Political Analysis, 9:3

Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking

Simon Jackman
*Department of Political Science,
Stanford University, Stanford, California 94305-6044
e-mail: jackman@stanford.edu*

Vote-specific parameters are often by-products of roll call analysis, the primary goal being the measurement of legislators' ideal points. But these vote-specific parameters are more important in higher-dimensional settings: prior restrictions on vote parameters help identify the model, and researchers often have prior beliefs about the nature of the dimensions underlying the proposal space. Bayesian methods provide a straightforward and rigorous way for incorporating these prior beliefs into roll call analysis. I demonstrate this by exploiting the close connections among roll call analysis, item–response models, and "full-information" factor analysis. Vote-specific discrimination parameters are equivalent to factor loadings, and as in factor analysis, they (1) enable researchers to discern the substantive content of the recovered dimensions, (2) can be used for assessing dimensionality and model checking, and (3) are an obvious vehicle for introducing and testing researchers' prior beliefs about the dimensions. Bayesian simulation facilitates these uses of discrimination parameters, by simplifying estimation and inference for the massive number of parameters generated by roll call analysis.

1 Introduction

IT IS WELL KNOWN that the analysis of roll call data generally results in statistical models with many parameters. Operationalizing the D-dimensional Euclidean spatial voting model (Enelow and Hinich 1984) with roll call data from n legislators over m roll calls generates a statistical model with $nD + m(D+1)$ parameters. For instance, fitting a unidimensional model to data from a recent U.S. Senate ($n = 100, m \approx 500$) creates a 1100-parameter problem (100 ideal points and 500×2 proposal parameters), while a two-dimensional model has 1700 parameters. Likelihood-based estimation and inference with this many parameters remain formidable even with the computing power now available to social scientists.[1]

Author's note: I thank John Londregan, Adam Mierowitz, Keith Poole, and, especially, my collaborators on this project, Joshua Clinton and Doug Rivers, for helpful discussion and comments. Errors and omissions remain my own responsibility.

[1] Direct MLE may be feasible with data sets that are small relative to the data sets generated by the contemporary U.S. Congress. For instance, see Londregan's (2000b) analyses of committees in the Chilean legislature, where

In a recent article in *Political Analysis* (Jackman 2000) I reported on recent work with Joshua Clinton and Doug Rivers, where we use Bayesian simulation (Markov chain Monte Carlo methods) to simplify estimation and inference for the large number of parameters arising in roll call analysis. In my earlier article I focused on the issue of inferences for legislators' ideal points in a unidimensional setting. Here I show how the Bayesian approach helps us deal with the added complexities of moving to higher-dimensional contexts. First, identification, estimation, and inference for ideal points become more complicated in the higher-dimensional setting, and I show how each task is accomplished in a Bayesian setting.

Second, I show how the *proposal parameters* assume more importance when we shift to higher dimensional settings. In most roll call analyses, proposal parameters are often considered nuisances, since the usual goal is measuring legislator's ideal points, so much so that the statistical analysis of roll call data is often referred to as "legislative scaling." In the Bayesian approach there is no real distinction between either type of parameter (legislators' ideal points or proposal-specific parameters), and Bayesian simulation methods easily provide estimates and inferences for both sets of parameters; contrast likelihood-based approaches that marginalize with respect to one set of parameters so as to obtain estimates and inference for the others (e.g., Bock and Aitken 1981). I show below that proposal parameters are analogous to factor loadings and can be put to the same uses as factor loadings. These include determining the qualitative nature of recovered dimensions (as in exploratory factor analyses) or a means for researchers either to impose or to test conjectures about the nature of the underlying dimensions (as in confirmatory factor analyses). I develop some diagnostics for assessing dimensionality based on the proposal parameters. In short, my goal here is to use Bayesian simulation to make the analysis of roll call data less a technical "scaling" exercise and more genuinely *data analytic*, in which researchers' conjectures or substantive expertise can alternately be tested, or used to guide the data analysis.

2 Operationalizing the Euclidean Spatial Voting Model

Assume a D-dimensional Euclidean proposal space. Each bill $j = 1, \ldots, m$ presents each legislator $i = 1, \ldots, n$ with a choice between a Yea position, $\boldsymbol{\zeta}_j$, and a Nay position $\boldsymbol{\psi}_j$. The recorded votes (roll calls) are binary indicators: $y_{ij} = 1$ if legislator i votes Aye on the jth vote and $y_{ij} = 0$ if legislator i votes Nay. The Euclidean spatial voting model drives the development of a statistical model for these data: legislators receive utilities from $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$ declining in the squared distance of the these points from each ideal point $\mathbf{x}_i$. It is well known that the *statistical* model implied by the Euclidean spatial voting model is the following *two-parameter item-response model*, used extensively in the educational testing literature:[2]

$$y^*_{ij} = U_i(\boldsymbol{\zeta}_j) - U_i(\boldsymbol{\psi}_j) = \boldsymbol{\beta}'_j \mathbf{x}_i - \alpha_j + \varepsilon_{ij}, \tag{1}$$

with the censoring rule $y_{ij} = 1 \iff y^*_{ij} > 0$, otherwise $y_{ij} = 0$. With the further assumption $\varepsilon_{ij} \sim N(0, 1)$ we have a hierarchical probit model with the complication that the ideal points $\mathbf{x}_i$ appear as unobserved predictors in Eq. (1), to be estimated along with the proposal parameters $\boldsymbol{\beta}_j$ and α_j.

$n < 10$ and (by assumption) the $m(D + 1)$ proposal parameters are reduced to a dramatically smaller set of *proposer* parameters.

[2]See Jackman (2000, pp. 317–323) and Londregan (2000a).

