The Statistical Analysis of Roll Call Data

Joshua Clinton[1]
Simon Jackman,
and
Douglas Rivers[2]

May 4, 2003

[1]Department of Politics, Princeton University. E-mail clinton@princeton.edu

[2]Department of Political Science, Stanford University, Stanford, California 94305-6044, USA. E-mail: jackman@stanford.edu and rivers@stanford.edu. We thank Larry Bartels, David Brady, Richard Brody, Jim Fearon, John Ferejohn, Rui de Figueiredo, Tim Groseclose, Keith Krehbeil, John Londregan, Adam Meriowitz, Ron Mitchell, Keith Poole, David Primo, and Mike Tomz for useful comments.

Abstract

We develop a Bayesian procedure for estimation and inference for spatial models of roll call voting. Our approach is extremely flexible, applicable to any legislative setting, irrespective of size, the extremism of the legislative voting histories, or the number of roll calls available for analysis. Our model is easily extended to let other sources of information inform the analysis of roll call data, such as the number and nature of the underlying dimensions, the presence of party whipping, the determinants of legislator preferences, or the evolution of the legislative agenda; this is especially helpful since generally it is inappropriate to use estimates of extant methods (usually generated under assumptions of sincere voting) to test models embodying alternate assumptions (e.g., log-rolling). A Bayesian approach also provides a coherent framework for estimation and inference with roll call data that eludes extant methods; moreover, via Bayesian simulation methods, it is straightforward to generate uncertainty assessments or hypothesis tests concerning any auxiliary quantity of interest or to formally compare models. In a series of examples we show how our method is easily extended to accommodate theoretically interesting models of legislative behavior. Our goal is to move roll call analysis away from pure measurement or description towards a tool for testing substantive theories of legislative behavior.

1. Introduction

Modern studies of legislative behavior focus upon the relationship between the policy preferences of legislators, institutional arrangements, and legislative outcomes. In spatial models of legislatures, policies are represented geometrically, as points in a low dimensional Euclidean space. Each legislator has a most preferred policy or *ideal point* in this space and his or her utility for a policy declines with the distance of the policy from his or her ideal point; see Davis, Hinich and Ordeshook (1970) for an early survey.

The primary use of roll call data---the recorded votes of deliberative bodies[1]---is the estimation of ideal points. The appeal and importance of ideal point estimation arises in two ways. First, ideal point estimates let us *describe* legislators and legislatures. The distribution of ideal points estimates reveals how cleavages between legislators reflect partisan affiliation or region, or become more polarized over time (e.g., McCarty, Poole and Rosenthal 2001). Roll call data serves similar purposes for interest groups, such as Americans for Democratic Action, the National Taxpayers Union, and the Sierra Club, to produce "ratings" of legislators along different policy dimensions. Second, estimates from roll call analysis can be used to *test theories* of legislative behavior. For instance, roll call analysis has been used in studies of U.S. Congress, both contemporary and historical (e.g., Canes-Wrone, Brady and Cogan 2002; Schickler 2000; Jenkins 1999), state legislatures (e.g., Wright and Schaffner 2002), studies of courts (Martin and Quinn 2001), comparative politics (Londregan 2000b) and international relations (Voeten 2000). In short, roll call analysis make conjectures about legislative behavior amenable to quantitative analysis, helping make the study of legislative politics an empirically-grounded, cumulative body of scientific knowledge.

Current methods of estimating ideal points in political science suffer from both statistical and theoretical deficiencies. First, any method of ideal point estimation embodies an explicit or implicit model of legislative behavior. Generally, it is inappropriate to use ideal points estimated under one set of assumptions (such as sincere voting over a unidimensional policy space) to test a different behavioral model (such as log-rolling). Second, the computations required for estimating even the simplest roll call model are very difficult and extending these models to incorporate more realistic behavioral assumptions is nearly impossible with extant methods. Finally, the statistical basis of current methods for ideal point estimation is, to be polite, questionable. Roll call analysis involves very large numbers of parameters, since each legislator has an ideal point and each bill has a policy location that must be estimated. Popular methods of roll call analysis compute standard errors that are admittedly invalid (Poole and Rosenthal 1997, 246) and one cannot appeal to standard statistical theory to ensure the consistency and other properties of estimators (see section 3, below).

In this paper we develop and illustrate Bayesian methods for ideal point estimation. Bayesian inference provides a coherent method for assessment of uncertainty and hypothesis testing in the presence of large numbers of parameters, and recent advances in computing put Bayesian modeling (via Monte Carlo simulation) well within the reach of social scientists. Using our approach, we

[1] The same model is applicable to courts, though the data are called "decisions," not "roll calls."

show how it is possible to extend the basic estimation technique to accommodate more complex behavioral assumptions. Our goal is to move roll call analysis away from pure measurement or description and to make it a tool for testing substantive theories of legislative behavior.

2. A Statistical Model for Roll Call Analysis

In theoretical work on spatial voting models, utility functions are usually deterministic and the precise functional form, aside from an assumption of convexity, is not specified. For empirical work, it is convenient to choose a parametric specification for the utilities and to add a stochastic disturbance. Several different specifications have been used, but all are quite similar. We assume a quadratic utility function for legislators with normal errors. Poole and Rosenthal (1997) assume Gaussian utilities with extreme value errors. Heckman and Snyder (1997) assume quadratic utilities with uniform errors for one of the alternatives and non-stochastic utility for the other. See Table 2 for a comparison of the specifications.

The data consist of n legislators voting on m different roll calls. Each roll call $j = 1, \ldots, m$ presents legislators $i = 1, \ldots, n$ with a choice between a "Yea" position $\boldsymbol{\zeta}_j$ and a "Nay" position $\boldsymbol{\psi}_j$, locations in $\mathbb{R}^d$, where d denotes the dimension of the policy space. Let $y_{ij} = 1$ if legislator i votes "Yea" on the jth roll call and $y_{ij} = 0$ otherwise. Legislators are assumed to have quadratic utility functions over the policy space, $U_i(\boldsymbol{\zeta}_j) = -||\mathbf{x}_i - \boldsymbol{\zeta}_j||^2 + \eta_{ij}$ and $U_i(\boldsymbol{\psi}_j) = -||\mathbf{x}_i - \boldsymbol{\psi}_j||^2 + \nu_{ij}$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the *ideal point* of legislator i, and η_{ij} and ν_{ij} are the errors or stochastic elements of utility, and $||\cdot||$ is the Euclidean norm. Utility maximization implies

$$y_{ij} = \begin{cases} 1 & \text{if } U_i(\boldsymbol{\zeta}_j) > U_i(\boldsymbol{\psi}_j), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The specification is completed by assigning a distribution to the errors. We assume that the errors η_{ij} and ν_{ij} have a joint normal distribution with $E(\eta_{ij}) = E(\nu_{ij})$, $\text{var}(\eta_{ij} - \nu_{ij}) = \sigma_j^2$ and the errors are independent across both legislators and rollcalls. It follows that

$$\begin{aligned} P(y_{ij} = 1) &= P\left(U_i(\boldsymbol{\zeta}_j) > U_i(\boldsymbol{\psi}_j)\right) = P\left(\nu_{ij} - \eta_{ij} < ||\mathbf{x}_i - \boldsymbol{\psi}_j||^2 - ||\mathbf{x}_i - \boldsymbol{\zeta}_j||^2\right) \\ &= P\left(\nu_{ij} - \eta_{ij} < 2(\boldsymbol{\zeta}_j - \boldsymbol{\psi}_j)'\mathbf{x}_i + \boldsymbol{\psi}_j'\boldsymbol{\psi}_j - \boldsymbol{\zeta}_j'\boldsymbol{\zeta}_j\right) = \Phi(\boldsymbol{\beta}_j'\mathbf{x}_i - \alpha_j) \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}_j = 2(\boldsymbol{\zeta}_j - \boldsymbol{\psi}_j)/\sigma_j$, $\alpha_j = (\boldsymbol{\zeta}_j'\boldsymbol{\zeta}_j - \boldsymbol{\psi}_j'\boldsymbol{\psi}_j)/\sigma_j$, and $\Phi(\cdot)$ denotes the standard normal distribution function. This corresponds to a probit model[2] with an unobserved regressor $\mathbf{x}_i$ corresponding to the legislator's ideal point. The coefficient vector $\boldsymbol{\beta}_j$ is the direction of the jth proposal in the policy space relative to the "Nay" position.

[2] A logit model results if the errors have extreme value distributions.

Given the assumptions of independence across legislators and roll calls, the likelihood is

$$L(\mathbf{B}, \alpha, \mathbf{X}|\mathbf{Y}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \Phi \left(\mathbf{x}_i' \boldsymbol{\beta}_j - \alpha_j\right)^{y_{ij}} \left(1 - \Phi \left(\mathbf{x}_i' \boldsymbol{\beta}_j - \alpha_j\right)\right)^{1-y_{ij}} \tag{3}$$

where $\mathbf{B}$ is an $m \times d$ matrix with jth row $\boldsymbol{\beta}_j'$, $\alpha = (\alpha_1, \ldots, \alpha_m)'$, $\mathbf{X}$ is an $n \times d$ matrix with ith row $\mathbf{x}_i'$, and $\mathbf{Y}$ is the $n \times m$ matrix of observed votes with (i, j)th element y_{ij}.

The model, as described above, is the simplest possible form and is a convenient starting point for more elaborate models. We show later how it is possible to add party effects to this specification. Clinton and Mierowitz (2001) modify this framework to study agenda dependence. It is also possible to incorporate vote trading and cue-taking into the model by making the utility of one legislator dependent upon either the utility or voting behavior of another.

The spatial voting model is equivalent to the two-parameter item response model used in educational testing,[3] where β_j is the item discrimination parameter, α_j is the item-difficulty parameter, but in the roll call context the latent trait or "ability" parameter x_i is the ideal point of the i-th legislator. There is a large literature in psychometrics on estimation of these models (e.g., Bock and Aitken 1981; Baker 1992), but the focus is usually on estimation of the β_j (the item parameters), which are used for test equating. In roll call analysis, however, primary interest almost always centers on the $\mathbf{x}_i$ (the ideal points), while in psychometrics the $\mathbf{x}_i$ (ability parameters) are usually treated as random effects.

