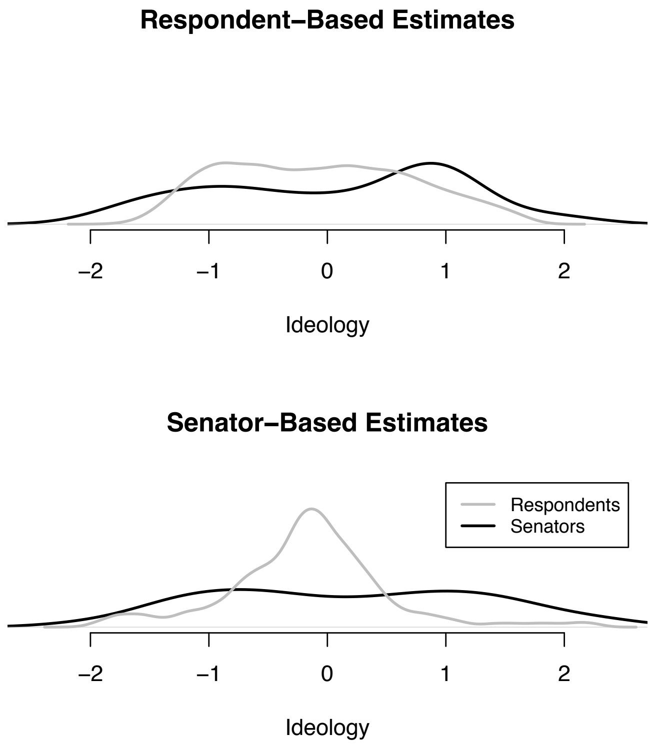

FIGURE 3. *Relative Densities of Ideal Point Estimates from Senate Representation Survey Show Large Differences Under Respondent-Based and Senator-Based Scalings.* Densities of respondent and senator ideal points are plotted from respondent-based and senator-based scalings of all respondents and senators from the Senate Representation Study.

Figure 3 also show significant differences in their overall characteristics. In particular, the senator-based estimates show a much more moderate (less spread out) distribution of respondent ideologies relative to those of senators, while the respondent-based estimates show only a slightly higher variance for the senator ideal point estimates as compared to respondents.

The densities of respondent-based ideal points, plotted in Figure 3, look similar to the pooled joint scaling in the top pane of Figure 1, while the senator-based densities look more similar to those based on all senators and only a subsample of 111 respondents seen in the bottom pane of Figure 1. This makes sense given that higher the proportion of respondents in the data, the more the estimated dimension will be similar to that for respondents. As senators make up a larger proportion of the data, the dimension will tend to be influenced more by the structure of senator preferences. The key advantages of the group-based procedure, however, are that the dimension being estimated can be chosen directly, rather than loosely affected by dropping some number of actors from one group, and also that ideal points are estimated for all actors, whether or not they are members of the group chosen to structure the estimated dimension.

One way to understand why these two sets of estimates differ is to examine the estimated item parameters under the two setups. If the ideological dimensions for senators and respondents are structured similarly, we should observe similar estimates of the item parameters for each policy whether from senator- or respondent-based scalings. Figure 4 shows the estimated discrimination parameters and cutpoints (β_j's and γ_j's from equation 3) for these two scalings. For the discrimination parameters, the posterior means are plotted, while for the cutpoints, the posterior medians are used. This is because the parameter transformation between equations 1 and 2 involves dividing the difficulty parameter α_j by the discrimination parameter β_j for each item j to obtain the cutpoint γ_j. For iterations of the Gibbs sampler in which β_j is near zero for a given item, the value of γ_j can become extremely large in magnitude. As the number of iterations of the sampler used approaches infinity, this will average out, but even with the extremely long runs used here, a small number of extremely large or extremely small draws for some γ_j's can dramatically affected the posterior means. By contrast, the posterior means and medians are nearly identical for item parameters β_j. Online Appendix, Section 2 provides more information about this issue and the characteristics of the posterior means and medians for each item parameter.

The discrimination parameters (β_j's) in Figure 4 show a positive association, but most do not line up close to the forty-five degree line indicating equality between the two sets of estimates. Although the signs of the discrimination parameter estimates are the same for 25 out of the 27 items, the estimates exhibit little if any association beyond this, suggesting that while policies seen as liberal (conservative) by senators also tend to be seen as liberal (conservative) by respondents, the degree of ideological distance perceived between supporting and opposing the policies does not seem to be similar for the two groups. To put it differently, the degree of ideological divisiveness for each of the items is not strongly associated between the two scalings.

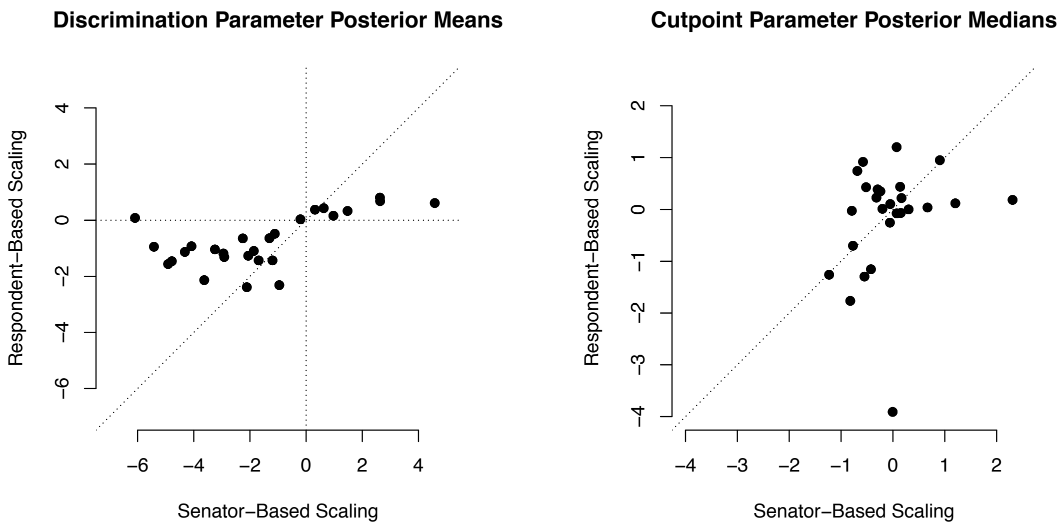

FIGURE 4. *Item Parameters from Senate Representation Study Show Large Differences Between Respondent-Based and Senator-Based Scalings.* Figure plots estimates of discrimination (left pane) and difficulty (right pane) parameters from respondent- and senator-based scalings of Senate Representation Study. Discrimination parameter estimates are posterior means, while cutpoint parameter estimates are posterior medians due to extreme draws from the posterior (see Online Appendix Section 2 for more information).