Given our Bayesian approach, estimation and inference for this model amounts to computing the posterior density $\pi(\boldsymbol{\theta} \mid \mathbf{Y})$, $\boldsymbol{\theta} = \{\mathbf{X}, \mathbf{B}, \boldsymbol{\alpha}\}$, where $\mathbf{X}$ is an $n \times D$ matrix of ideal points, $\mathbf{B}$ is an $m \times D$ matrix of discrimination parameters, and $\boldsymbol{\alpha}$ is an m-vector of intercepts (or item-difficulty parameters). A Gibbs sampler generates arbitrarily many samples from this joint posterior density, which are then summarized for inference; a more detailed description of the Gibbs sampler for this problem is given by Clinton *et al*. (2000), drawing on work in the item–response context by Johnson and Albert (1999) and Albert (1992).

Finally, note that the parameters α_j and $\boldsymbol{\beta}_j$ in Eq. (1) are functions of the unknown Yea and Nay locations, $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$. Without further identifying restrictions the proposal parameters themselves cannot be uniquely recovered from α_j and $\boldsymbol{\beta}_j$,[3] and instead we recover the "indifference hyperplane," the set of points equidistant from $\boldsymbol{\zeta}_j$ and $\boldsymbol{\psi}_j$. Researchers often find it convenient to use these cutting planes when presenting results on specific sets of votes (e.g., Pool and Rosenthal 1997, Chap. 7); my focus here is on the underlying proposal parameters themselves.

3 Inference for Proposal-Specific Parameters

The most common goal of roll call analysis is *measurement*: estimation and inference for the ideal points $\mathbf{x}_i$. Nonetheless, the slope parameters $\boldsymbol{\beta}_j$ are substantively interesting in their own right and are equivalent to *item discrimination* parameters in the item–response literature. In the legislative context it is also useful to think of the $\boldsymbol{\beta}_j$ as discrimination parameters, literally tapping the extent to which the jth roll call discriminates among legislators along the various dimensions of the proposal space. For instance, with a unidimensional model, Eq. (1) is $y^*_{ij} = \beta_j x_i - \alpha_j + \varepsilon_{ij}$, and β_j taps how change in x_i translates into support for proposal j. A plausible hypothesis is that support for proposal j is unrelated to movement in x_i (i.e., β_j is indistinguishable from zero) or, in substantive terms, that support for proposal j is unrelated to the underlying policy continuum. In fact, the substantive content of proposals that have large and statistically significant β_j supply the substantive content of the underlying policy continuum.

Inspecting the discrimination parameters $\boldsymbol{\beta}_j$ is thus extremely useful in assessing the *qualitative* character of each dimension. But, in addition, the discrimination parameters also let us assess the *dimensionality* of the proposal space. For instance, if a one-dimensional model yields a large number of discrimination parameters indistinguishable from zero, then the researcher might well be prompted to fit a higher-dimensional model. In addition, those roll calls with insignificant discrimination parameters can be examined to help the researcher determine the likely substantive content of the higher dimensions.

3.1 *Discrimination Parameters as Factor Loadings*

At this stage it is helpful to note a close connection between IRT models and factor analytic models. Recall that in exploratory factor analysis (EFA), the researcher typically has weak prior beliefs as to the substantive content of the dimensions underlying a set of variables. The estimated factor loadings tell the researcher which variables are explained by which dimension. In fact, in a genuinely *exploratory* factor analysis, variables with large factor

[3] See Clinton and Mierowitz (2001) for an example of assumptions that permit recovery of the Yea and Nay proposal.

loadings on dimension d supply the substantive content of that dimension. In this sense discrimination parameters are extremely similar to factor loadings; the substantive content of the dimensions recovered by an ideal point model is revealed by inspecting the proposals that discriminate with respect to each dimension.

The similarities between multidimensional item–response (MIRT) models and factor analytic models for binary data are well known (e.g., Takane and de Leeuw 1987; Reckase 1997), and indeed, MIRT is sometimes referred to as "full information item factor analysis" (Bock *et al*. 1988).[4] And in the specific context of roll call data, Heckman and Snyder (1997) use factor analytic methods to estimate legislators' preferences, although motivated by a different set of assumptions than typically used in roll call analysis.[5]

The connection between MIRT and factor analysis is helpful in terms of interpreting the model parameters and, in particular, how discrimination parameters help us discern the nature of the recovered dimensions. But having noted this analogy, three important distinctions between our MIRT model and a traditional factor analytic model warrant mention. First, factor analysis is fundamentally a model for correlations and not a model of the individual level responses. Factor analyses collapse individual level responses to form a correlation matrix, discarding information about the means and the variances of the input variables. Information is necessarily lost in this way, making it difficult to learn simultaneously about the locations of legislators (the $\mathbf{x}_i$) and properties of the proposals (the discrimination parameters $\boldsymbol{\beta}_j$ and the difficulty parameters, α_j). Contrast the MIRT framework, where the individual binary responses (the Yeas and Nays, y_{ij}) are modeled directly as functions of the parameters of substantive interest.

Second, factor analytic models for binary data are much harder to estimate than traditional factor analytic models presuming multivariate normal data; for example, see Heckman and Snyder's (1997, p. 161) discussion of the difficulties they encounter estimating their factor analysis model with binary roll call data. In contrast, our IRT/MIRT model confronts the binary character of roll call data directly, since it is a direct operationalization of the presumed (binary) data generation mechanism, the Euclidean spatial voting model.

Third, in our Bayesian approach, we supply informative priors on certain $\boldsymbol{\beta}_j$ to "predefine" the specific dimensions. The factor analytic analogue is *confirmatory* factor analysis (CFA), where researchers have beliefs about how specific variables relate to specific dimensions. As I show below, priors are especially helpful in resolving a second dimension in the 105th U.S. Senate data.