2.1. Identification

As it stands, the model (3) is not identified. For example, suppose we rotate the matrix of ideal points $\mathbf{X}$ by premultiplying by an orthogonal $d \times d$ matrix $\mathbf{R}$ (with $\mathbf{R}'\mathbf{R} = \mathbf{I}_d$) and apply the same transformation to the direction vectors $\mathbf{B}$, $\mathbf{X}^* = \mathbf{RX}$ and $\mathbf{B}^* = \mathbf{R}^{-1}\mathbf{B}$. Then the likelihood $L(\mathbf{B}, \alpha, \mathbf{X}|\mathbf{Y}) = L(\mathbf{B}^*, \alpha, \mathbf{X}^*|\mathbf{Y})$ for all possible voting patterns $\mathbf{Y}$: no data can distinguish between the different parameter values because any translation or rotation of the ideal points and proposals leaves the distances between ideal points and alternatives unchanged. This is not unexpected since otherwise the underlying utilities are not recoverable.

Identification is essential for standard methods of estimation, such as maximum likelihood, which are inconsistent when the model is unidentified. The role of identification in Bayesian estimation is more controversial. Bayesian procedures can be applied to unidentified models, though the data are only informative about identified parameters (e.g., Neath and Samaniego 1997). However, in many cases it is difficult to formulate a reasonable prior for problems involving arbitrary rescalings. For example, we may have some prior information about tomorrow's temperature, but it is very difficult to quantify this information unless we agree in advance whether temperature is measured on the Fahrenheit, Centigrade, or some other scale. This is a simple example of *normalization*. The same

[3] This equivalence has been noted by several authors, including Poole and Rosenthal (1997, 247), Bailey and Rivers (1997) and Londregan (2000a).

problem occurs in policy spaces since both the origin and metric are arbitrary.

Rivers (2003) derives necessary and sufficient conditions for identification of multidimensional spatial models based upon *a priori* restrictions on the ideal point matrix **X**.[4] In the case of a unidimensional policy space, the identifying conditions are straightforward: two linearly independent restrictions on the ideal point matrix **X** are required. One possibility is to constrain the positions of two legislators at arbitrary positions, *e.g.*, Kennedy at -1 and Helms at +1. Alternatively, we can constrain the ideal points to have mean zero and standard deviation one. This is sufficient for local, but not global, identification (since the left-right direction can be reversed by reflecting the ideal points around the origin and reversing the policy directions).

In d dimensional choice spaces, $d(d+1)$ linearly independent *a priori* restrictions on the ideal points **X** are required for identification. Thus, in two dimensions, it is necessary to fix the positions of three legislators (three ideal points, each with two elements). In general, identification can be achieved by fixing the positions of $d+1$ legislators. Estimation becomes progressively more difficult in higher dimensions. In addition to the necessary identifying restrictions, it is also beneficial to add other *a priori* information. We provide an example using the interest group ratings in section 6.1 below.

3. Estimation and Inference

The classical or frequentist approach treats ideal points as fixed but unknown parameters. An estimation technique, such as maximum likelihood, is evaluated by considering its sampling distribution. We imagine the ideal points and other parameters to be fixed and draw repeated samples from the same data generating process. Each of these samples is a hypothetical roll call governed by the same ideal points and bill parameters. Because voting is probabilistic (see equation 2), each sample yields different votes and hence different estimates of the ideal points and other parameters. The sampling distribution of an estimated ideal point is its distribution across a set of hypothetical roll calls.

The Bayesian approach, in contrast, treats the unknown ideal points and other parameters as random variables and conditions upon the observed roll call data. We represent any *a priori* information by a *prior distribution* over the parameters. Bayes' formula describes how to combine the prior information with the observed data to obtain a *posterior distribution* which summarizes our information about the parameters having seen the roll call data. The Bayesian approach, as we will see, allows us to make probability statements, such as "Kennedy is more likely than O'Connor to be the median Justice on the Supreme Court." Of course, this kind of statement is meaningless from the frequentist perspective, which treats the ideal points as fixed.

Bayesian methods are often thought of primarily as a way to use non-sample information in estimation. (See Western and Jackman 1994 for some examples from political science.) Our

[4]Rivers also assumes two additional regularity conditions. The columns of **X** are linearly independent and do not contain the unit vector in their span; and the matrix of direction vectors **B** has full column rank.

motivation for using Bayesian methods in roll call analysis, however, is rather different. Roll call data sets are usually very large, so in most cases the benefit to introducing *a priori* information is slight. Instead, the real benefit to the Bayesian approach is that it turns a very difficult classical estimation problem into a fairly routine application of Markov chain Monte Carlo (MCMC) simulation. Moreover, the Bayesian approach lets us make inferences about ideal points and substantive hypotheses that were intractable with classical techniques. And in addition, the Bayesian model and estimation procedures are easily extended to handle more complex formulations.

To understand better the computational challenge that roll call analysis presents for classical estimation, consider the number of parameters that need to be estimated in some typical applications. With data from n legislators voting on m roll calls, a d-dimensional spatial voting model gives rise to a statistical model with $p = nd + m(d + 1)$ parameters. Table 1 presents values of p for five different data sets. A moderately sized roll call data set (say the 105th U.S. Senate) with n=100, m=534 non-unanimous roll calls and d = 1 yields p = 1,168 unknown parameters, while a two dimensional model has p = 1,802 parameters. A typical House of Representatives (e.g., the 93rd House) set has n = 442 and m = 917, and so a one dimensional model has p = 2,276 parameters, while a two dimensional model has p = 3,635 parameters. Pooling across years dramatically increases the number of parameters: for instance, Poole and Rosenthal (1997) report that fitting a two dimensional model to roughly two hundred years of U.S. House of Representatives roll call data gave rise to an optimization problem with $p > 150,000$ parameters.

	Legislators	Roll Calls	Dimensions (d)		
Legislature	n	m	1	2	3
U.S. Supreme Court, 1994-97	9	213	435	657	879
105th U.S. Senate	100	534	1,168	1,802	2,436
93rd U.S. House	442	917	2,276	3,635	4,994
U.S. Senate, 1789-1985	1,714	37,281	76,276	115,271	154,266
U.S. House, 1789-1985	9,759	32,953	75,485	118,017	160,549

TABLE 1. Number of Parameters in Roll Call Analyses

The proliferation of parameters causes several problems. The usual optimality properties of conventional estimators, such as maximum likelihood, may not hold when, as in this case, the number of parameters is a function of the sample size (see Lancaster 2000, for a recent survey). In particular, the customary asymptotic standard error calculations, using the inverse of the information matrix, are not valid. As a practical matter, the size of the information matrix is too large for direct inversion. Poole and Rosenthal (1997, 246) take the obvious shortcut of fixing the bill parameters at their estimated values before calculating standard errors for the ideal point estimates. They point out that this is invalid, but it reduces the computational burden by an order of magnitude.

The Bayesian methods of estimation and inference proposed here are valid for finite samples and do not employ any large sample approximations. The number of parameters is fixed for any particular estimation problem by the *actual* number of legislators and roll calls and Bayes' Theorem gives the exact posterior distribution of the parameters conditional upon the observed data. The only approximation that is required involves the simulation of the posterior distribution and this approximation can be made to any desired degree of accuracy by increasing the number of simulations (*not* the sample size).

Details of the estimation procedure are provided in an appendix, but a brief heuristic explanation may be useful. The fundamental difficulty in roll call analysis is that everything other than the votes is unobservable: the ideal points, bill parameters, and utilities are unknowns. But if it were possible to impute values to the bill parameters and utilities, then the ideal points could be estimated by regression. By the same logic, if we were able to impute values for the ideal points and utilities, the bill parameters could also be estimated by regression. The MCMC algorithm repeatedly performs such imputations starting from an arbitrary point and alternating between simulation of the ideal points, bill parameters, and utilities, and is described in full in the Appendix. Under a wide set of conditions (e.g., Tierney 1996) MCMC algorithms are guaranteed to generate samples from the posterior density of the model parameters, regardless of where in the parameter space the algorithm is initialized.

We use intentionally vague priors for most of the parameters. For each application below, we describe the actual prior used, but, except where noted, the results are insensitive to choice of prior.

4. Comparison with Other Methods of Ideal Point Estimation

Having detailed our approach, Table 2 provides a summary of the differences between our approach, NOMINATE, and the factor-analytic approach of Heckman and Snyder. Our approach has more in common with NOMINATE than the Heckman-Snyder factor analysis approach. The Heckman-Snyder factor-analytic approach is distinctive in that the statistical model does not follow as neatly from a formal model of legislative voting[5], but provides ideal point estimates relatively cheaply; indeed, factor analysis supplies starting values for both the NOMINATE algorithms and our Bayesian simulation approach.

4.1. Example 1: 106th U.S. House of Representatives

To illustrate the differences and similarities between existing approaches and our simulation-based Bayesian estimator, we first analyze roll calls from the 106th U.S. House of Representatives via several methods. We fit a one dimensional model to these data using principal components (extremely similar to the Heckman-Snyder estimator)[6], W-NOMINATE, as well as our Bayesian

[5]The difficulty is that the Heckman-Snyder factor analytic statistical model results from a linear probability model for roll call voting, in turn driven by the assumption that the stochastic component of the legislators utility differential (the net utility a legislator has for voting "Yea" over voting "Nay") follows a uniform distribution; alas, rationalizing this statistical model requires some *ad hoc* assumptions as to legislator's utilities, since there does not exist a distribution such that the difference of two independent realizations from it yield a quantity with a uniform distribution.