The two policies for which the discrimination parameters are estimated to have opposite signs for respondents and senators are themselves quite different. S.Amdt. 3158, which proposed that a planned round of military base closures should be restricted solely to bases outside of the United States, had an estimated discrimination parameter of -.20 in the senator-based scaling and .03 in the respondent-based scaling, with the 95% highest posterior density regions (HPDs) for both estimates overlapping zero.[12][13] This suggests that although the signs of the estimates differ between the two groups, the policy does not seem to be very ideologically divisive for either group. Therefore, this might not be thought to be a severe violation of the assumption that the item parameters are the same for the two groups. By contrast, the senator-based and respondent-based discrimination parameters for S.Amdt. 3107, which would have reversed proposed regulations that eliminated overtime pay for workers making over \$100,000 a year or those making over \$23,600 while working as administrators or in professional "white collar" positions, show much larger differences.

[12]HPDs are defined as the smallest region of the parameter space that contains the specified posterior probability (in this case 95%) for a given parameter. HPDs can loosely be thought of as a Bayesian analogue of confidence intervals.

[13]The posterior probability that the senator-based and respondent-based discrimination parameters for S.Amdt. 3158 have the same sign is .19.

The estimated value for senators of -6.09 indicated that the measure was highly ideological, with support for the amendment being the more liberal position. For respondents, the discrimination parameter is estimated to be .08, which implies that support for the amendment was a conservative position, albeit a very mildly divisive one.[14]

The estimated cutpoints (γ_j's) also show considerable variation between senators and respondents. The biggest outlier among the cutpoint estimates is clearly H.R. 1308, whose posterior medians are -.01 and -3.91 for senators and respondents, respectively. There is, however, a large amount of uncertainty in these estimates, particularly for the respondent-based scaling. This is due to the fact that the discrimination parameter for respondents is estimated to be quite close to zero. Therefore, it is difficult to tell if the large discrepancy between the two estimates is due to sampling error or whether it reflects a true difference between the relationship between ideology and positions on this policy between senators and respondents. The second largest outlier (albeit a much milder outlier) among the cutpoint parameter estimates is S. Amdt. 1977 which proposed to prohibit torture of detainees in U.S. military custody, limiting interrogation techniques to those authorized in the U.S. Army Field Manual on Intelligence Interrogation. Even after accounting for uncertainty in the cutpoint estimates, it seems clear that the cutpoint for respondents is much closer to zero than for senators, indicating that moderate respondents are more likely to be indifferent or close to indifferent on this measure, while moderate or even slightly conservative senators were far more likely to support than oppose the measure. This is made clear by the fact that all of the 10 "Nay" votes on the amendment came from Republicans, with most of them being cast by the most conservative members of the party.

One approach to dealing with the differential item functioning indicated by the outlying points in both panes of Figure 4 is to drop the worst of such offenders. In the present analysis, this might suggest omitting S.Amdt. 3107 and H.R. 1308, which had the largest differences in estimated discrimination and difficulty

[14]The 95% HPDs for the senator-based and respondent-based discrimination parameter for S. Amdt. 3107 both did not overlap zero.

parameters, respectively, between the two group-based scalings. The key question in this exercise would be whether the estimated ideal points from the senator-based and respondent-based scalings showed a higher degree of correspondence once one or both of these items was dropped from the scaling. When re-estimating the two group-based scalings dropping the item with the most outlying discrimination parameter (S.Amdt. 3107), the sMSD between the two sets of ideal point estimates remains roughly the same (.94) for senators and rises to .80 for respondents. Dropping H.R. 1308, which has by far the most discrepant cutpoint estimates, results in virtually no change in the correspondence between the estimates (sMSD of .94 and .73 for senators and respondent ideal points, respectively). Finally, dropping both of these items simultaneously produced roughly the same degree of correspondence between senator and respondent ideal point estimates from the two scalings (sMSDs of .93 and .80, respectively) as dropping S.Amdt. 3107 alone. Although this selective item deletion approach shows some promise to increase the correspondence between respondent- and senator-based estimates, it is not clear where to stop given that several remaining policies have a similar level of discrepancy, suggesting that we should either stop after dropping one or two, or proceed to drop many more, neither of which is a particularly satisfying option.

One way to think about how to interpret the sMSD values in this application is to ask how similar group-based scalings would be according to this metric if the ideological dimension were actually identical for the two groups. With the aim of answering this question, Online Appendix section 3 presents a set of simulations in which position data is simulated for all senators and respondents, setting ideal points and item parameters equal to their posterior means from the full joint scaling. 100 such simulated vote matrices are simulated and senator- and respondent-based scalings are estimated for each one. The distribution of sMSD values from these simulations ranged roughly from .97 to .99 for senators and from .95 to .99 for respondents. Both of the observed values (.93 and .73 for senators and respondents, respectively), fall well outside of these ranges. Although the aim of these group-based scaling evaluations is not to provide sharp hypothesis tests of identical dimensions between the two groups, these results emphasize that the discrepancy observed in the Senate Representation Survey suggests that there are significant differences between the dimensions structuring policy views for senators and survey respondents.

Overall, the parameters from these two sets of full scalings show similarities, but are far from identical. The relationship between the senator-based and respondent-based discrimination parameters appears roughly linear on average, but with considerable variation. The magnitude of respondent-based β's is clearly smaller than those from the senator-based scaling. This suggests that while respondents and senators tend to agree on which policy proposals are liberal or conservative, senators discriminate much more sharply based on ideology in their position taking. That the magnitude of the discrimination parameters for respondents might be a fraction of those for senators is equivalent to the error variance in the utility-based voting model being higher for respondents than senators (in the standard CJR model and the group-based model here, the error variance is fixed to one for all actors on all items). This could be seen as unsurprising given that legislators are essentially professional position takers who might be expected to do so with relatively low amounts of error. The cutpoint estimates, while not wildly divergent in most cases, also did not show strong correspondence between the two groups. Perhaps most importantly, the estimated ideal points, which are typically the focus of interest in political science scaling applications, were somewhat similar, but far from identical under the two group-based estimations.