4 Moving to Higher Dimensions: Identification Strategies

Almost all models for latent traits have unidentified parameters. The two-paramter IRT model is no exception. In one dimension this is rather obvious: without any constraints on the model parameters, $p_{ij} = F(x_i\beta_j - \alpha_j) = F(x_i^*\beta_j^* - \alpha_j)$, where $x_i^* = cx_i$ and $\beta_j^* = c^{-1}\beta_j$, for any scaling factor $c \neq 0$. That is, the model parameters $(x_i, \boldsymbol{\beta}_j)$ are not identified. Non-Bayesian analyses of roll call data solve this *scale invariance* problem in two ways. Poole and Rosenthal (1997) constrain their estimates of $\mathbf{x}_i$ to lie in the $[-1, 1]^D$ hypercube. Londregan (2000b) sets specific $\mathbf{x}_i$ to fixed values, treating the corresponding legislators as "reference legislators." In the Bayesian context we solve this problem with a proper prior density over the $\mathbf{x}_i$; we specify $\mathbf{x}_i \stackrel{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{I}_D)$, $\forall i$, or more formally, $\pi(\mathbf{X}) = \prod_{i=1}^{n} \boldsymbol{\phi}_D(\mathbf{x}_i)$,

[4] Moreover, the invention of maximum-likelihood factor analysis by Lawley in the early 1940s was in the context of analyzing item–response data.

[5] Note that Heckman and Snyder construct their model so as to recover the ideal points $\mathbf{x}_i$ as the factor loadings.

where $\boldsymbol{\phi}_D$ is a D-dimensional standard normal density. These priors are innocuous in that they solve the scale invariance problem and no more; each legislator is assigned the same prior, and the selection of prior variance is arbitrary.

Precisely because these priors are so innocuous, the resulting posterior densities are generally not unimodal. That is, while solving *scale invariance*, these priors do not rule out *rotational invariance*. Again, consider the unidimensional case. Our priors do not preclude multiplying all model parameters by -1, reversing the orientation of the recovered policy dimension, but yielding an identical fit to the data. That is, the resulting posterior density $\pi(\theta \mid \mathbf{Y})$ is a two-component *mixture*, one component with a mode at θ_0 and another "mirror image" component with a mode at $-\theta_0$; put differently, our priors ensure local identification, not global identification (Bollen 1989, p. 248). In practice, this is not a problem for most unidimensional models of roll calls. Our experience is that with (a) a reasonably large number of roll calls and (b) reasonable discrimination along the underlying policy continuum (as occurs in even mildly partisan legislatures), the two component modes are extremely distant from one another, and the component densities are compact around their respective modes. This gives us a choice of which mode to use for inferences; we choose the mode that places Democrats on the left and Republicans on the right, in accord with widely accepted senses of "left" and "right."

Recall that we use a Gibbs sampler to explore this posterior density randomly; our experience is that the sampler quickly moves from its starting values to one of these component modes, and *never* visits the other component.[6] Accordingly, a practical solution to the rotational invariance problem is simply to start the Gibbs sampler with all Democratic legislators on the left of the origin and all Republican legislators on the right. If the components of the posteriors for all the ideal points were diffuse and centered close to zero, then this strategy would be impractical (since the two components of the posterior could not be clearly resolved), but this is seldom the case for roll call data sets. In short, the fact that model parameters are unidentified is rarely of practical importance in unidimensional settings.

But moving to two dimensions dramatically increases the number of observationally equivalent rotations. We have just seen that in one dimension, our $x_i \sim N(0, 1)$ prior generates a bimodal posterior density. But in two dimensions, our priors $\mathbf{x}_i \sim N(\mathbf{0}, \mathbf{I}_D)$ admit *eight* observationally equivalent rotations; each dimension can be reversed, while at the same time the dimensions can be swapped, yielding $2 \times 2 \times 2 = 8$ rotations.[7] Put differently, the posterior density of any ideal point $\mathbf{x}_i = (x_{i1}, x_{i2})'$ is a mixture of eight component densities.

Figure 1 shows four possibilities for the case of $D = 2$, with each bivariate posterior density shown as a surface in the left column and as a contour plot in the right column. When the components are tightly distributed around their respective modes (as in the top row in Fig. 1), it is easy to resolve the individual components, and researchers can simply pick a mode for inference and communicating results to readers. But when the component densities are not tightly distributed and overlap (lower panels in Fig. 1), the component modes are no longer easily resolved and may disappear. As the precision of the posterior

[6]The Gibbs sampler is effectively "trapped" at the mode closest to its initial state, and the probability of jumping the wide region of the parameter space separating the two modes (where the posterior density is essentially zero) is *extremely* low and, hence, never occurs even in an extremely long run of the Gibbs sampler.

[7]Let the $n \times 2$ matrix $[\mathbf{a}\ \mathbf{b}]$ denote a joint mode of one of the components of the posterior for the ideal points. Then the following points are also component modes: $[\mathbf{a}\ -\mathbf{b}]$, $[-\mathbf{a}\ \mathbf{b}]$, $[-\mathbf{a}\ -\mathbf{b}]$, $[\mathbf{b}\ \mathbf{a}]$, $[\mathbf{b}\ -\mathbf{a}]$, $[-\mathbf{b}\ \mathbf{a}]$, and $[-\mathbf{b},\ -\mathbf{a}]$.