[6]We implement the principal components estimator as follows: (1) compute a n-by-m matrix $\mathbf{D}$ by double-centering the roll call matrix (subtracting out column and row means), (2) compute $\mathbf{R}$, a n-by-n correlation matrix (the correlation

CJR	NOMINATE	Heckman-Snyder
Legislators' utilities, deterministic component:		
quadratic	Gaussian, with fixed scale	unknown
Legislators' utilities, stochastic component:		
normal (yielding a probit model) or Type 1 extreme value (yielding logit)	Type 1 extreme value (logit)	difference of utilities for "Yea" and "Nay" alternatives has a stochastic component with $U[0,1]$ distribution, yielding a linear probability model
Identification for One-Dimensional Case:		
fix two legislators' ideal points at -1 and 1	constrain legislators' ideal point to [-1,1]	identified only up to scale or reflection
Estimation:		
Markov chain Monte Carlo	alternating conditional maximum likelihood	eigen-decomposition (principal components factor analysis)
Uncertainty Assessments/Standard Errors:		
arbitrarily exact; can be approximated to any desired degree of accuracy via repeated sampling from joint posterior density of model parameters	approximate for ideal points, after conditioning on estimates for bill parameters	none

TABLE 2. Comparison of Ideal Point Estimation Methods

approach.[7] After discarding lop-sided votes, W-NOMINATE uses 871 roll calls, and does not fit an ideal point for Livingston (R-LA) who resigned from Congress in February 1999 after voting on nineteen roll calls in the 106th House. Lop-sided votes and short voting records pose no problems in the Bayesian approach; we estimate ideal points for all legislators and include all but unanimous roll calls, yielding m=1,073 roll calls in all, comprising 444,326 individual voting decisions. With predicted probabilities of 0.5 as a classification threshold we correctly classify 89.9% of the individual voting decisions,[8] and find that 1,007 of the 1,073 (93.8%) roll calls discriminate with respect to the single latent dimension.[9] Of the 66 roll calls that fail to discriminate with respect to the recovered dimension, only two roll calls were decided by margins closer than 60%-40%. In short, a one-dimensional model appears to be a very good characterization of these roll call data.

Figure 1 plots the three sets of ideal point estimates against one another. This figure exemplifies a pattern we have seen in many other roll call data sets: when n and m are both reasonably large and a low dimensional model fits the data well, there is extremely little difference in the ideal point estimates produced by W-NOMINATE and our Bayesian estimator. In this specific example, $n = 440$, $m = 1,073$ and a one dimensional model gives an extremely good fit to the data (as is typical of recent U.S. Congresses), and the ideal point estimates correlate at .996. Nonetheless, by retaining more of the lop-sided votes than NOMINATE, our Bayesian estimator can discriminate among extremist Democrat legislators (in the left tail), effectively "stretching" the distribution of the Democrats ideal points relative to NOMINATE's estimates. The comparison of both NOMINATE and our Bayesian estimator with the principal components estimator reveals the linearity of the factor analytic model, with the two non-linear models both generating more discrimination among extremist legislators.

4.2. Example 2: U.S. Supreme Court, 1994-97

But now consider a much smaller roll call data set: $m = 213$ non-unanimous decisions of the seventh "natural" Rehnquist court (n=9) consisting of Justices Rehnquist, Stevens, O'Connor, Scalia, Kennedy, Souter, Thomas Ginsberg and Breyer appearing in the Spaeth (2001) data set. The decisions of the justices are coded as y_{ij}=1 if justice i joins the majority on case j, and y_{ij}=0 if he or she dissents; there are ten abstentions in the data.

Table 3 compares the results we obtain fitting a unidimensional model via Bayesian simulation and W-NOMINATE. Our Bayesian estimates (posterior means and standard deviations) are reported in the first column. In this instance we constrain the justices' ideal points to have mean zero and

matrix of $\mathbf{D}'$, using pairwise deletion of missing data), and (3) take the first d eigenvectors of $\mathbf{R}$ as the ideal point estimates for a d-dimensional model.

[7] We use the probit version of our model, and constrain the legislators' ideal points to have mean zero and variance one across legislators.

[8] This classification rate is a function of the unknown model parameters, and so is itself subject to uncertainty; here we report the classification rate averaging over uncertainty in the model parameters. See the discussion of auxiliary quantities of interest in section 5.

[9] That is, these 1,007 roll calls all had slope coefficients (β_j, the equivalent of item-discrimination parameters) whose 90% posterior confidence intervals did not cover zero.

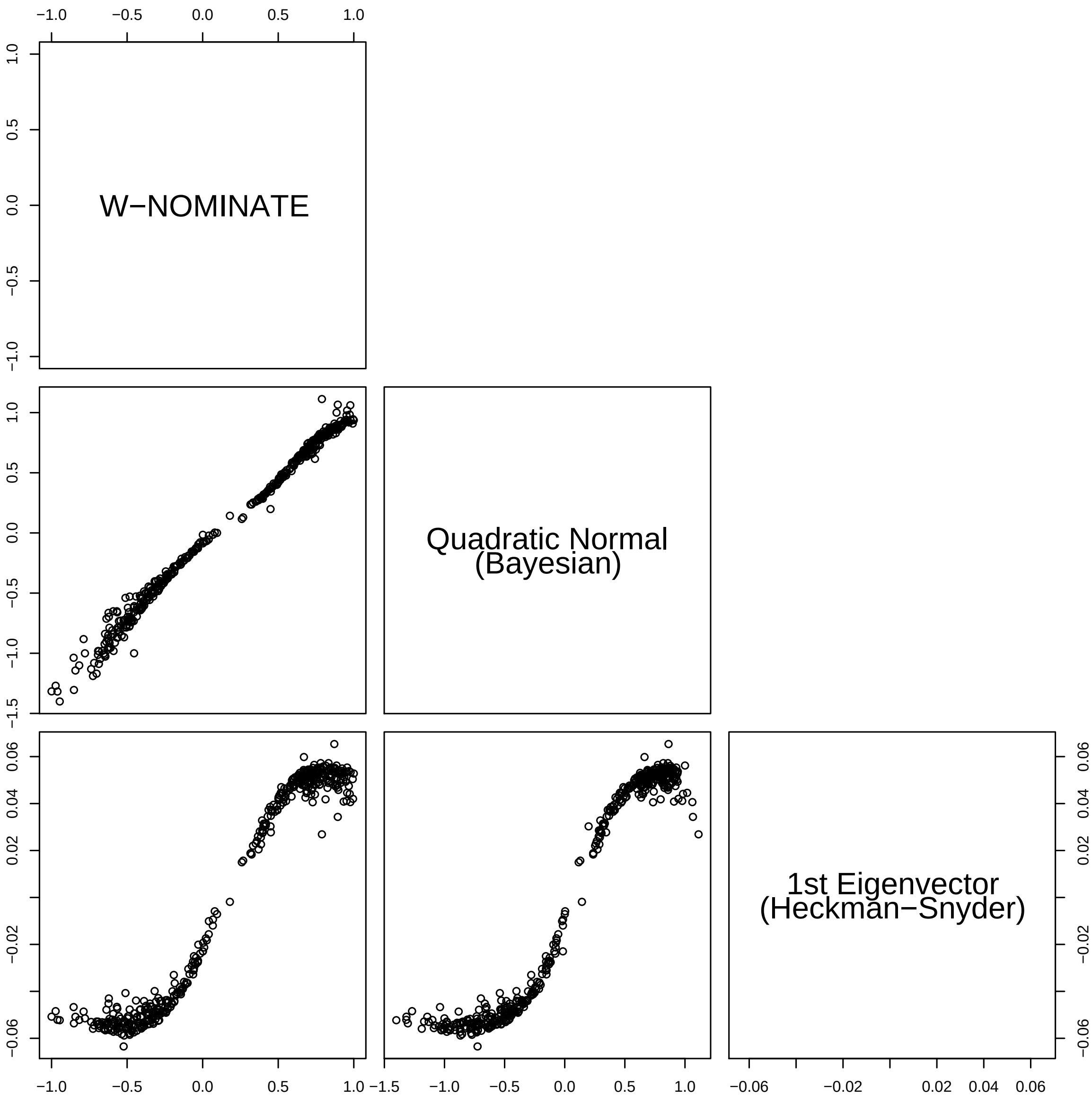

FIGURE 1. Comparison of W-NOMINATE, Bayesian and principal components factor analysis estimates of ideal points, one-dimensional model, 106th U.S. House.

unit variance across justices.[10] Of the 213 cases analyzed, 179 (84%) discriminate with respect to the recovered dimension,[11] and using a predicted probability of 0.5 as a classification threshold we correctly predict 86.6% of the 1,907 individual decisions. On the other hand, W-NOMINATE places all of the ideal point estimates at its boundary constraints of -1 and 1, and reports pseudo standard errors of zero, indicative of the difficulties W-NOMINATE may run into when working with small data sets.[12]

	Bayesian	W-NOMINATE
Rehnquist	.47	1.00
	(.046)	(.00)
Stevens	-2.40	-1.00
	(.065)	(.00)
O'Connor	.33	1.00
	(.043)	(.00)
Scalia	.83	1.00
	(.074)	(.00)
Kennedy	.30	1.00
	(.042)	(.00)
Souter	-.09	-1.00
	(.057)	(.00)
Thomas	.99	1.00
	(.11)	(.00)
Ginsberg	-.21	-1.00
	(.064)	(.00)
Breyer	-.22	-1.00
	(.067)	(.00)

Table 3. Ideal Point Estimates, U.S. Supreme Court, 1994-97. The Bayesian estimates are estimated posterior means. Quantities in parentheses are posterior standard deviations for Bayesian estimates, implied standard errors for W-NOMINATE. See text for details of estimation.

The Bayesian estimates largely accord with common wisdom about the Court: Stevens is unambiguously the most liberal justice, with Kennedy and O'Connor occupying the middle ground of the court, and Scalia and Thomas the two most conservative justices. Since our Bayesian approach recovers the joint posterior density of all parameters, we can test hypotheses of the sort "justice *a* is to the left/right of justice *b*"; hypotheses of this type require working with the joint density of

[10] The estimates in the table are based on five million iterations of the MCMC algorithm, discarding the first 100,000 iterations as "burn-in", and then thinning the MCMC output by a factor of 500. Further details appear in the Appendix.

[11] Again, in the sense that the 90% confidence interval for the discrimination parameter β_j does not overlap zero.