It bears keeping in mind that the Senate Representation Survey poses a "hard test" for the assumption that ordinary citizens and political elites have their ideologies structured in the same way. Many of the policy items included in this survey could be considered obscure or technical, regarding topics ordinary citizens have not formed solid opinions about. Even with this hard test, however, the item parameters are found to have large positive correlations, albeit with a slope clearly less than one, and the estimated ideal points, which are typically the main parameters of interest in political science applications, are quite similar, particularly after dropping the most problematic of the items. These mixed results beg the question of how common-scale ideal point estimation between citizens and political elites might fare when applied to different types of data including those focused on the types of issues ordinary citizens are likely to have encountered and formed meaningful opinions on.

6. Data: 2008 Cooperative Congressional Election Study

The 2008 Cooperative Congressional Election Study (CCES) was fielded to an online sample of 32,800 respondents from the Polimetrix/YouGov online panel during October and November 2008. Various versions of the CCES study have been fielded since 2006, but the 2008 version was chosen specifically because it contains the largest number of policy questions that pertain to specific House and Senate roll call votes, making it particularly well suited for our purposes. In total, the 2008 CCES included eight items that directly corresponded to specific votes taken during the 110th House and Senate sessions. Table 2 lists the policies as well as the House and Senate vote margins and the percentage of respondents supporting, opposing, and saying "don't know" to each.[15] Although the number of bridging items in the 2008 CCES is smaller than in the Senate Representation Study, it has two attractive features. First, the CCES sample contains more than five times as many more respondents as the Senate Representation Survey. Second, the CCES contains items for which we know the positions of respondents, senators, and House members, allowing for three group-based scaling possibilities. We can thus compare not only how the structure of respondent and legislator ideology may differ, but also how the structure of House and Senate ideology may differ.

While the Senate Representation Study deliberately included a range of issues, the items asked in the CCES tended to pertain to more straightforward policies. As such, the CCES might be expected to be an "easier test" for the bridging assumptions implied by common-space scaling. Another way to think of this is that the policy items asked in the Senate Representation Study were closer to the type of proposals routinely voted on by legislators, while the CCES mostly asked about items that ordinary citizens might encounter and think about more routinely. It is instructive, then, to ask how the ideological dimensions underlying

[15]The 2008 CCES also included an item about a constitutional amendment to define marriage as between one man and one woman, but this was not included in the analysis, because it was not directly voted on by the House and Senate during the 110th Congress.

Table 2. List of Policy Items from 2008 CCES

Policy	Representatives Yea-Nay Votes	Senators Yea-Nay Votes	Respondents Y-N-DK %
Withdrawing troops from Iraq within 180 days	171-256	28-71	47-41-13
Increasing minimum wage to $7.25	315-117	95-3	72-21-7
Allow federal funding of stem cell research	247-177	63-35	53-30-16
Allow U.S. spy agencies to eavesdrop on overseas terrorist suspects without first getting a court order	294-129	70-28	59-27-14
Fund a $20 billion program to provide health insurance for children in families earning less that $43,000	265-160	68-32	58-26-16
Federal assistance for homeowners facing foreclosure and large lending institutions at risk of failing	241-173	84-13	39-39-22
Extend the North American Free trade Agreement (NAFTA) to include Peru and Columbia	286-132	78-18	31-34-35
U. S. Government's $700 Billion Bank Bailout Plan	264-171	75-25	20-54-26

Table shows House and Senate vote totals and percentages of 2008 CCES survey respondents supporting, opposing, and saying "don't know" to each surveyed policy.

the preferences of legislators and ordinary citizens differ in the CCES data and how the overall character of these results compares to those for the Senate Representation Survey.

In order to answer these questions, three versions of the group-based ideal point model are estimated on the CCES data, letting the positions taken by House members, senators, and respondents, respectively structure the dimension on each of the estimates. The scales from these three separate estimations are identified by post-processing each iteration of the sampler so that the mean and standard deviation of legislator (House members and senators together) ideal points have mean zero and variance one, and that higher ideal point values indicate more conservative positions. This allows for direct comparisons of the all estimated parameters—ideal points, discrimination parameters and cutpoint parameters—from the three different group-based scalings.

Figure 5 plots the relationship between the estimated ideal points from these three scalings for House members, senators and respondents separately. The most obvious feature across all of these plots is the very high degree of similarity between the estimates across all scalings and all types of actors. The sMSDs for these nine comparisons range from .94 to .99, which is quite high particularly since these ideal points are estimated based on only eight items and therefore contain a considerable amount of measurement error.