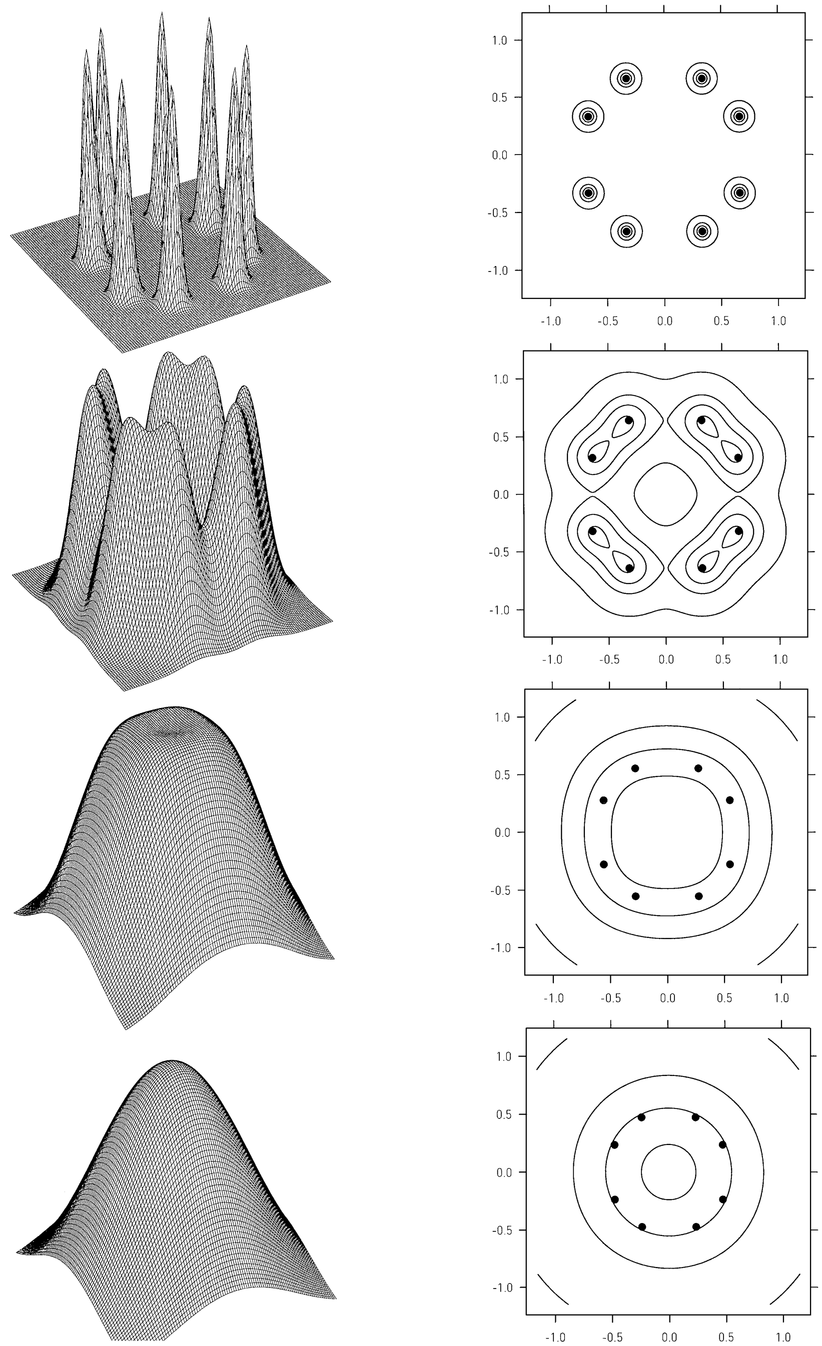

Fig. 1 Rotational invariance and the posterior density for $\mathbf{x}_i$. Four posterior densities are shown (surface plots in the left column, contour plots on the right). The posterior density for each $\mathbf{x}_i$ is a mixture of eight component densities (the mode of each component is marked with a filled dot in the contour plots). If the components have low dispersion (i.e., the data are highly informative about the ideal points), then the mode of each component is clearly resolved (top row). As the data become less informative (moving down the page), the eight individual component densities are no longer apparent, and the posterior density tends toward the $N(\mathbf{0}, \mathbf{I}_D)$ prior (bottom panels).

components diminshes (i.e., the roll call data are increasingly uninformative about the ideal points), we wind up with the limiting case seen in the bottom row in Fig. 1, a unimodal bivariate density centered on the origin [i.e., the $N(\mathbf{0}, \mathbf{I}_D)$ prior].

Again, this lack of identification is of no substantive consequence *if the data are highly informative about the ideal points*. But this is not always the case when fitting higher-dimensional models to roll call data. If a unidimensional model provides a good fit to the data, then it is quite likely that there is little information in the data about the second dimension parameters. In this case there is a real risk that the posteriors for the ideal points start to resemble the cases described in the lower panels in Fig. 1. This is especially relevant given our use of a Gibbs sampler, which (if let run long enough) will visit each component density and faithfully reproduce the mixture densities shown in Fig. 1. If the component densities overlap then a naive use of a Gibbs sampler might result in the inferences that are technically correct but substantively misleading. It is worth stressing that this can hardly be considered a weakness of the Bayesian approach; this problem stems from the attempt to estimate a model with unidentified parameters! If anything, the Bayesian approach lets us see the nature of this problem more clearly, by focusing our attention on the entire posterior density rather than on a search for local/global maxima.

In the Bayesian context, priors are an obvious way to reduce the number of components in the posterior density. Legislator-specific priors are one way to proceed, say, by specifying priors that constrain an Edward Kennedy to be on the “left” on one dimension and, on that same dimension, a Jesse Helms constrained to be on the “right” (a Bayesian analogue of Londregan’s “reference” legislator approach). Of course, this is harder to do in higher dimensions, particularly when the researcher has only weak ideas as to the nature of those dimensions or who might reasonably serve as reference legislators on those dimensions.

Our preferred solution is to use informative priors on the discrimination parameters, $\boldsymbol{\beta}_j$. That is, we treat various proposals as “reference proposals” for specific dimensions, with the substantive content of those proposals providing an initial guess as to the nature of the corresponding dimension. Only a small number of proposals need informative priors so as to reduce the number of components in the posterior. For instance, consider priors that make proposal j discriminate with respect to dimension 1 but provide no discrimination in dimension 2 ($\beta_{j1} > 0, \beta_{j2} = 0$). Then if (a, b) is a posterior mode for $\mathbf{x}_i = (x_{i1}, x_{i2})'$, then $(-a, b)$ and $(-a, -b)$ cannot be posterior modes, since this would violate the a priori constraint that $\beta_{j1} > 0$. Likewise, any swapping of dimensions is ruled out by this prior. This leaves $(a, -b)$ as the only other possible posterior mode, which the researcher can either (a) ignore, which is feasible if the two posterior modes are well separated and easily resolved, or (b) eliminate via informative priors on another “reference proposal.” In the next section I show how priors help fit a two-dimensional model to the 105th U.S. Senate data.