[12] Keith Poole provided us with a special version of W-NOMINATE for working with "small-*n*" roll call matrices, such as these court data. The ideal point estimates produced by this version of W-NOMINATE are more sensible.

ideal points x_a and x_b, which we obtain from our Bayesian simulation algorithm.[13] For instance, the probability that Justice O'Connor is to the right of Justice Kennedy is .83, while the probability that Justice Thomas is to the right of Justice Scalia is .95; i.e., the preponderance of the evidence suggests that Thomas is more conservative than Scalia, and O'Connor more conservative than Kennedy. Only for the Scalia-Thomas comparison would we reject the null hypothesis that the respective pairs of justices had the same ideal points at conventional 95% levels of statistical significance.

5. Estimation and Inference for Auxiliary Quantities of Interest

An advantage of the Bayesian approach is that it is straightforward to estimate posterior distributions over any auxiliary quantity of interest that is a function of the model parameters. These quantities of interest can be any function of the model parameters, as we now demonstrate.

5.1. Example 3: Pivotal Senators in the 106th U.S. Senate

The notion of pivotal legislators is critical to many theories of legislative behavior. For instance, super-majorities are often needed for extraordinary legislative action, such as the two-thirds majority required to override a presidential veto, or the 60 votes needed to pass cloture motions in the U.S. Senate. Order statistics of ideal points play a prominent role in theories of legislative politics: e.g., the "gridlock interval" is defined as the region between the filibuster pivot and the veto pivot, and in the case of a liberal president the gridlock interval is bounded on the left by the veto pivot (the 33rd senator) and on the right by the filibuster pivot (the 60th senator) (e.g., see Krehbiel 1998, Figure 2.2). Formal theories of legislative politics make sharp and exact predictions on the basis of these pivotal locations: e.g., proposals that attempt to change status quo policies located in the gridlock interval will not succeed. To operationalize these theoretical predictions, the gridlock interval is usually computed using the estimated ideal points of the corresponding legislators (e.g., Howell et al. 2000). Similarly, Schickler (2000) characterizes the parties' ideological positions with the ideal points of the median legislator within each party. Given the importance of individual legislators such as the chamber median or the "filibuster pivot" (i.e., the 40th Senator), it is straightforward to generate posterior estimates for both the identity and the spatial location for such legislators in our Bayesian simulation approach.

Consider the task of uncovering the identity of the "pivotal" Senators in the 106th Senate (n=102, m=596 non-unanimous roll calls; there are 102 Senators because of the replacement of John Chafee (RI) by his son Lincoln Chafee and the replacement of Paul Coverdell (GA) by Zell Miller). To determine which Senators are critical for invoking cloture or which are the median Senators requires recovering the posterior distribution of the rank of each Senator's ideal point. We compute this by repeating the following scheme an arbitrarily large number of times: (1) sample the legislators' ideal points x_i

[13] To compute the probability the $x_a > x_b$ we simply note the proportion of times we observe that event in many draws from the joint posterior density of the ideal points. Note that while W-NOMINATE reports an estimated pseudo standard error for each estimated ideal point, it does not report the full variance-covariance matrix of the ideal points estimates (with covariances in the off-diagonal elements). This means when using NOMINATE for pairwise comparisons of ideal points, the ideal point estimates are implicitly assumed independent (zero covariances), which in general is not true.

from their joint posterior density; (2) rank order the sampled ideal points; (3) note which legislator's occupies a particular pivot or order statistic of interest. We then report the *proportion* of times the *i*-th legislator's ideal point is the pivot or order statistic of interest over these repeated samples from the posterior of the ideal points. Since we are working with an arbitrarily accurate approximation of the joint posterior density for the ideal points, inferences as to the ranks and the identity of the legislators occupying particular ranks are also arbitrarily accurate.[14]

Figure 2 summarizes the posterior densities over the *identities* of the senators at key pivot points. We omit the "replacement senators" Chafee and Miller from these calculations. There is almost no doubt as to identity of the chamber median: Republican Senators Snowe and Collins are the only Senators with positive probability[15] of being the 50-th Senator, with Collins overwhelmingly most likely to be the 50-th (p = .98). Twenty-two Senators have discernible probability of being the veto pivot (the 33rd Senator), with 10 Senators having probabilities greater than 5% of being the veto pivot: Senators Baucus (p = .12), Biden (p = .11), Johnson (p = .08), Graham (p = .08), Bayh (p = .08) and Cleland (p = .08) account for roughly half of the uncertainty as to the identity of the veto pivot, but clearly no one Senator is unambiguously the veto pivot. Thirteen Republican Senators have positive probability of being the filibuster pivot (the 60th Senator), five of whom have $p > .05$, with Stevens the most likely candidate for the filibuster pivot (p = .41), followed by Warner (p = .26) and Campbell (p = .09).

A similar computation can be performed to recover the estimated spatial *location* of pivotal legislators. Figure 3 shows the location of the median, veto pivot and filibuster pivot (with confidence intervals), along with the ideal point estimate of adjacent senators. Again, it is apparent that there is little uncertainty as to the median (Senator Collins). But an interesting result arises for the veto pivot: although we are unsure as to the *identity* of the veto pivot, we are quite sure as to the veto pivot's *location*. A similar result is also apparent for the filibuster pivot, comparing Figures 2 and 3. While we may be able to pin down the location of either pivot with some precision, we don't know which legislators will be the veto and filibuster pivots will be on any given vote. This is a seldom noticed feature of contemporary U.S. congresses, but one with implications for lobbying and legislative strategy: i.e., relatively precise knowledge of where the pivots lie does not correspond to knowing the identity of pivotal legislators, whose votes are necessary to guarantee cloture or veto-proof majorities.

5.2. Example 2 (continued): Who is the Median Justice?

Pivotal legislators can also be defined for deliberative bodies such as courts. For instance, tremendous interest centers on the identity and location of the *median justice*, who is pivotal in

[14] In principle, one could implement a similar procedure with W-NOMINATE estimates, two complications arise: (1) all covariances among ideal points are implicitly set to zero since W-NOMINATE only reports pointwise standard errors on each ideal point and (2) an asymptotically-valid normal approximation is assumed to characterize ideal point uncertainty (less pressing when working with large roll call data sets).

[15] In the sense that over many repeated samples from the posterior density for the model parameters, at least in one sample we observe the sampled value of the respective Senator's ideal point being ranked 50-th.

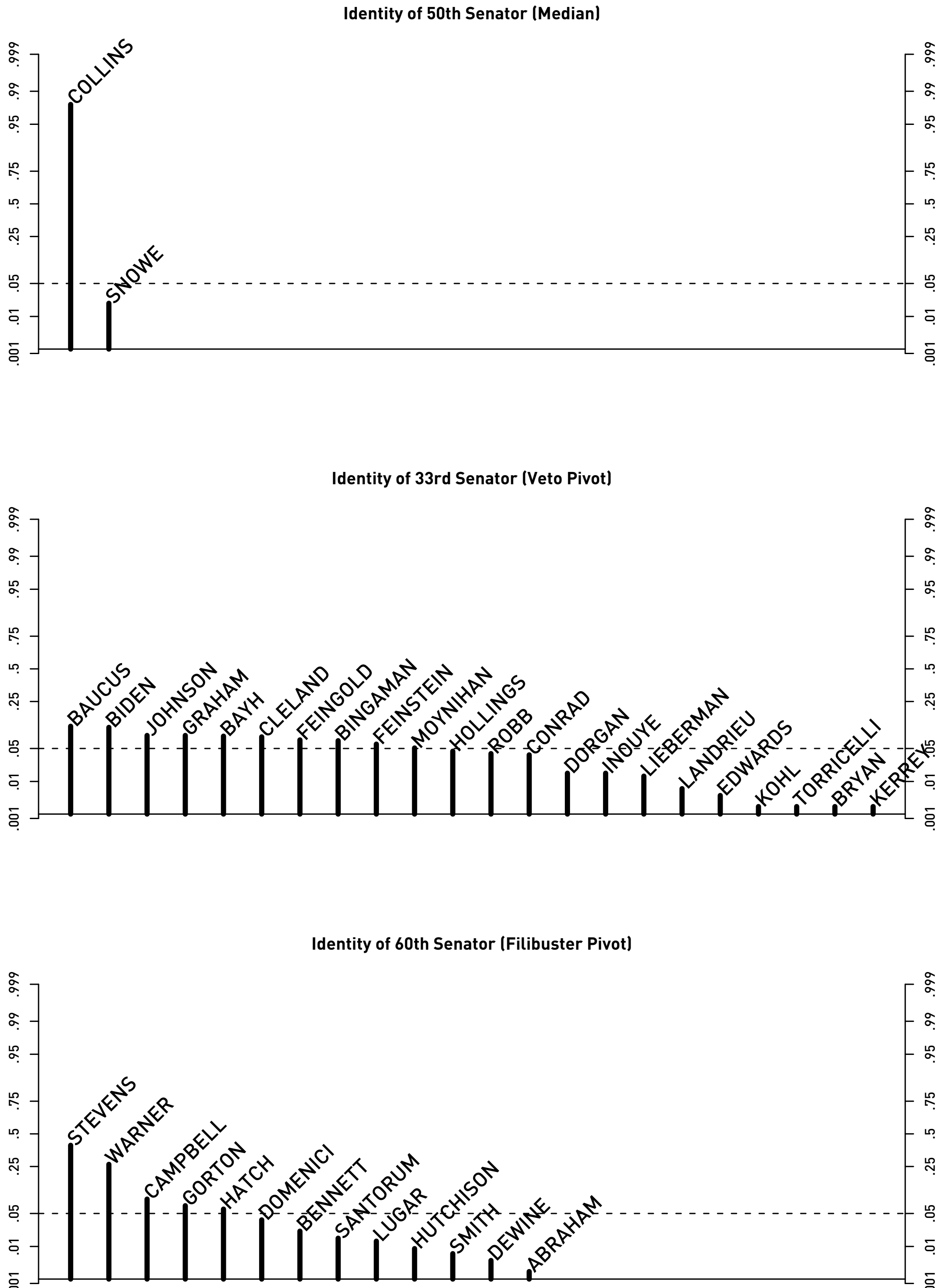

FIGURE 2. Posterior Densities over Identity of Pivots, 106th U.S. Senate. Each bar indicates the probability that the indicated Senator occupies the particular rank, for the the 100 Senators that originally comprised the 106th U.S. Senate (i.e., omitting Lincoln Chafee and Zell Miller). Vertical axis is scaled non-linearly, via the probit transformation, so as to highlight differences among small probabilities.