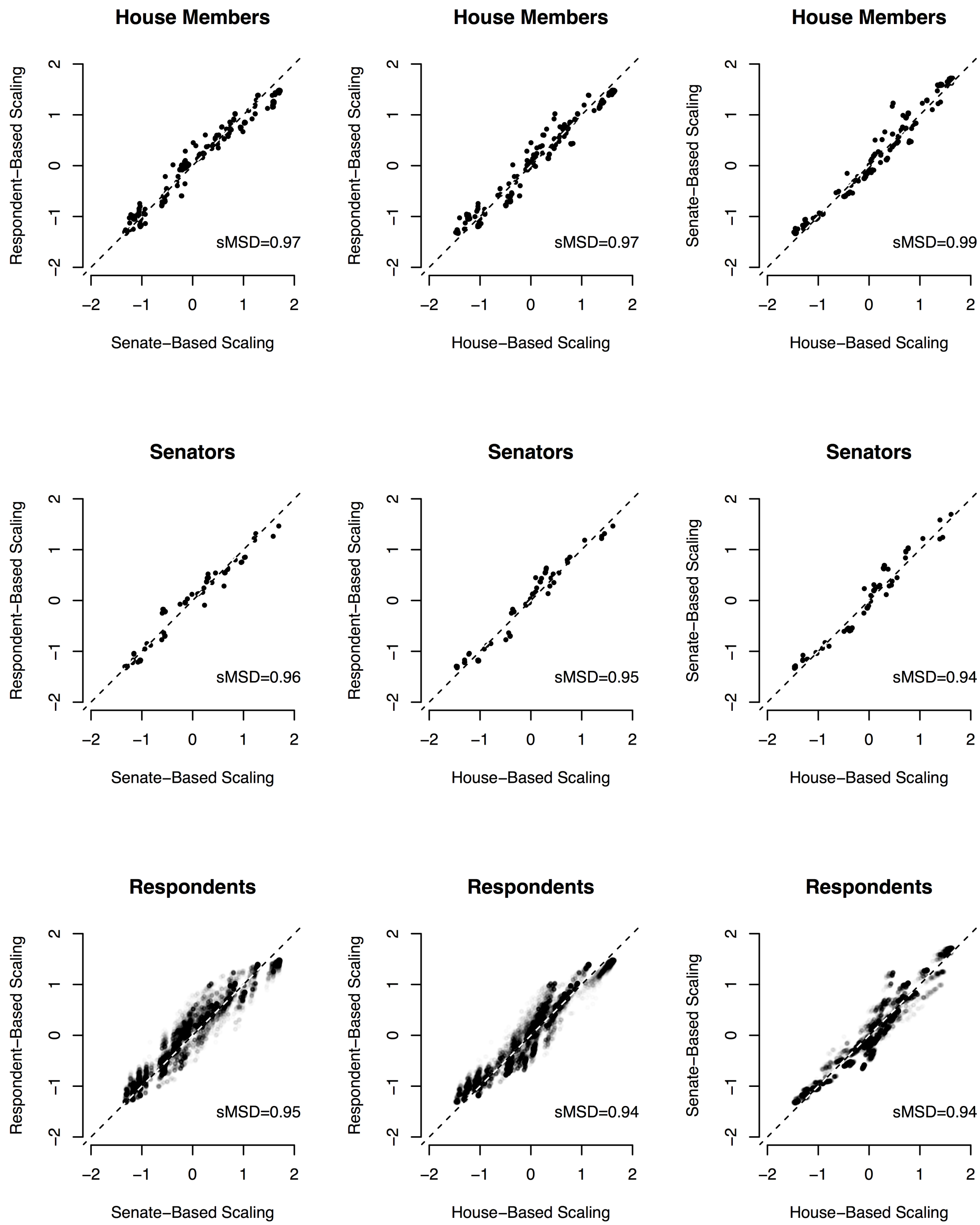

FIGURE 5. *2008 CCES Ideal Point Estimates are Similar Under House- Senate- or Respondent-Based Scalings.* Plots show estimates (posterior means) for ideal points of House members, senators and respondents under House- Senate- and respondent-based scalings. Respondent estimates are plotted with transparency to show overlapping points.

The estimated item parameters from these three scalings are plotted in Figure 6. The three panes in the left column show estimates of the discrimination parameters, comparing each possible pair from the estimates structured by House members, senators and respondents. In all three cases, there is a positive relationship between the estimates. The respondent-based discrimination parameters are very similar to those based on House member and senator preferences, with a strong linear relationship between the estimates and a relatively small amount of error. It is clear, however, that the slope of this linear relationship is less than one, meaning that the discrimination parameters used by respondents are smaller in magnitude (closer to zero) than those used by either House members or senators. This pattern is similar to the one in Figure 4 from the Senate Representation Study, but the relationship is much stronger in the CCES data. As above, this can be interpreted to mean that respondents are "noisier" position takers, but that the basic pattern of how relatively liberal or conservative each issue is tends to be quite similar for all three of the groups considered here.

The right column of Figure 6 shows the relationships between the estimated cutpoint parameters from the three group-based scalings estimated for the CCES data. Although there is a fairly strong correspondence between the cutpoint estimates for senators and House members (lower-right pane), the relationships for respondents and both sets of legislators (upper-right and middle-right panes) are much weaker. As in the Senate Representation Survey, however, the estimated cutpoints tend to be clustered near the middle of the ideological scale. This means that even though there is not a high correlation between the cutpoints estimated for legislators and respondents, the actual distance between the estimates tends to be small, with one or two notable exceptions. The biggest outlier between respondent and both House or Senate estimates is H.R. 1424, the "Emergency Economic Stabilization Act of 2008" (commonly known as the federal "bailout"). Although only 27 percent of respondents who took a position supported this policy, majorities voted for it in both the House and Senate. A more mildly outlying cutpoint between legislators and respondents is H.R. 3688, extending the NAFTA to include Columbia and Peru, which was supported by large majorities