5 Data: 105th U.S. Senate

I begin with a moderately large roll call data set analyzed by Jackman (2000): $m = 486$ nonlopsided roll calls from the 105th U.S. Senate, which sat from January 1997 through October 1998.[8] Estimates of each of the $n = 100$ senators’ ideal points (accompanied by confidence bounds) for a one-dimensional model are given in my earlier *Political Analysis*

[8]Roll calls with fewer than three legislators voting for or against the proposal were dropped from the analysis, consistent with Poole and Rosenthal’s definition of lopsidedness.

article (Jackman 2000, Fig. 5); the striking feature of those results is that there is no partisan overlap on the recovered policy dimension, even acknowledging the uncertainty in the estimated ideal points.

6 Assessing Dimensionality via Discrimination Parameters

With a one-dimensional model, 446 of the 486 β_j discrimination parameters are distinguishable from zero (i.e., the 5th and 95th percentiles of the posterior for the respective β_j lie on the same side of zero). That is, 91.8% of the proposals discriminate on the unidimensional policy continuum recovered by this model. In the language of factor analysis, these 446 bills load onto the single factor, the policy dimension recovered by the unidimensional model.

Figure 2 shows that of the 40 proposals that do not discriminate among legislators (Fig. 2, right), most are highly lopsided votes, with close to all legislators voting for the proposals. These lopsided votes are equivalent to test items that are too easy, which most test-takers pass, and hence provide little discrimination. Note that there are hardly any proposals that result in close roll calls while simultaneously failing to discriminate among legislators. Again, this suggests that moving to a higher-dimensional model is unlikely to produce a marked improvement in fit to the data. In addition, the left panel of Fig. 2 shows the converse, that close roll calls provide the greatest discrimination among legislators. The greatest discrimination occurs for votes decided by 55–45 margins, which was the partisan composition of the 105th Senate, confirming the strong partisan polarization recovered by the one-dimensional model.

In short, these inspections suggest that a one-dimensional model is a reasonable fit to the data. Roughly 92% of the roll calls discriminate with respect to the the recovered policy

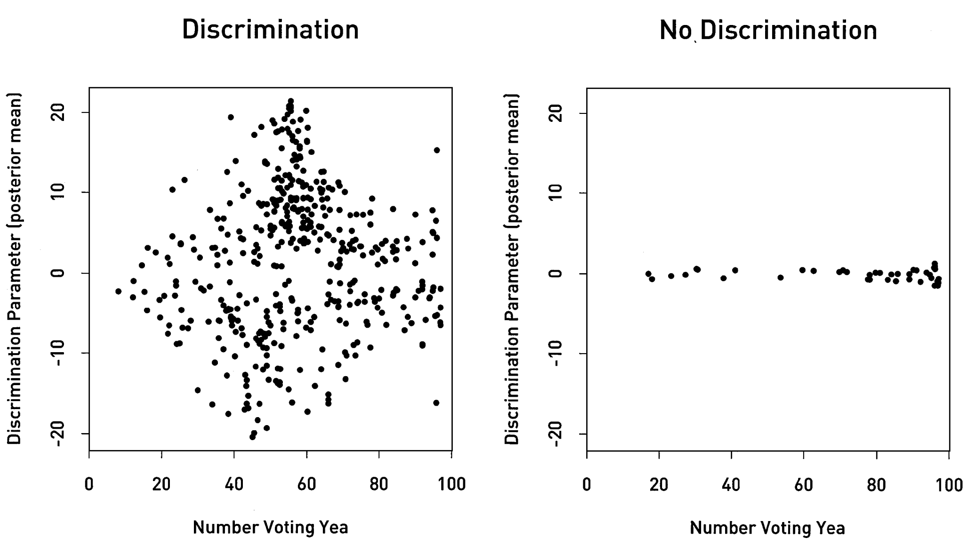

Fig. 2 Discrimination parameters and roll call margins, one-dimensional model, 105th U.S. Senate. The posterior means for the discrimination parameters [the slope parameters $\boldsymbol{\beta}_j$ in Eq. (1), equivalent to factor loadings] are plotted on the vertical axis, and the number of legislators voting Yea on the horizontal axis. Roll calls with statistically significant discrimination appear on the left (90% confidence bounds do not overlap zero); 40 roll calls (8.2% of the total number of roll calls analyzed) with no statistically significant discrimination appear on the right panel.

continuum, and 87.78% of the individual votes (discarding abstentions and absences) are correctly classified.[9] Below I consider whether these suggestions are correct, by considering the performance of a two-dimensional model.

7 Priors for a Two-Dimensional Model

Initial attempts to fit two-dimensional models with vague priors on the proposal parameters generated posterior densities like those in the second and third rows in Fig. 1. The data simply are not sufficiently informative about the model parameters to let us resolve the different components of the posterior density. Consistent with the discussion in Section 4, I reduce the complexity of the posterior density via informative priors over selected proposal parameters. But which proposals?

Consider the substantive content of the roll calls that *do not* discriminate in one dimension. Do these proposals have anything in common substantively, accounting for why they fail to discriminate on the highly partisan dimension recovered by the unidimensional model? If so, this common basis is likely to form the basis of a second dimension. Of the 40 roll calls that fail to discriminate, five occur on one day, April 30, 1998 (roll calls 111, 112, 115, 117, and 118, 2nd session, 105th Senate), and one is a reasonably close vote (roll call 112; failing 41–59). Inspection of *Congressional Record* shows that with the exception of 118, these roll calls were all on amendments and final passage of a resolution ratifying the expansion of NATO to include Poland, Hungary, and the Czech Republic. I let two of these bills supply my initial guess as to the substance of the second dimension in the data. I do this by specifying informative priors on the discrimination parameters for roll calls 111 (a proposal to defer ratification until Poland, Hungary, and the Czech Republic joined the EU) and 117 (final passage; passing 80–19); these votes are near-mirror images of one another, with just four senators voting Nay on both roll calls and just two senators voting Yea on both. Thus I am quite confident that these two roll calls tap the same underlying policy dimension; strong associations are also apparent with the other NATO-expansion roll calls. The priors assigned to discrimination parameters for these roll calls are