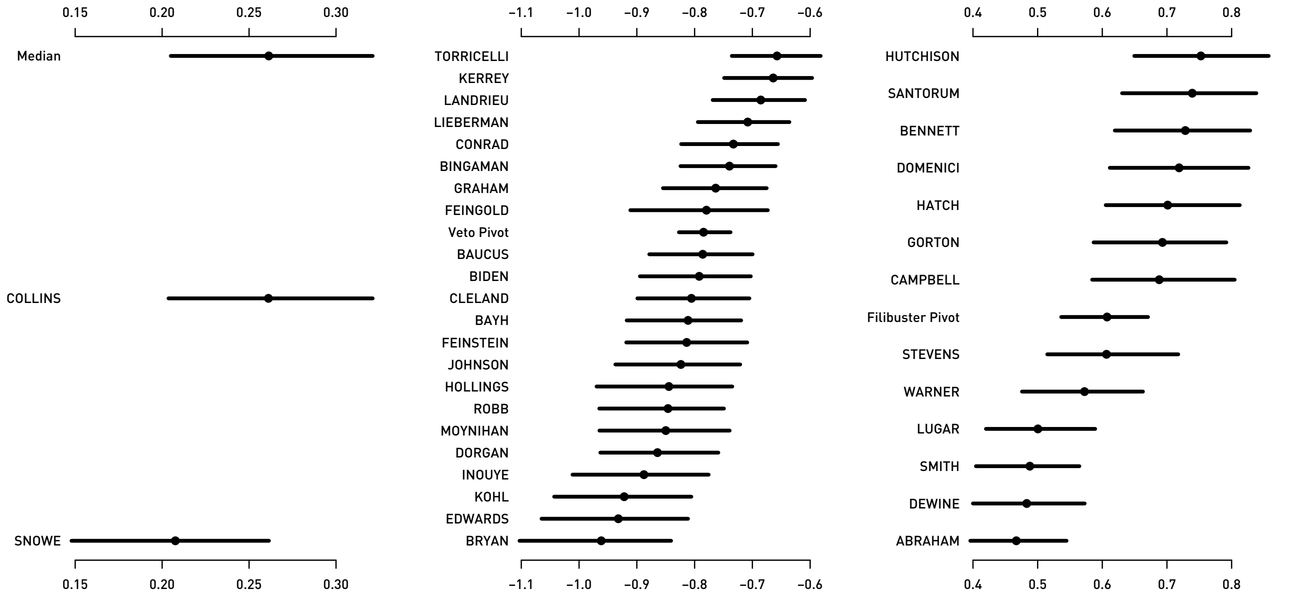

FIGURE 3. Location of Pivots, and Adjacent Senators, 106th U.S. Senate. Points are posterior means, and the horizontal lines cover 95% confidence intervals.

cases that divide the court. Using the estimates reported in section 4.2 we find only two justices are assigned positive probability of being the median justice: Kennedy (p = .82) and O'Connor (p = .18).

6. Extending the Spatial Voting Model

A major advantage of the Bayesian simulation approach lies in its extensibility: additional substantive information about roll calls and legislators can be brought to bear on analyses of roll call data. This feature of our approach is especially helpful given that the roll call data is merely a matrix of "Yeas" and "Nays", possessing no information as to the sequence and substantive content of votes, characteristics of legislators and their constituencies, and the substantive content of the latent policy dimensions. This information is critical in scholarly accounts of legislative behavior, but is absent in almost all existing models and analyses purporting to estimate ideal points.[16] Our estimator is readily extended to incorporate this information, effectively integrating the *measurement* of legislative preferences with the *analysis* of legislative behavior.

6.1. Example 4: Prior Beliefs over Ideological Dimensions

The number of dimensions and their substantive content are critical considerations in studying legislative politics. Most existing methods for analyzing roll call data are agnostic on these questions; the substantive content of the policy dimensions are deduced *ex post* by inspecting scatterplots of the ideal point estimates and/or overlaying cutting planes for specific votes. But

[16]Some notable exceptions include Londregan's analyses of small n committees (Londregan 2000b) and the Constitutional Convention, where the identity of the legislator making the j-th proposal is incorporated into the analysis.

students of congressional politics often have beliefs or conjectures as to the substantive nature of the dimensions thought to underlie roll call data. For instance, certain roll calls are known or strongly suspected to tap a particular dimension more so than other dimensions; scholars of particular periods in American political history are able to point to specific roll calls on, say, slavery, civil rights, or trade liberalization, that cut across a traditional economic "left-right" dimension. In a similar vein, Supreme Court decisions are explicitly classified into substantive categories (e.g., Spaeth 2001).

One source of plausible information that can be used to define the estimated dimensions *a priori* are the "key votes" interest groups identify as being definitive on the issue(s) they care most about. As an illustration, we estimate a four dimensional model for the 106th U.S. House, with reference roll calls drawn from the key votes lists of (a) the AFL-CIO, (b) Americans for Democratic Action, (c) the Sierra Club; and (d) the Chamber of Commerce. Each group's list of key votes serve as our reference roll calls for a specific dimension. In other words, using the 91 votes as reference roll calls permits us to estimate a multi-dimensional model in which it is possible to examine how legislators' revealed preferences in the "Sierra Club" dimension relate to those in the "Chamber of Commerce" dimension.[17]

Incorporating this type of prior information is a natural part of the Bayesian approach. To exploit the substantive content of these key roll calls, we specify informative priors on the relevant discrimination parameters: if bill j is a reference bill for dimension k, we use the priors $\beta_{jk} \sim N(\pm 4, .5^2)$, with the sign of the prior mean depending on the polarity of the reference bill[18] and $\beta_{jk'} \sim N(0, .025^2)\, \forall\, k' \neq k$, ensuring that bill j supplies substantive content to dimension k but not to other dimensions. For all non-reference bills we use the diffuse prior $\beta_{jk} \sim N(0, 10^2)\, \forall\, k$.

Figure 4 presents point estimates (posterior means) for the ideal points as a matrix of dimension-against-dimension scatterplots. The most striking feature of these results are the high correlations across dimensions. It appears that the there is considerable (if not complete!) redundancy in the interest groups "dimensions": i.e., a legislator's position on one dimension is an extremely good predictor of their position on the other dimensions, so much so that we could collapse these four dimensions onto one dimension, without any appreciable loss in fit to the data. However, using substantive information to identify the dimensions permits scholars to interpret the the measure of legislator preferences in terms of known political characteristics such as labor and environmental policies. The ability to define the recovered dimensions may be particularly important when scholars face political environments that are less structured than the contemporary Congress (e.g., Clinton and Meirowitz 2003).

[17]The fact that 14 roll calls are identified as key by more than one interest group suggests that the preferences of these groups are not mutually orthogonal. Consequently, we can expect high correlations among the ideal points on the recovered dimensions.

[18]If a "Yea" ("Nay") vote is consistent with the group's position then we assign a positive (negative) prior.

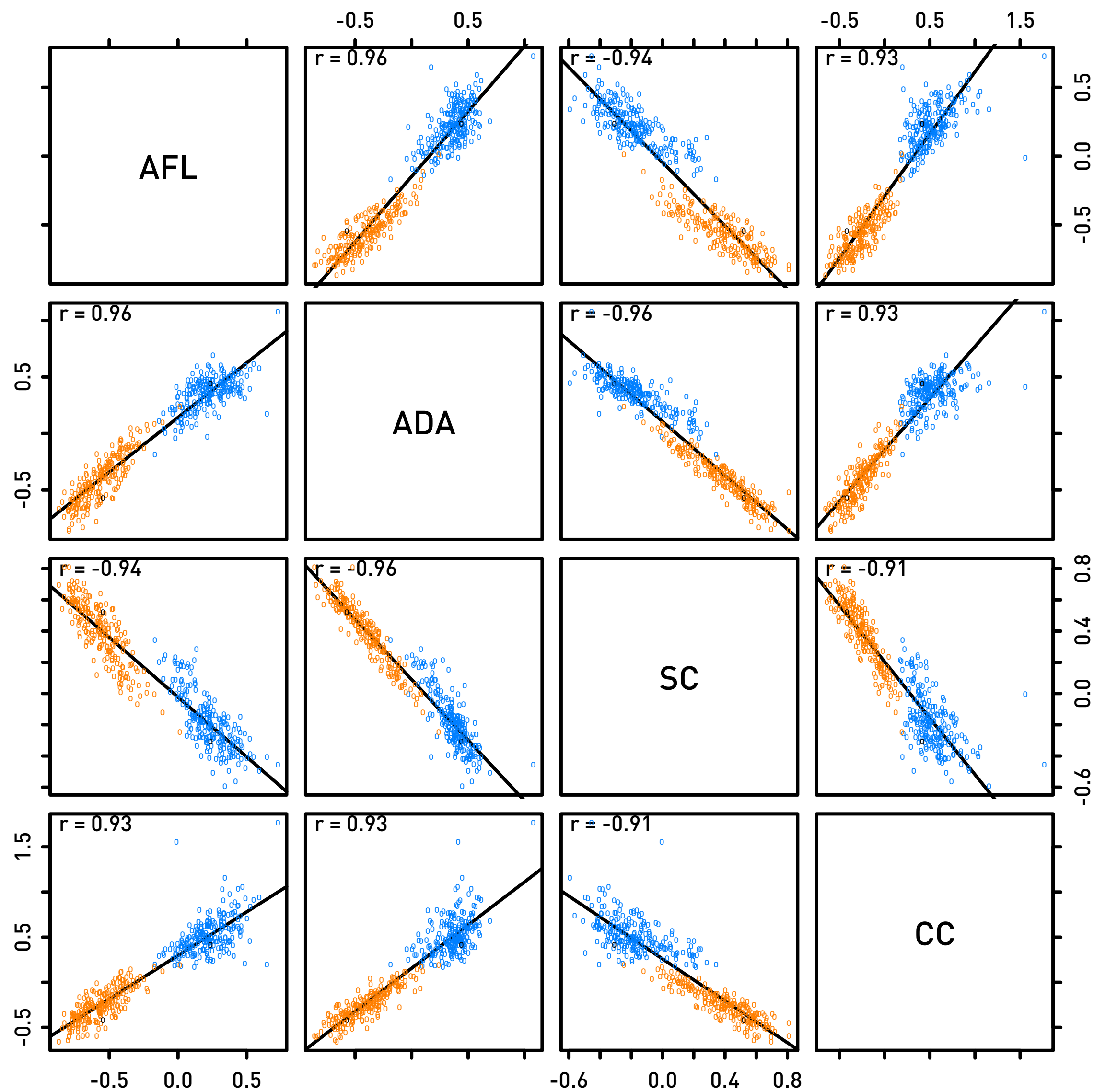

FIGURE 4. Scatterplot Matrix for Ideal Point Estimates (Posterior Means), Four-Dimensional Interest-Group Model, 106th U.S. House. Lighter points indicate Democrats (generally to the left); darker points indicate Republicans (generally to the right). The solid lines indicate linear regression fits. AFL = AFL-CIO, ADA = Americans for Democratic Action, SC = Sierra Club, CC = Chamber of Commerce.