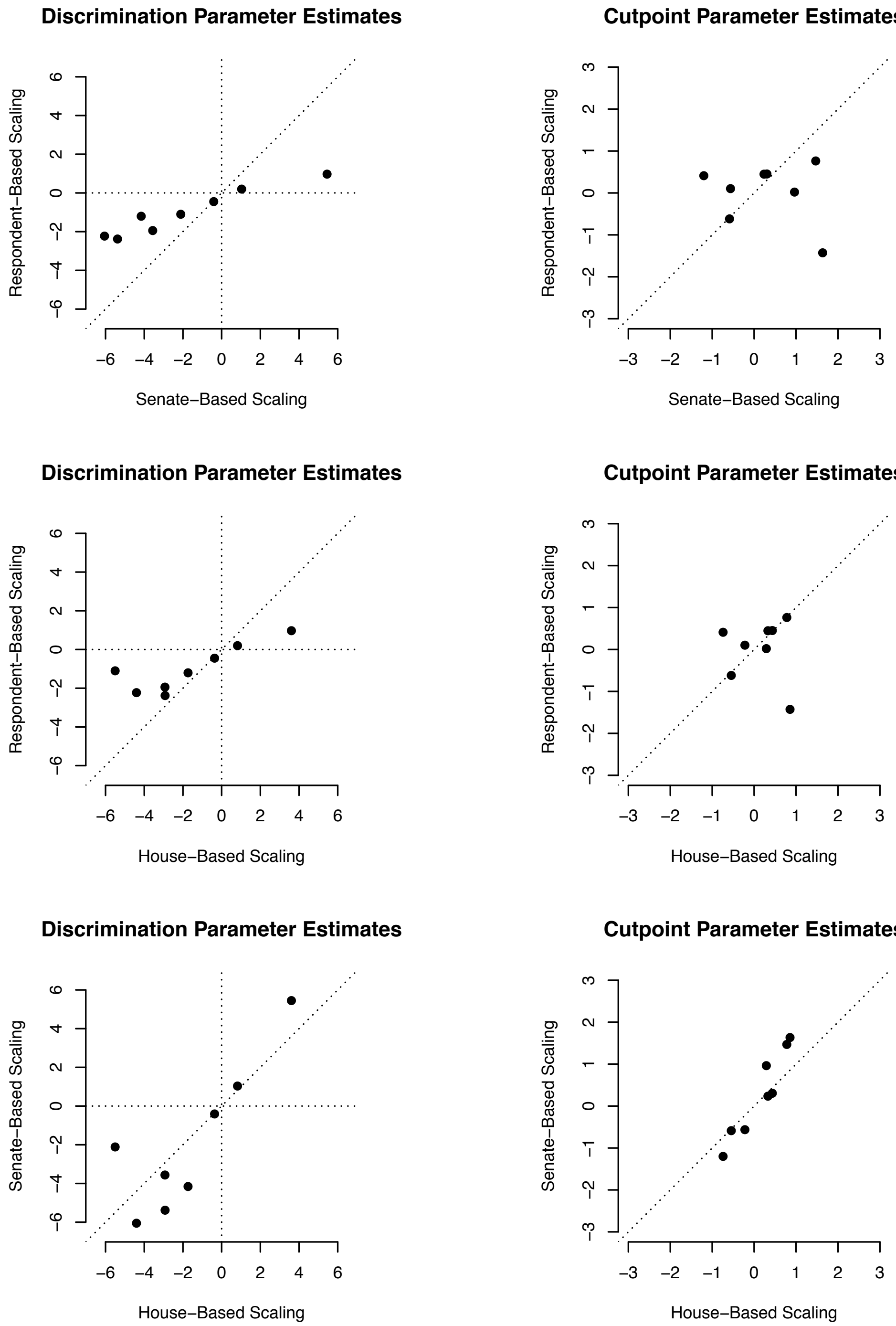

FIGURE 6. *2008 CCES Item Parameter Estimates Show Some Differences Under House-Senate- or Respondent-Based Scalings.* Plots compare estimates for item parameters under House- Senate- and respondent-based scalings of 2008 CCES data. Discrimination parameter estimates are posterior means, while cutpoint parameter estimates are posterior medians due to extreme draws from the posterior (see Online Appendix Section 2 for more information).

in the House and Senate, but a minority of respondents. Overall, the cutpoints for the House- and Senate-based estimates are more similar to each other than the respondent-based estimates are to either of the legislator-based cutpoints.

7. "Hard" vs. "Easy" Issues and the Performance of Joint Scaling

One way to assess whether the differing results for the Senate Representation Survey and the 2008 CCES were driven primarily by the different types of policy items used in the two surveys is to separately look at the "hard" and "easy" policy items on the Senate Representation Survey. Because this survey included a much larger number of policy items than the CCES, we can create two different "sub-surveys", each having 8 policy questions, the same number as CCES. To do this, I select the 8 items that are the most straightforwardly ideological and which respondents are most likely to have thought about, calling them "easy" issues. I also select the 8 issues that respondents are the least likely to have thought about and formed meaningful opinions about, call these "hard" questions. The leftmost column of Table 1 shows these classifications. While they are somewhat subjective, these categorizations were made without reference to the estimated item parameters (for example, the "easy" questions were not simply chosen to be those whose discrimination parameters were largest in magnitude). Respondents who answered at least 5 of a given issue type were included in this analysis, yielding 1,091 respondents for the "easy" issue scaling and 749 for the "hard" issue scaling.

The correspondence between respondent-based and senator-based ideal point estimates is indeed much higher when using the "easy" issue items than when using the "hard" ones. The sMSD for the "easy" issue scaling was .96 for senators and .86 for respondents. By contrast, the senator and respondent sMSD was .75 and .51 for senators and respondents, respectively. Although these figures are affected somewhat by the smaller number of respondents, it seems clear that certain types of issues yield more directly comparable scales for the two types of actors. This finding should not be surprising given that legislators are essentially professional position takers, while most citizens are unlikely to have thought about many of the policies that are decided on by Congress.

8. Discussion

Ideal point models are a valuable tool for political scientists. But these models are not magic. There is no guarantee that an ideal point model will find the dimension of interest to researchers when fed a set of data. The burden is on researchers to provide clear thought about both the indicators chosen and what underlying dimension is relevant for a given application. The selection of a measurement approach is as much a substantive or theoretical issue as a statistical one.

In the context of studies bridging citizen and legislator ideology, the results presented above make clear that the dimension estimated can vary in unexpected ways. For example, given that the number of legislators is often fixed in these studies, the characteristics of estimated ideal points can in some cases depend strongly on the number of respondents in the survey data analyzed. This is problematic in and of itself, but also suggests that the underlying dimensions for legislators and citizens are not identical, a hypothesis further corroborated looking at the estimated item parameters from separate scalings of these groups.

So how, broadly speaking, should researchers respond to the findings presented above? One response is to admit defeat. If the primary ideological dimension underlying citizens' policy views and the one underlying legislators' roll call voting are not the same, one could argue that it is meaningless to talk about comparing the positions of citizens and legislators.

Another response is to recognize that the results of joint scalings are typically used in subsequent analyses, often some variant of a regression model, to test hypotheses of interest. In these situations, researchers could remain agnostic as to which measure is most appropriate and attempt to validate their hypotheses using both the citizen-based and respondent-based ideology estimates as described above. If the findings are similar using both measures, this may provide some reassurance that the findings hold under both conceptualizations of ideology.

Yet another potential response is to view similarity of ideal point estimates between different groups as the primary criterion for selecting policy items in joint scalings. As discussed above, it seems clear that the use of "easier" policy items produces ideology scales that are more similar between legislators and citizens. One could go even farther by pilot testing a large set of policy items in a survey and selecting for inclusion in a larger survey those items that produce the most similar scales according to measures such as sMSD, similarity of discrimination and difficulty parameters, or other criteria.

The most productive response, I would argue, is to recognize that while many latent dimensions may exist in joint scaling applications, researchers should seek to estimate the dimension that is relevant for whatever theory or hypothesis they are studying. This dimension is unlikely to be "whatever the ideal point model estimates when given a set of indicators," which, loosely speaking, is the dimension that explains the most variation in the data. To determine which dimension is appropriate, we must look to the theoretical framework motivating our hypothesis test. For example, in tests of spatial voting theory (e.g. Jessee 2012), which posit that voters are more likely to cast their ballot for candidates whose ideological positions are closest to their own, a citizen-based ideological dimension might be most relevant given that the theory describes voters calculating ideological distances on which to base their decisions. Conversely, if one was interested in how candidates for office would be estimated alongside sitting legislators if they joined Congress, a legislator-based scaling in which the candidates' policy positions were matched to specific roll call votes may be appropriate. The group-based scaling technique presented here provides a way to impose such restrictions on the dimension being estimated in joint scaling applications.

A related issue is how literally ideal point models should be taken. On one hand, a key advantage of these models is that they are often built up from clear microfoundations of spatial utility. These foundations correspond nicely to many broader theoretical frameworks for analyzing politics, including representation and spatial voting. On the other hand, it is clear that these models are abstractions, rather than literally true representations of how specific policy positions are generated. These models could even be viewed as data reduction or summarization tools. The most useful way to evaluate them is not to verify that the models are strictly true, but rather to assess whether they represent useful, rather than misleading, simplifications.