$$\pi(\boldsymbol{\beta}_{111}) = N\left(\begin{bmatrix} 0 \\ -4 \end{bmatrix}, \begin{bmatrix} .01 & 0 \\ 0 & 4 \end{bmatrix}\right) \quad \text{and} \quad \pi(\boldsymbol{\beta}_{117}) = N\left(\begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} .01 & 0 \\ 0 & 4 \end{bmatrix}\right)$$

such that support for the expansion of NATO is associated with positive movement on dimension two. Note the tight prior around zero for the dimension 1 discrimination parameter, ensuring that this roll call discriminates exclusively on the second recovered dimension. The choice of 4 as the magnitude of the informative prior mean on the second dimension is somewhat arbitrary. Guidance in choosing the prior mean comes from inspecting the statistically significant β_j for the one-dimensional model: the magnitudes of these parameters have an interquartile range of 3.8 to 10.5, and so my choice of 4.0 is toward the lower end of this range, but fairly typical for bills decided by 80–20 margins (see Fig. 2, left). The prior variances are chosen such that the prior means are two prior standard deviations away from zero.[10] As in the one-dimensional model, all other

[9]Classifications were performed by generating predicted probabilities of each individual voting decision with all model parameters set to their posterior means and using a classification threshold of 0.5.

[10]An informative prior on just one of these roll calls produces results that are almost indistinguishable from those reported here. Note that these priors do not rule out $\boldsymbol{\beta}_{111,2} > 0$ or $\boldsymbol{\beta}_{117,2} < 0$, and in fact, the prior probability of either event is roughly 0.025. Thus each prior alone is not sufficient to stop the Gibbs sampler visiting posterior

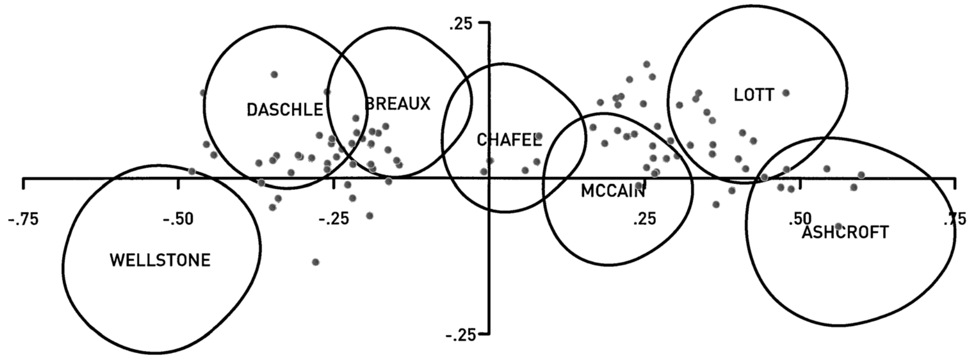

Fig. 3 Posterior means, ideal points of the 105th U.S. Senate, two-dimensional spatial voting model. Lighter points indicate Democrats, to the left of the policy space; darker points indicate Republicans, to the right of the policy space. Ninety-five percent confidence regions are displayed for the indicated senators; the names of these senators are centered on the location of the respective posterior means.

discrimination parameters and all the intercept parameters α_j are assigned a vague $N(0, 100)$ prior.

A Gibbs sampler was used to sample from this posterior density; 1.5 million samples were generated, with the first 100,000 iterations discarded as "burn-in" (letting the sampler move away from its initial values to the neighborhood of the joint posterior mode) and then every 2000th Gibbs sample retained for inference. The retained Gibbs samples generally have low to moderate autocorrelations across iterations. The results reported below are thus based on 700 reasonably independent samples from the joint posterior density for the model parameters; summaries of these samples (e.g., means and confidence intervals) can thus be regarded as valid characterizations of the joint posterior. Further details on the Gibbs sampler are available upon request, as is the computer code for the sampler.

8 Results

Figure 3 shows the posterior means of the $\mathbf{x}_i$ for all 100 senators, with 95% confidence regions displayed for seven senators.[11] Since the use of priors solves the scale invariance problem, the metric information conveyed by Fig. 3 is meaningful. That is, distances on the horizontal dimension are in the same metric as distances on the vertical dimension. In this case, the fact that there is considerably more variation along the horizontal dimension is a direct reflection of the fact that the horizontal dimension is a more important determinant of the roll calls than the vertical dimension.

---

modes with, say, $\beta_{117,2} < 0$ but is sufficient to make visits to those modes rare and clearly distinguishable from the "desired" mode with $\beta_{117,2} > 0$. But with priors on both parameters, the joint event $\beta_{111,2} > 0$ *and* $\beta_{117,2} < 0$ becomes very rare and is sufficient to ensure the posterior density is unimodal.

[11] The confidence regions are formed by first fitting a two-dimensional density over the output of the Gibbs sampler, using Loader's (1999) local fitting density routines (using a tricube kernel and a $\alpha = 0.50$ nearest-neighbor bandwidth). The fitted density is then used to generate a 95% highest-density region (HDR), using algorithms of Hyndman (1996). Formally, then, each plotted confidence region is an estimate of that region of the parameter space supporting the upper 95% of the posterior density for the respective ideal point. As the the number of roll calls (m) increases, these posteriors tend to a normal density (Chang and Stout 1993), and each HDR converges to an ellipse. Since $m = 486$, the slight irregularities in the HDRs in Fig. 3 could reflect a departure from asymptotic normality but more likely reflect Monte Carlo error; with a longer run of the Gibbs sampler, the HDRs will tend toward ellipses.

Table 1 Model comparisons: spatial voting models fit to the 105th U.S. Senate[a]

	One dimension	*Two dimensions*
log-likelihood	−12,965.14	−11,844.19
$df = nD + (m+1)D$	1,072	1,658
AIC	28,074.28	27,004.38
Schwarz	30,867.03	31,323.75
Classification rates (%)		
All	87.78	88.62
Yeas	90.43	90.85
Nays	84.04	85.49
Democrats	86.74	87.72
Republicans	88.62	89.36
Best, by senator	94.70 (Kennedy)	95.89 (Akaka)
Worst, by senator	71.13 (Byrd)	71.13 (Byrd)
Worst, by roll call	54.0	58.6
Perfect roll calls (No.)	13	14
Mean PRE	0.524	0.561

[a]All quantities are generated by generating predicted probabilities with model parameters set to their posterior means. The classification threshold is 0.5. Perfect roll calls are those with no classification errors. Following Weisberg (1978), Poole and Rosenthal (1991), and Heckman and Snyder (1997), PRE is defined relative to a null model, in which all senators vote with the majority on each roll call, and then averaged across the m roll calls.