6.2. Example 5: Party Switchers and the "Party Influence" Hypothesis

A major advantage of our modeling approach is the ability to extend the model to encompass alternative models of legislative behavior. For instance, thus far we assumed a Euclidean spatial voting model, in which conditional on legislators' unobserved ideal points (which are considered constant over the period spanned by the roll call data), voting is independent across legislators and roll calls. In the next set of examples we consider alternatives to this basic setup, all of which are easily accommodated in our approach.

A question of particular interest to congressional scholars is the influence of political parties on the voting records of legislators. Party switchers -- that is, legislators who change parties between elections, while continuing to represent the same geographic constituency -- provide something akin to a natural experiment: the switcher's change in party affiliation helps identify a "party effect", since many other determinants of roll call voting remain constant (e.g., characteristics of the legislators' constituency). The typical investigation of party switchers (e.g., Nokken 2000) uses a "pre/post" or "differences-in-differences" design, comparing changes in ideal point estimates for the party switcher relative to the changes among the switcher's fellow legislators, or a random selection of non-switchers (McCarty, Poole and Rosenthal 2001).[19] By definition, splitting the roll call data into "pre" and "post" switching periods gives us less data than in the entire legislative session, and, as a consequence, ideal point estimates based on the "pre" and "post" sets of roll calls will be less precise than those based on all roll calls. Any comparison of change in the ideal points ought to properly acknowledge the relative imprecision arising from the smaller sets of roll calls available for analysis. A strength of our Bayesian simulation approach is that we routinely obtain uncertainty assessment for all model parameters, and all inferences will reflect the drop in precision occasioned by slicing the roll call matrix around the party switch.

The flexibility of our approach lets us formally embed a model for change in ideal points in a statistical model for roll call data, greatly facilitating investigation of the party switching hypothesis. Let $s \in \{1, \ldots, n\}$ designate the party-switcher and x_{i1} and x_{i0} be the ideal points of legislator i in the post-switch and pre-switch periods, respectively. Then a weak form of the party-switcher hypothesis is that $\delta_s \equiv x_{s1} - x_{s0} \neq 0$ (i.e., the party-switcher's ideal point changes, and, presumably, in a direction consistent with the change in parties). A strict form of the party-switcher hypothesis involves the $n - 1$ additional restrictions $\delta_i \equiv x_{i1} - x_{i0} = 0, \forall\ i \neq s$ (i.e., the party-switcher is the only legislator whose pre-switch and post-switch ideal points differ). An intermediate version of the switching hypothesis maintains that legislative ideal points may "drift", but that the party-switcher's δ_s is non-zero and larger than non-switchers' δ_i. In any event, we require estimates of each legislators' δ, either by running two separate analyses (splitting the roll call data around the time of the party switch), or a more direct (but equivalent) approach in which we parameterize the post-switch ideal points as

[19] An obvious threat to this approach is self-selection into the "treatment" of party-switching, say, if party switching is motivating by change in the personal preferences of the legislator; in this case we could not distinguish any party effect from an effect due to the personal preferences of the legislator, but we do not pursue this issue here. Other analyses of party switchers have also noted this problem: see McCarty, Poole and Rosenthal (2001, 686).

$x_{i0} + \delta_i, i = 1, \ldots, n$.

Since the pre and post switch roll calls do not refer to identical proposal and status quo positions, some normalization is required to compare the resulting ideal point estimates. Our solution is to focus on relative changes, since without further assumptions, any "global" or "uniform" shift in the legislative agenda or preferences around the the party switch is unidentified. That is, the average post-switch legislative proposal may be more liberal than the average pre-switch proposal by distance ω, or (observationally equivalent), all legislators may move ω to the right (or more simply, there is no guarantee that the pre and post ideal point estimates are comparable).[20]

We illustrate our method with the 107th U.S. Senate. On May 24, 2001, Senator James M. Jeffords (VT) announced that he would leave the Republican party and become an independent. The switch was particularly consequential, catapulting the Democrats into the majority: the post-switch partisan composition was 50 Democrats, 49 Republicans and 1 Independent (Jeffords). One hundred and forty-eight non-unanimous votes were recorded in the 107th Senate prior to Jeffords' switch, and 349 were recorded post-switch.

Figure 5 depicts the joint distribution of pre-switch and post-switch ideal point estimates (top panel) and rank orderings (lower panel). We find evidence that the ideal points underlying Jeffords' voting behavior differ across the switch: the estimate of Jefford's post-switch ideal point is statistically distinguishable and more liberal than the pre-switch estimate, indicating that Jeffords shifted his policy position as well as his party.

Of course, there are several other senators who do *not* switch parties but whose ideal points do move markedly. Points above (below) the 45 degree lines in Figure 5 indicate senators moving in a conservative (liberal) direction. Robert Torricelli (D, NJ) moves about 23 places to the left after the Jeffords switch; we speculate that Torricelli's ethics problems perhaps induced him to adopt more liberal positions, perhaps so as to head off a primary challenge in the 2002 election.[21] Particularly noteworthy is the fact that the ex-Majority Leader (Trent Lott, R-TX) becomes relatively more extreme and the new Majority Leader (Tom Daschle, D-SD) becomes relatively more moderate after the change in partisan control of the Senate. In particular Lott jumps 20 places, from the 77th most conservative senator the 97th senator, when arrayed from left-to-right (liberal to conservative); further, the 95% confidence bounds on Lott's rank orderings are 68th to 87th pre-switch (i.e., Lott's ideal point as majority leader was indistinguishable from the Republican party median), but 95th to 100th post-switch, so this movement is statistically significant. On the other hand, Daschle changes from the 13th (95% bound: 2nd-21st) to the 42nd (35th-45th) senator, jumping from being unambiguously on one side of the Democratic senate median to the other. These results are based

[20] In fact, solutions to this scaling problem abound: one could try to find bills with identical "Yea" and "Nay" locations in both periods, or assert that particular legislators do not move across the party switch. The identifying restriction we adopt is to estimate subject to the constraint that the legislators' ideal points have mean zero and unit variance in each period and that we interpret the δ_i as relative changes.

[21] As it turned out, Torricelli was unopposed in the Democratic primary, but dropped out of the race five weeks before the general election.

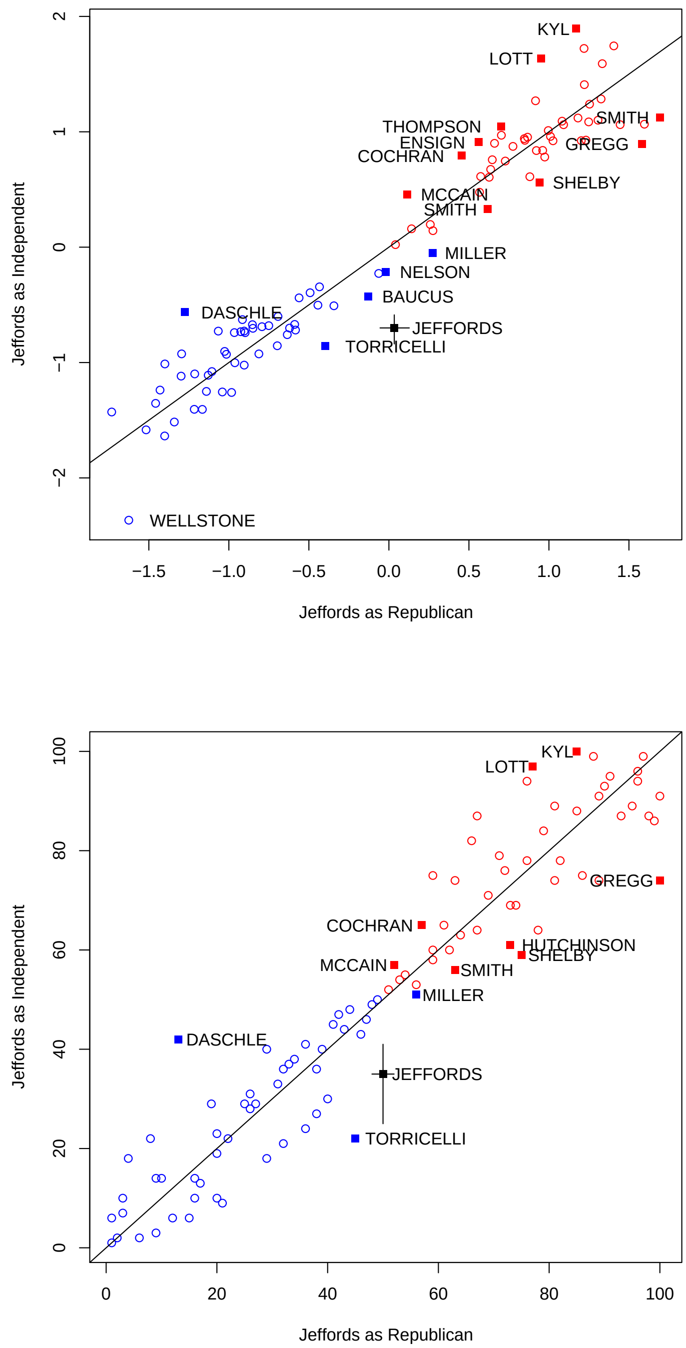

FIGURE 5. Comparison of Ideal Point Estimates, 107th U.S. Senate, pre and post Jeffords switch. Top panel compares relative spatial locations; lower panel compares rank order. The plotted points are posterior means. Squares indicate significant change. The diagonal line is a 45 degree line (i.e., if there was no relative change, then all the data points would lie on the line). The estimates for Jeffords are accompanied by 95% confidence intervals.

on a unidimensional fit to a single Congress, making us cautious about reaching for any broad conclusion; nonetheless, they are consistent with policy moderation by majority leaders, perhaps in order to more effectively represent and articulate party policy positions, or even so as to secure the majority leadership in the first instance.