Along these lines, one could imagine proposing a heteroskedastic model under which the variance of the utility disturbances for citizens and legislators are potentially different. This would be equivalent to allowing for the discrimination parameters for respondents to have a different magnitude than those for legislators, paralleling the results shown in Figures 4 and 6. In fact similar models have been estimated in the context of testing spatial voting in presidential elections by Jessee [2009] and in the context of congressional roll call voting by Lauderdale [2010]. But this model, while coherent within the ideal point framework, fundamentally changes the meaning of the estimated dimension. For example, if respondents have larger error variances than legislators, it would be possible for a legislator to have a more conservative ideal point than a respondent, but for the respondent to be more likely to support a given conservative proposal. More seriously, rather than remedying the issues with joint scaling identified here, such extensions have the potential to make the problems even worse. As an example, a heteroskedastic ideal point model applied to the Senate Representation Survey data produces dramatically different estimates of ideal points and of the relative error variances for the two groups depending on how many respondents are included in the analysis (see Online Appendix section 4). These results suggest that while thinking seriously about the formal assumptions of ideal point models obviously remains important, we should not conclude that problems with specific formalizations in ideal point models can always be solved by further elaboration of the statistical models used. In fact, this approach has the potential to worsen these problems while also taking estimates farther away from measuring desired concepts.

This paper has examined some of the most important questions about the validity of bridging ideal point models. But the set of issues considered here is by no means complete. Although analyses estimating ideology across multiple groups have shown great promise in recent years, researchers must continue to work to assess their applicability, robustness and validity. In the end, the central concern for researchers using this approach must be that estimates represent as closely as possible the substantive dimension(s) being studied. Although statistical tests can speak to this issue, they are far from a panacea. Careful thought about the data used for joint scaling as well as the assumptions of the model and structure imposed on the estimation remain the most important considerations in joint scaling applications.

References

Christopher H. Achen. Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science*, 21(4):805–815, 1977.

James F. Adams, Samuel III Merrill, and Bernard Grofman. *A Unified Theory of Party Competition*. Cambridge University Press, 2005.

John H. Aldrich and Richard D. McKelvey. A method of scaling with applications to the 1968 and 1972 presidential elections. *The American Political Science Review*, 71(1):111–130, 1977.

Michael R. Alvarez and Jonathan Nagler. Economics, issues and the perot candidacy: Voter choice in the 1992 presidential election:. *American Journal of Political Science*, 39:714–44, 1995.

Stephen Ansolabehere. Cooperative congressional election study, 2008: Common content, July 2011.

Joseph Bafumi and Michael C. Herron. Leapfrog representation and extremism: A study of american voters and their members in congress. *American Political Science Review*, 104(3):519–542, 2010.

Michael A. Bailey. Comparable preference estimates across time and institutions for the courts, congress, and presidency. *American Journal of Political Science*, 51(3):433–448, 2007.

Adam Bonica. Mapping the ideological marketplace. *American Journal of Political Science*, 58(2):367–386, 2014.

Henry E. Brady and Paul M. Sniderman. Attitude attribution: A group basis for political reasoning. *The American Political Science Review*, 79(4):1061–1078, 1985.

Melvin J. Hinich Cahoon, Lawrence and Peter C. Ordeshook. *Graphical Representations of Multivariate Data*, chapter A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting, pages 243–278. Academic Press, 1978.

Joshua Clinton, Simon Jackman, and Douglas Rivers. The statistical analysis of roll call data: A unified approach. *American Political Science Review*, 98:355–70, 2004.

Anthony Downs. *An Economic Theory of Democracy*. HarperCollins, 1957.

James Enelow and Melvin J. Hinich. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press, 1984.

Tim Groseclose and Jefferey Milyo. A measure of media bias. *Quarterly Journal of Economics*, CXX(4): 1191–1237, 2005.

James J. Heckman and James M. Snyder. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics*, 28:142–89, 1997.

Harold Hotelling. Stability in competition. *The Economic Journal*, XXXIX:41–57, 1929.

Simon Jackman. *pscl: Political Science Computational Laboratory (R library)*, 2009.

Stephen A. Jessee. Spatial voting in the 2004 presidential election. *American Political Science Review*, 103 (01):59–81, 2009.

Stephen A Jessee. *Ideology and Spatial Voting in American Elections*. Cambridge University Press, 2012.

Benjamin E. Lauderdale. Unpredictable voters in ideal point estimation. *Political Analysis*, 18(2):151–171, 2010.

Jeffrey B Lewis and Chris Tausanovitch. Has joint scaling solved the achen objection to miller and stokes? Working Paper, 2013.

Keith T. Poole and Howard Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357–384, 1985.

Boris Shor. All together now: Putting congress, state legislatures, and individuals in a common ideological space. Working Paper, 2009.

Alexander M. Tahk. The signals in the noise: A cross-validation approach to estimating the dimensionality of roll call voting. Unpublished, 2005.

Chris Tausanovitch and Christopher Warshaw. Representation in municipal government. *American Political Science Review*, 108(03):605–641, 2014.

Herbert F. Weisberg and Jerrod G. Rusk. Dimensions of candidate evaluation. *The American Political Science Review*, 64(4):1167–1185, 1970. ISSN 0003-0554.

ONLINE APPENDIX: (HOW) CAN WE ESTIMATE THE IDEOLOGY OF CITIZENS AND POLITICAL ELITES ON THE SAME SCALE?

This appendix provides supplementary information for the paper "(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?" Section 1 provides information about the estimation procedure for the group-based ideal point model discussed in the paper. Section 2 provides information about the behavior of the Gibbs sampler, including why posterior medians, rather than means, are used as estimates of the cutpoint parameters.

1. Description of Group-Based Scaling Gibbs Sampler

This section describes a modified version of the Gibbs sampler for the quadratic-normal ideal point model from Clinton et al. [2004] (CJR). The modification restricts the sampler to estimate the item parameters (α, β and, by transformation γ) to be structured based only on the preferences of a pre-specified group of actors. The sampler is identical to the CJR sampler except for the sampling step for the item parameters. This appendix shares notation with CJR. For more information about the standard version of the model and sampler, see the CJR paper.

Starting with a set of initial values for the ideal points x and item parameters α and β, we can write the steps of the Gibbs sampler for the group-restricted ideal point model as:

(1) sample y^* from $p(y^*|x, \alpha, \beta, y)$

(2) sample γ, β from $p(\alpha, \beta|x_{(g)}, y^*_{(g)})$

(3) sample x from $p(x|\alpha, \beta, y^*)$

where y^* denotes the latent utility differences between the two alternatives for each voter on each vote which are sampled at each iteration as a data augmentation step and the (g) subscript for x and y in the second sampling step denotes the values only for the group chosen to structure the ideological dimension. Note that we estimate the model in the specification from equation 1 in the paper and then transform $\gamma = \alpha/\beta$ at each iteration of the sampler to obtain the specification in equation 2.