Another significant feature of these data is the strong separation by party along the horizontal dimension. Even taking into account the reasonably large confidence intervals, there is extremely little overlap by party. On the other hand, senators are not clearly distinguished on the vertical dimension; not only are the senators less dispersed vertically than horizontally, but the relatively wide confidence intervals mean that it is extremely difficult to authoritatively order the legislators along this dimension. This is one indication that little has been gained from fitting this second dimension.

Table 1 presents a detailed set of model comparisons, formed by generating predicted probabilities with the model parameters set to their posterior means. The use of informative priors means that a classical likelihood ratio test cannot be interpreted in the usual way (the reported log-likelihoods are actually mean posterior likelihoods; space precludes consideration of formal Bayesian model comparisons). With this important caveat in mind, the critical value of the χ^2 test is 643.42, while twice the difference in the recovered log-likelihoods is 1120.95, overwhelmingly rejecting the restrictions implied by the one-dimensional model. This improvement in likelihood more than offsets the penalty for lack of parsimony tapped by the Akaike information criterion, which prefers the two dimensional model. On the other hand, the Schwarz criterion imposes a bigger penalty for additional parameters and prefers the simpler one-dimensional model. Comparing a variety of classification measures also suggests that little is gained in the way of predictive power in moving to higher dimensions. For instance, the rate of correct classification for this two-dimensional model is 88.62%, compared with 87.78% for the one-dimensional model, a proportional increase of

Table 2 Comparison of discrimination parameters, one- and two-dimensional models, 105th U.S. Senate[a]

One-dimensional model	*Two-dimensional model* “Pure” dimension 1	*“Pure” dimension 2*	*Both dimensions*	*No discrimination*	*Total*
Discrimination	282	12	151	1	92%
No discrimination	0	26	0	14	8%
Total	58%	8%	31%	3%	

[a]Proposals are considered to discriminate if the 90% confidence interval of the posterior density for the respective slope parameter does not overlap zero.

just 0.97%, or in absolute terms, 400 more correct classifications of 47,739 individual voting decisions. This is an extremely modest increase given that the two-dimensional model has $n + m = 586$, or 54.6% more parameters than the one-dimensional model.

Another way to compare the two models is to ask if moving to a two-dimensional model increases the number of roll calls with nonzero discrimination parameters? That is, does the two-dimensional model explain more roll calls than the one-dimensional model? And if so, *which* roll calls discriminate with respect to which dimensions?

Table 2 shows the cross-classification of roll calls by whether they discriminate among legislators for the two models presented here. Discrimination parameters are again considered indistinguishable from zero if the 90% confidence interval of the respective posterior density overlaps zero, the equivalent of a classical one-tailed test at $p = .05$. In the two-dimensional model, roll call j can either (1) discriminate solely on dimension 1 ($\beta_{j1} \neq 0$, $\beta_{j2} = 0$), (2) discriminate solely on dimension 2 ($\beta_{j1} = 0$, $\beta_{j2} \neq 0$), (3) discriminate on both dimensions ($\beta_{j1} \neq 0$, $\beta_{j2} \neq 0$), or (4) fail to discriminate on either dimension ($\beta_{j1} = 0$, $\beta_{j2} = 0$). The first feature to note is that, in moving to the two-dimensional model, additional roll calls are rationalized in terms of the recovered spatial structure. Ninety-seven percent of all roll calls discriminate on either or both of the dimensions recovered by the two-dimensional model, a net improvement of 25 roll calls over the one-dimensional model. One hundred eighty-nine (39%) roll calls load on dimension 2, of which 38 are “pure” dimension two roll calls.[12] Of these 38 roll calls, 26 (roughly two-thirds) did not discriminate in the one-dimensional model, underlining the fact that this dimension is nearly orthogonal to the dimension recovered by the one-dimensional model, via the use of informative priors on selected roll calls.

These 38 roll calls supply the substantive content of the second dimension. Given the priors specified earlier, the NATO ratification roll calls discriminate with respect to dimension two and will supply part of whatever substantive content vests in the dimension. But what else? That is, having estimated the model, is it possible to replace the label “dimension 2” with a politically meaningful label? Of course, in the Bayesian approach it is always possible to assign substantive labels to dimensions a priori, by specifying informative priors on the discrimination parameters for more roll calls. Given my spartan use of priors (just

[12]Another useful way to think about “pure” dimension 1 or “pure” dimension 2 roll calls is to note that their indifference lines or cutting planes are parallel with the dimensions. The slope of the jth cutting plane is $-\beta_{j1}/\beta_{j2}$, so pure dimension 2 roll calls have horizontal cutting planes, while pure dimension 1 roll calls have vertical cutting planes.

Table 3 Description of selected "pure" dimension 2 roll calls

Date	*Number*	*Yea–Nay*	β_{j2}[a]	*Substantive area*
6/27/97	155	30–69	−6.0	Ethanol tax exemption
7/23/97	196	53–47	4.8	Tobacco subsidy
7/23/97	199	59–40	10.8	Market Access Program (promoting U.S. goods in foreign markets)
10/1/97	264	55–45	9.7	Senator's pay
10/30/97	287	69–30	7.9	Military spending, line-item veto override
11/4/97	292	69–31	6.9	Reciprocal Trade Agreement ("fast track"), cloture
11/5/97	294	68–31	8.0	Reciprocal Trade Agreement ("fast track")
3/11/98	27	71–26	6.8	Ethanol tax exemption
3/23/98	39	61–31	11.9	Miscellaneous items from natural disaster spending
4/30/98	112	41–59	−6.5	NATO expansion
5/21/98	145	37–61	5.4	Tobacco industry liability
7/15/98	201	53–46	−5.1	Trade sanctions
7/15/98	202	70–29	11.5	Market Access Program
7/16/98	206	71–28	11.7	Declarative resolution ("sense of the Senate") concerning various trade-related items, including fast-track, IMF funding, sanctions, regulations on farming
10/21/98	314	65–29	10.4	Miscellaneous items from omnibus emergency appropriations

[a]The posterior mean of the respective dimension 2 discrimination parameter.