This example models change in ideal points around a recent, vivid instance of party switching. But other models of change in ideal points can be easily fit into this framework. Party switchers perhaps provide the most direct evidence of party effects (although our results suggest that a misleading picture might be obtained if one were to to focus solely on the change of the switcher), but the methodology applied to examining relative change elsewhere (e.g., across congressional sessions, congresses, or politically consequential "shocks" in American political history).

6.3. Example 6: a two cutpoint, "party-influence" model

This last example again examines the question of party pressure in legislative behavior. The standard model assumes that conditional on a legislator's ideal point x_i and the vote parameters $\boldsymbol{\beta}_j$, vote y_{ij} is independent of $y_{i'j'}, \forall\, i \neq i', j \neq j'$: i.e., shocks making legislator i more likely to vote "Yea" do not make legislator i' any more or less likely to vote "Yea". Accordingly, the recovered ideal points need to interpreted with some caution. Party influence or "whipping" is one way that conditional independence assumption can be breached: e.g., legislators whose preferences might lead them to vote "Yea" are persuaded to vote "Nay", and vice-versa. Modeling roll call data without considering these possibilities leads to ideal point estimates that absorb a shock common to members of a given party (or, say, a whipped subset of a party's members); to the extent party influence is an unmodeled common shock, then the recovered x_i display greater partisan polarization than exists in the "true" x_i. Note that while party influence is a plausible mechanism for generating party-specific utility shocks, we are wary of inferring the presence of party pressure given evidence of party-specific utility shocks; we acknowledge that other mechanisms may generate party-specific utility shocks (e.g., lobbying by activists or interest groups that targets legislators from one party more than the other) and so we refer to "party-specific inducements" rather than "party-pressure."

It is reasonably straightforward to augment the standard model to allow for party-specific inducements. For instance, suppose that in addition to the standard quadratic spatial utilities, legislator i receives $\delta_j^{P(i)}$, a net incentive to vote "Yea" vis-a-vis "Nay" on vote j, but where the net incentive is specific to i's party affiliation $P(i) \in \{$"Democrat", "Republican"$\}$. If D and R are mutually exclusive and exhaustive party labels, then α_j and both $\delta_j^{P(i)}$ terms are unidentified: the problem is identical to trying to simultaneously estimate a grand mean and K group means (i.e., a design matrix of rank K but $K + 1$ columns). At best we can estimate a *net* difference in party-specific inducements, $\delta_j = \delta_j^D - \delta_j^R$ (see also Krehbiel 2003): i.e., we estimate $y_{ij}^* = x_i\beta_j - \alpha_j + \delta_j D_i + \varepsilon_{ij}$, where D_i is a binary indicator coded one if legislator i is a Democrat and zero otherwise, and for this example we assume the ε_{ij} are iid logistic (logit rather than probit).[22] Since the standard model nests as a special case

[22] McCarty, Poole and Rosenthal (2001) refer to this class of model as a "two-cutpoint" model, since it implies separate points where legislators of each party are indifferent between voting "Yea" and "Nay".

of the two cut-point model, it is straightforward to assess whether the restrictions implicit in the standard model are valid, by testing the joint null hypothesis $H_0 : \delta_j = 0, \forall j$. In addition, we also need to identify a set of votes in which the net party-specific inducements are not relevant (or can be reasonably assumed to be zero), since if *every* roll call was assumed to be potentially subject to party influence, then there is no way to compare the recovered ideal points of Democrats and Republicans.[23]

Several implementations of similar models appear in the literature. Snyder and Groseclose (2000) use a two-stage procedure: first, using lop-sided votes (those decided by more than 65/35 margins), estimate x_i via the standard model using the linear factor-analysis model due to Heckman and Snyder (1997); second, on non-lop-sided votes, estimate the linear regression of votes on x_i and a dummy variable for Democratic senators, with the coefficient on the dummy variable interpreted as a coefficient of party influence. Aside from (1) equating net party-specific inducements with party influence; (2) assuming no net party-specific inducements for lop-sided votes, this approach makes some strong additional assumptions: (3) the use of the Heckman-Snyder factor-analytic estimator in the first stage; (4) the linear functional form in the second stage (Synder and Groseclose use a Huber-White robust variance-covariance estimator to correct for the resulting heteroskedasticity); (5) the fact that the x_i result from a measurement procedure (the first-stage) and generate an "errors-in-variables" problem, which Synder and Groseclose tackle via instrumental variables.[24]

In contrast, our approach provides a direct way to test for party influence; our modeling approach is easily extended to let us embed parameters tapping party influence. That is, we fit one model to the data, with the statistical specification following directly from augmenting the utility functions of the legislators to permit vote-specific and party-specific inducements. In this way there is no need to break the estimation into separate pieces (one to recover ideal point estimates free of party effects, the other to recover estimates of effects, conditional on the ideal points recovered from the first stage): uncertainty in the recovered ideal points estimates propagates into uncertainty in the estimates of the vote-specific parameters (β_j, α_j and δ_j), and vice-versa.

We estimate our expanded model with 534 non-unanimous roll calls from the 105th U.S. Senate (55 Republicans, 45 Democrats). Senators Kennedy and Helms are constrained to have ideal points of -1 and 1, respectively and we impose the additional identifying constraint $\delta_j = 0$ among roll calls decided with majorities of 85% or more: 108 of the 534 roll calls meet this criteria. Of the remaining 426 roll calls, 124 (29.1%) have net party-specific inducements that are significantly different from zero[25]. Figure 6 shows that net party-specific inducements are at their largest for roll calls decided

[23] By way of analogy, consider standardized test items that (say, due to cultural biases) are suspected to be easier for group A than for group B: a phenomenon known as *differential item functioning* (DIF). Does better test performance by group A reflect higher ability than group B, or DIF? Unless we can identify a set of test items that are known to be DIF-free and use these items to pin down ability, then there is no way to distinguish differences in estimated ability from DIF.

[24] Another implementation of the two cut-point approach appears in McCarty, Poole and Rosenthal (2001), using a non-parametric optimal classification algorithm (Poole 2000): they compare the increase in classification success in moving from a one to two cutpoints.

[25] Again, in the sense that the 90% confidence interval does not overlap zero.

along party lines,[26] and declines as the votes become more lop-sided. Of the 277 roll calls decided by margins closer than 65-35, 102 (36.8%) appear to be shaped by party-specific inducements. And even among bills decided by margins more lop-sided than 65-35, we find 21 roll calls with party-specific inducements, amounting to 20% of all roll calls decided by margins between the 65-35 Snyder-Groseclose cutoff and our 85-15 cutoff (beyond which we assume there are no party-specific inducements). These estimates are generally smaller than those reported by Snyder and Groseclose (2000), and closer to the proportions found by Cox and Poole (2002).

Figure 7 plots the ideal points obtained from the augmented model against those from the standard model assuming no party-specific inducements. The standard model finds considerable separation between the parties, since any voting shaped by party-specific inducements is attributed to differences in the legislators' ideal points. Once we admit the possibility of party-specific inducements, the two partisan groupings (Democrats lying above the 45 degree line in Figure 7, Republicans below it) actually overlap: Breaux, the most conservative Democrat has roughly the same ideal point as the most liberal Republicans, Specter and Jeffords. Even after estimating net party-specific inducements specific to each roll call, we still find considerable partisan polarization, but substantially less polarization than the levels recovered from a conventional analysis.

7. Conclusion

Roll call analysis and the statistical operationalization of the Euclidean spatial voting model is a critical component of the scientific study of legislative politics. Although existing methods for the statistical analysis of roll call data have widespread acceptance and have been employed in many settings, the Bayesian simulation approach we present builds upon and improves extant methods in several ways. First, Bayesian methods permit auxiliary information to be brought to bear on roll call analysis in a straightforward and theoretically consistent way; this auxiliary information may include (but is not restricted to) expert judgments about dimensional structure, the location of extremist legislators, legislator-specific covariates, or the evolution of the legislative agenda. Second, the methodology we present is sufficiently flexible so as to easily accommodate alternative models of legislative behavior. For example, it is possible to permit ideal point estimates to change over time by modeling the process associated with that change (e.g., legislators switching affiliations between political parties). Finally, Bayesian simulation exploits tremendous increases in computing power available to social scientists over the last decade or so: estimation and inference via simulation --- long known to be an attractive statistical methodology (e.g., Metropolis and Ulam 1949) --- is now a reality. Consequently, our model works in any legislative setting, irrespective of the size of the legislature or its agenda.

Thus Bayesian methods make roll call analysis less a mechanical scaling exercise in which

[26] Indeed, for perfect party line votes, the data will drive the respective δ_j parameters towards $\pm\infty$, causing classical estimation procedures to break down; our use of proper priors alleviates this problem, bounding the posterior density away from numerically unstable regions.