The first step simply involves sampling from a normal with mean $x_i\beta_j$ and variance one, truncated at zero either above or below based on each observed value of y_{ij}. The second step is a set of regressions for each item j conditional on the values of x and y^*, only using the parameters from members of the group selected for structuring the ideology scale. The final step of the sampler involves transforming to obtain a set of regressions in which the ideal points x become regression coefficients for each actor i conditional on the item parameters and latent utility differences.

The sampler is implemented through a modification of the pscl function in the ideal library in R Jackman [2009]. Code is available from the author upon request.

2. Gibbs Sampler Performance

The sampler appears to converge rapidly (within several thousand iterations or, in most cases, much sooner) for the models and datasets used in the paper. The main issue involves the transformation of the item parameters to obtain cutpoint estimates (γ_j's) for each parameter by dividing the difficulty parameter α_j by the discrimination parameter β_j. Although this makes the parameters more easily interpretable, it induces undesirable behavior for posterior samples for the cutpoint parameters. This occurs for iterations of the sampler in which discrimination parameters take values near zero, causing the cutpoint parameters to become extremely large in magnitude.

In order to obtain more stable estimates of the cutpoint parameters that are not dramatically affected by the small number of extreme draws, I use the median (rather than mean) of posterior draws for the cutpoints as my estimates. Figure A1 shows an example of this, plotting posterior medians and means for each item in the Senate Representation Study for both Respondent-based and Senator-based ideal point estimations. Discrimination parameter estimates are nearly identical when using means or medians. Cutpoint estimates, however, differ sharply in some cases. The figure also shows examples of traceplots for the most discrepant

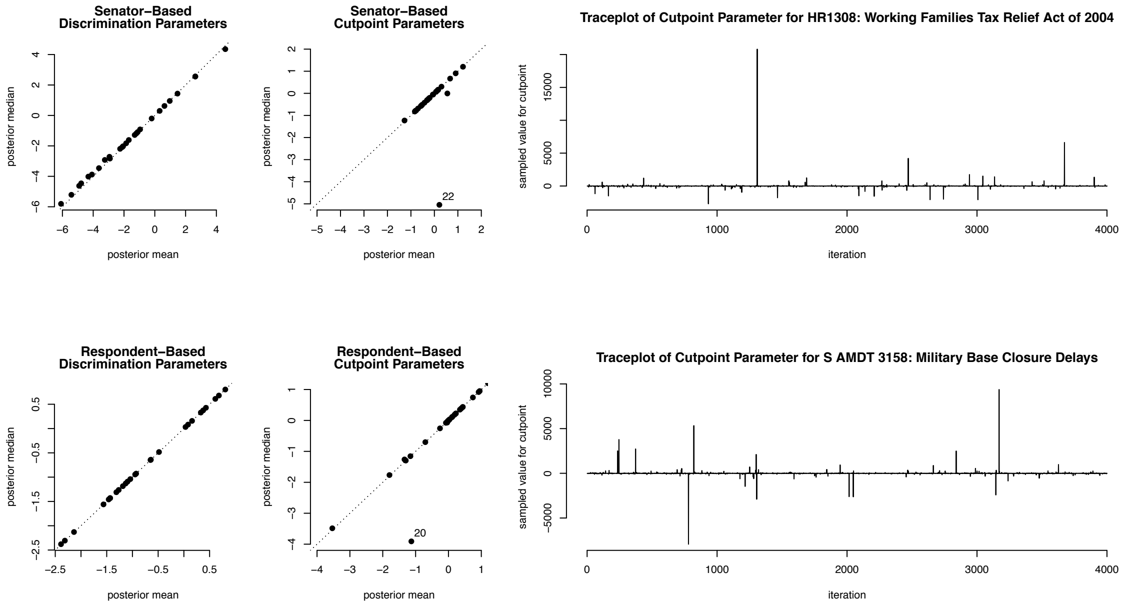

Figure A1. *Cutpoint Parameter Posterior Means are Highly Volatile for Items with Discrimination Parameter Estimates Near Zero.* Left (center) column plots means vs. medians of Gibbs sampler draws for discrimination (cutpoint) parameters for each item under group-based scalings of Senate Representation Study. Right column shows traceplots for the most divergent cutpoint parameters under each scaling.

cutpoints under each scaling. It is apparent that a relatively small number of iterations in which the discrimination parameters (traceplot not shown) are sampled near zero produce very high or very low values for the cutpoint parameters, having a strong affect on the resulting posterior mean estimates. This occurs only for the few items on which there is nontrivial posterior mass near zero.

3. sMSD Monte Carlo Simulations

This section describes a set of Monte Carol simulations for the Senate Representation Survey and 2008 CCES data in which all ideal points (x_i) and item parameters (β_j and α_j) are fixed at their posterior means from the full joint scaling of legislators and respondents, then for each simulation, a vote matrix is simulated by sampling from the distribution of votes given the parameter values and legislator- and respondent-based scalings are run separately for the simulated vote matrix. This process is repeated 100 times [NOTE: ONLY HAVE 20 SO FAR FOR CCES, SO RESULTS ARE LESS RELIABLE] and the sMSD between the two group-based scalings is noted.

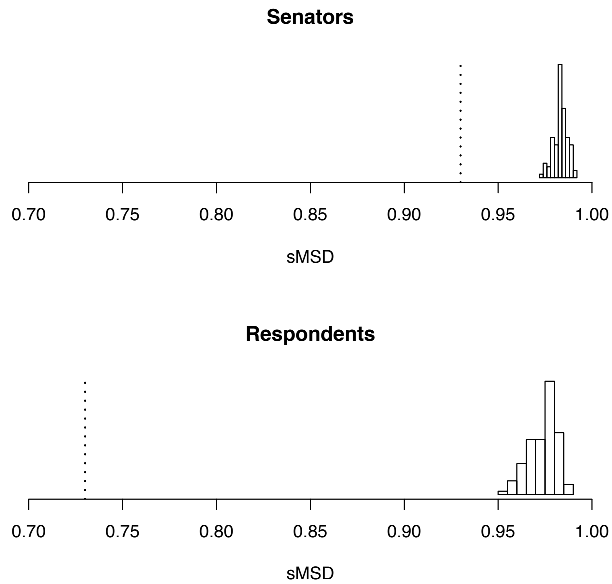

FIGURE A2. sMSD Monte Carlo Simulations, Senate Representation Survey Data

Figures A2 and A3 show histograms of the distributions of sMSD values for these simulations for the two datasets. On each histogram, the sMSD between the corresponding group-based scalings of the actual data are denoted by vertical lines. It is clear that the observed sMSD for the Senate Representation Survey is much lower for both senators and respondents, indicating a much larger discrepancy between the estimated ideological dimensions than we would expect to occur if the joint model were true. For the CCES data, however, the observed sMSD values fall much closer to those that would be expected if the joint model were true. While this is not intended as a sharp test of the hypothesis of identical dimension structure between the two groups, it does provide a way of assessing the magnitude of the discrepancy under each of the two group-based scalings.