4 of the $m \cdot D = 972$ discrimination parameters are assigned informative priors), the data supply most of the substantive content of the recovered dimensions.

Table 3 lists a selection of "pure" dimension two proposals with reasonably nonlopsided votes, along with a summary of the substantive content of each proposal. Trade dominates this list (especially in agricultural-related trade issues), although a number of issues appear in the list: e.g., overriding President Clinton's line-item veto of certain military spending projects, attempts to remove pork-barrel spending from emergency appropriations, and senators' own salaries. In sum, my priors on two of the NATO ratification votes supply enough information for a second dimension to be resolved in the data. Inspection of the roll calls reveals this second dimension to be akin to a "free-trade/internationalist" dimension, but this label does not exhaust the substantive content of the dimension.

9 Conclusion

My focus in this article has been on parameters specific to roll calls ($\boldsymbol{\beta}_j$) rather than legislator-specific ideal points ($\mathbf{x}_i$). Discrimination parameters almost always play a secondary role in the analysis of roll call data, the primary goal usually being the measurement of the ideal points. But discrimination parameters are the functional equivalents of factor loadings, and just as in factor analysis, they (1) enable researchers to discern the substantive content of the recovered dimensions, (2) can be used for goodness-of-fit assessments, and (3) are an obvious vehicle for introducing prior information into roll call analyses. These interpretations of discrimination parameters are greatly facilitated by Bayesian simulation methods, which simplify estimation and inference for the massive number of parameters generated by roll call analysis.

In the specific data set at hand, priors were needed to help resolve a second-dimension in the 105th U.S. Senate data. The informative priors I specified for just two roll calls were adopted largely for convenience, reducing the number of components in the posterior density, thereby "improving" identifiability. But priors need not play a mere technical role, and indeed, I chose priors by closely inspecting the substance of the roll calls. And more generally, researchers with interests in, say, trade, foreign policy, or environmental policy can specify informative priors for the discrimination parameters of key roll calls; in an extremely crude way, this is how interest group ratings get constructed. These priors effectively "prelabel" one or more of the dimensions presumed to underlie the roll calls. Analysis can then proceed entirely consistent with the spatial voting model, but incorporating the researcher's prior beliefs about the dimensions underlying the policy space. Researchers might then investigate what other roll calls discriminate with respect to these dimensions, which may be helpful in better understanding the politics underlying the observed roll calls and understanding legislative politics more generally. Other models might be considered, embodying an alternative set of prior beliefs linking specific roll calls and the dimensions of the proposal space.

This is not to say that Bayesian approaches are the only way to make roll call analysis substantively richer. For instance, Londregan's (2000b) innovative analyses of voting in Chilean Senate committees are wholly within a classical MLE framework, albeit with small roll call data sets and assumptions that (rather dramatically) reduce the number of proposal parameters being estimated. On the other hand, Clinton and Mierowitz's (2001) roll call analysis incorporates an assumption of how the legislative agenda progresses over time (using the first U.S. Congress as an example) and is wholly within a fully Bayesian framework. In short, the goal that Bayesian methods make plausible is a transformation of roll call analysis, from a technical *scaling* or *measurement* problem best left to psychometricians (witness the canonical status of NOMINATE scores) to something that scholars motivated primarily by substantive concerns can do for themselves.

References

Albert, James. 1992. "Bayesian Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling." *Journal of Educational Statistics* 17:251–269.

Bock, R. D., and M. Aitken. 1981. "Marginal Maximum Likelihood Estimation of Item Parameters: Application of An EM Algorithm." *Psychometrika* 46:443–459.

Bock, R. D., R. Gibbons, and E. J. Muraki. 1988. "Full Information Item Factor Analysis." *Applied Psychological Measurement* 12:261–280.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Chang, H. H., and W. Stout. 1993. "The Asymptotic Posterior Normality of the Latent Trait in an IRT Model." *Psychometrika* 58:37–52.

Clinton, Joshua D., and Adam Mierowitz. 2001. "Agenda Constrained Legislator Ideal Points and the Spatial Voting Model." *Political Analysis* 9:242–259.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2000. "The Statistical Analysis of Legislative Behavior: A Unified Approach," Paper presented to the Southern California Area Methodology Program, University of California, Santa Barbara, May 12–13.

Enelow, J., and M. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. New York: Cambridge University Press.

Heckman, James J., and James M. Snyder. 1997. "Linear Probabilty Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *RAND Journal of Economics* 28:S142–S189.

Hyndman, Rob J. 1996. "Computing and Graphing Highest Density Regions." *American Statistician* 50:120–126.

Jackman, Simon. 2000. "Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8:307–332.

Johnson, Valen E., and James H. Albert. 1999. *Ordinal Data Modeling*. New York: Springer-Verlag.

Loader, C. 1999. *Local Regression and Likelihood*. New York: Springer.

Londregan, John. 2000a. "Estimating Legislators' Preferred Points." *Political Analysis* 8:35–56.

Londregan, John. 2000b. *Legislative Institutions and Ideology in Chile's Democratic Transition*. New York: Cambridge University Press.

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

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Reckase, Mark D. 1997. "The Past and Future of Multidimensional Item Response Theory." *Applied Psychological Measurement* 21:25–36.

Takane, Yoshio, and Jan de Leeuw. 1987. "On the relationship between item response theory and factor analysis of discreteized variables." *Psychometrika* 52:393–408.

Weisberg, Herbert F. 1978. "Evaluating Theories of Congressional Roll-Call Voting." *American Journal of Political Science* 22:554–577.