4. Heteroskedastic Ideal Point Model Performance

This section presents the results of applying a heteroskedastic ideal point model to the Senate Representation Survey data. The model assumes that the probability of actor i supporting policy j can be written

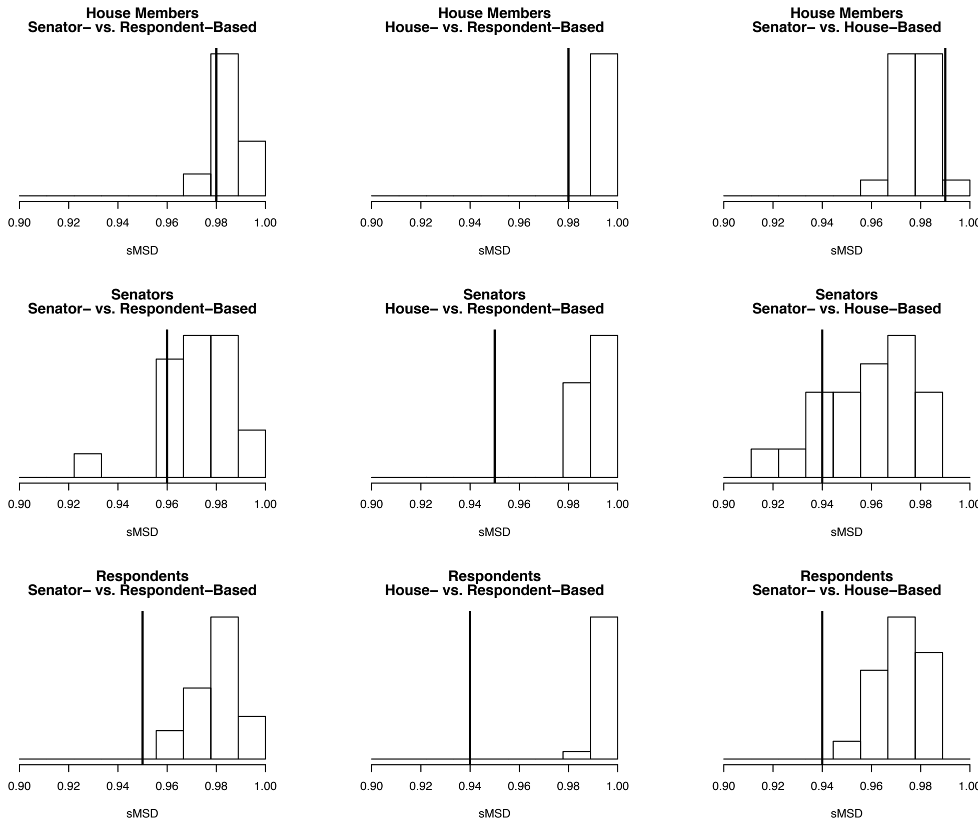

FIGURE A3. sMSD Monte Carlo Simulations, 2008 CCES Data

as

$$P(y_{ij} = 1) = \Phi(\frac{x_i\beta_j - \alpha_j}{\sigma_j}) \tag{1}$$

where σ_j is fixed at 1 for senators, while it is given a uniform prior from 0 to 100 for respondents, with all respondents assumed to have the same σ_j. Ideal points x_i are given independent standard normal priors, and both β_j and α_j are given independent normal priors with mean 0 and variance 25. The model is estimated using JAGS for the full data as well as for 100 randomly sampled subsets of respondents for sizes 2,000, 500 and 111.

Figure A4 plots the posterior means for ${}^{\sigma_{resp}}/_{\sigma_{sen}} = \sigma_{resp}$ (since σ_{sen} is fixed at 1) against the ratio of the standard deviations of respondent and senator idea ideal point estimates. The results make clear that,

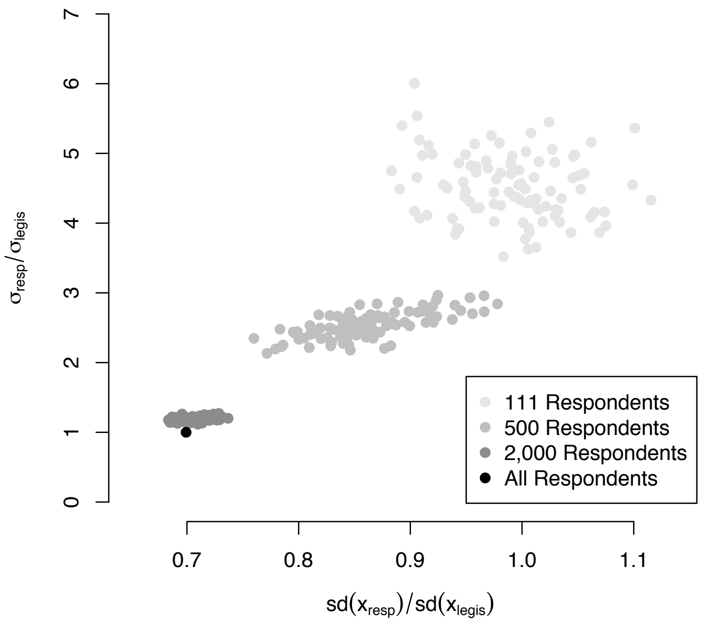

FIGURE A4. Heteroskedastic Ideal Point Model Simulations Varying Number of Respondents

far from solving the common dimension problem, this additional modeling feature results in perhaps *more* problematic pathologies whereby both the ideal point estimates and the relative error variances are sensitive to the number of survey respondents included in the data.

REFERENCES

Joshua Clinton, Simon Jackman, and Douglas Rivers. The statistical analysis of roll call data: A unified approach. *American Political Science Review*, 98:355–70, 2004.

Simon Jackman. *pscl: Political Science Computational Laboratory (R library)*, 2009.