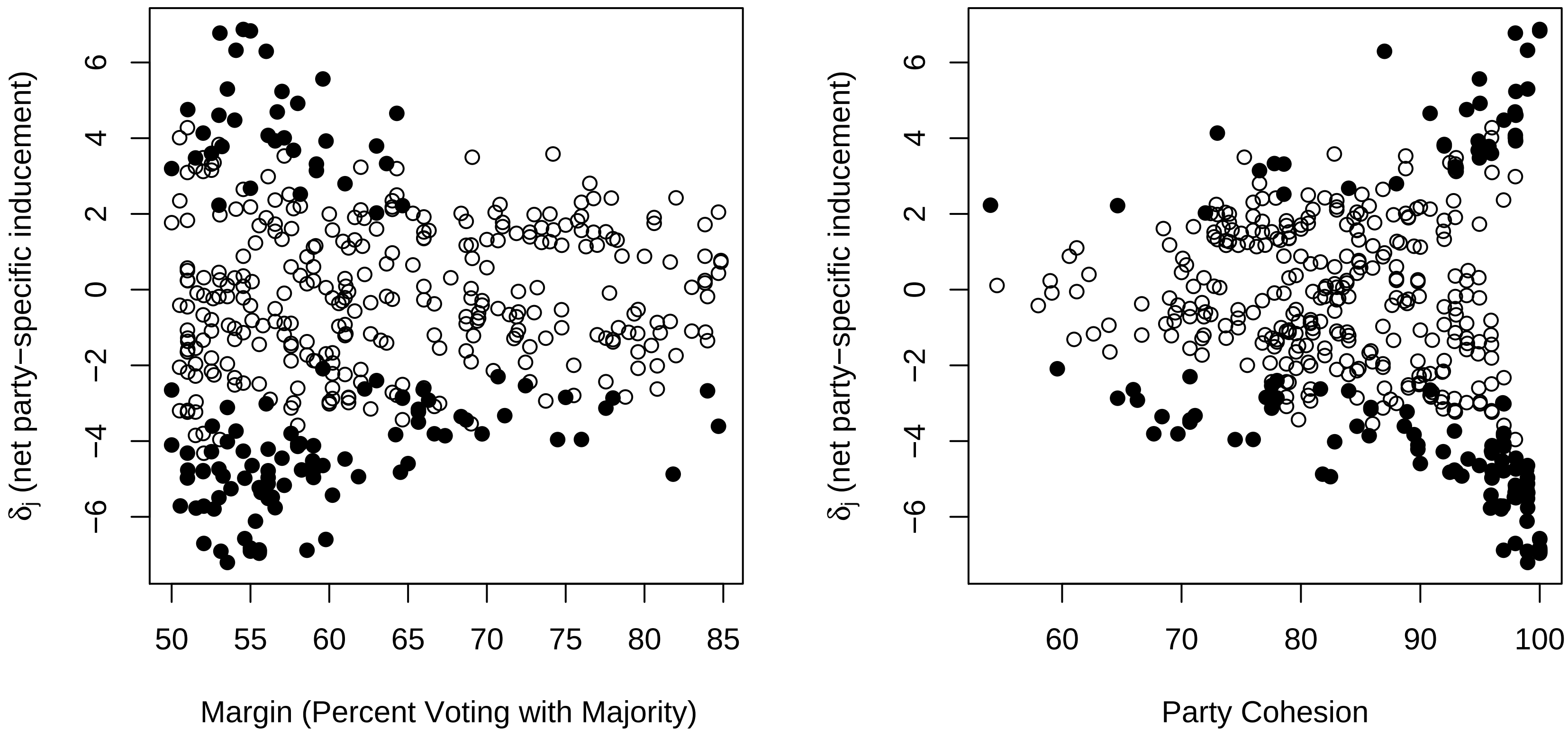

FIGURE 6. Net Party-Specific Inducements, 105th U.S. Senate, by Roll Call Margin and Party Cohesion. Each point represent posterior means of the net party-specific inducements parameters (δ_j), plotted against roll call margin and party cohesion (proportion of Democrats voting with majority of Democrats plus proportion of Republicans voting with majority of Republicans). Solid points indicate parameter significantly different from zero at conventional 95% level. The largest and most consistently significant net party-specific inducement estimates occur when the Senate splits on (or close to) party lines (55R-45D).

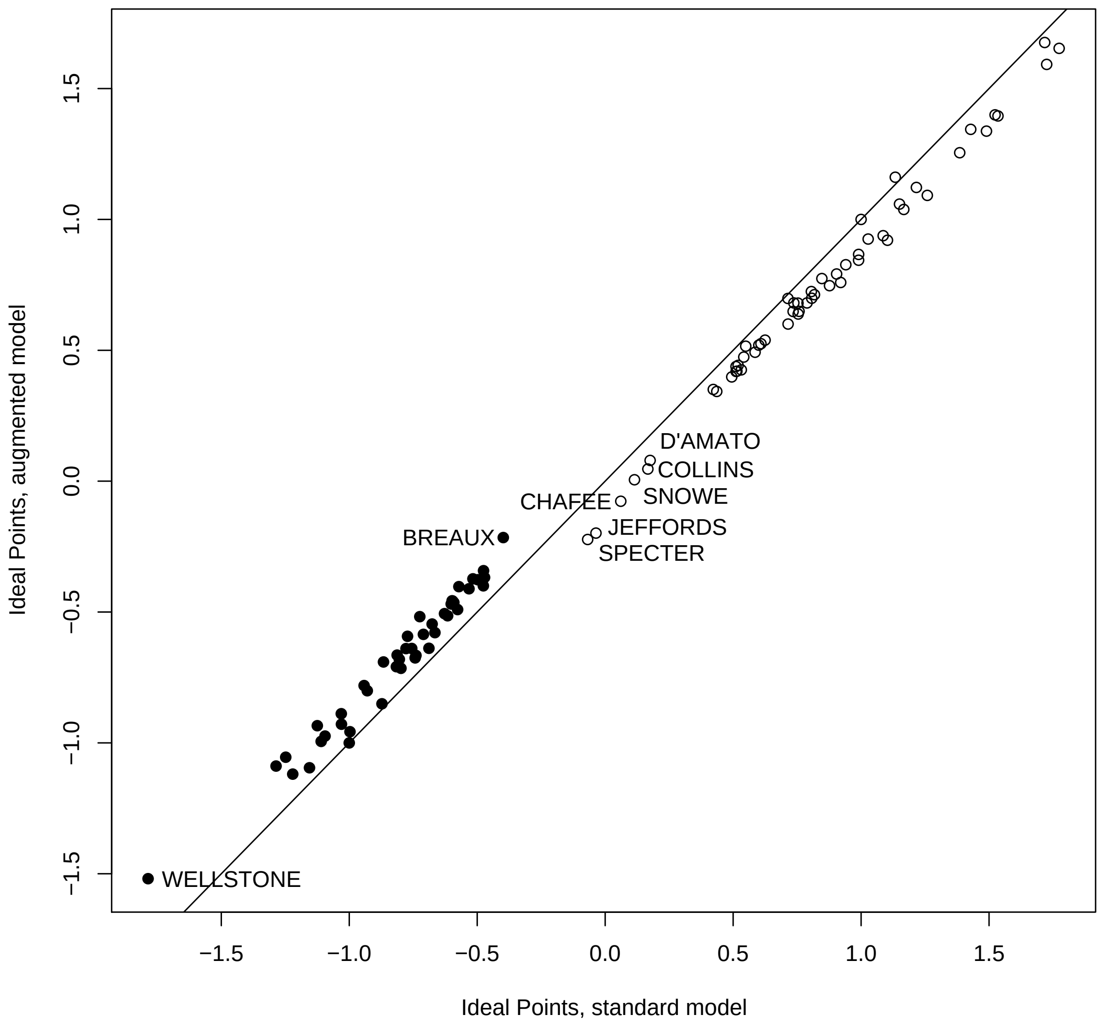

Figure 7. Ideal Point Estimates (Posterior Means), 105th U.S. Senate, standard model versus augmented model. Estimates from the standard model (assuming no party-specific inducements) are plotted against the horizontal axis; estimates from the model augmented to include party-specific inducements are plotted against the vertical axis. Democrats are solid points; Republicans are open circles. The diagonal line is a 45 degree line.

scholars simply feed roll call data to a "black box" algorithm, and more a way to empirically test theoretically interesting models of legislative behavior. In sum, the Bayesian simulation methodology we present lets scholars (1) incorporate substantive information about the proposals being voted upon or (2) about the preferences that structure the ideal points being estimated, (3) impose theoretically implied constraints on the standard model; (4) easily estimate and test alternative and models of legislator voting.

A. Appendix

A.1. Markov chain Monte Carlo Algorithm

The difference between the utilities of the alternatives on the jth roll call for the ith legislator is

$$y_{ij}^* \equiv U_i(\boldsymbol{\zeta}_j) - U_i(\boldsymbol{\psi}_j) = \boldsymbol{\beta}_j'\mathbf{x}_i - \alpha_j + \varepsilon_{ij}$$

where, for simplicity, we have set $\sigma_j = 1$. If $\boldsymbol{\beta}_j$ and α_j are given, then $\mathbf{x}_i$ is a vector of "regression coefficients" that can be imputed from the regression of $y_{ij}^* + \alpha_j$ on $\boldsymbol{\beta}_j$ using the m votes of legislator i. If $\mathbf{x}_i$ is given, then we use the votes of the n legislators on roll call j to impute $\boldsymbol{\beta}_j$ and α_j. Then given $\mathbf{x}_i$, $\boldsymbol{\beta}_j$, and α_j, the latent utility differences y_{ij}^* are simulated by drawing errors from a normal distribution subject to the constraints implied by the actual votes (if $y_{ij} = 1$ then $y_{ij}^* > 0$ and if $y_{ij} = 0$ then $y_{ij}^* < 0$), and the process repeats.

In our Bayesian approach, priors are required for all parameters: $\boldsymbol{\beta}_j$ and α_j, $j = 1, \ldots, m$ and $\mathbf{x}_i, i = 1, \ldots, n$. For the probit version of our model we have standard normal errors ε_{ij} and normal priors for the ideal points and the $\boldsymbol{\beta}_j$ and α_j parameters, leading to simple expressions for the conditional densities that drive the MCMC algorithm. For $\boldsymbol{\beta}_j$ and α_j, we denote the priors as $N(\boldsymbol{\tau}_0, \mathbf{T}_0)$; we generally choose vague priors by setting $\boldsymbol{\tau}_0 = \mathbf{0}$ and $\mathbf{T}_0 = \kappa \cdot \mathbf{I}_{d+1}$, with κ a large positive quantity (e.g., $\kappa = 5^2$); see 6.1 for the use of informative priors on particular "reference roll calls". For the latent traits, we use the normal prior $\mathbf{x}_i \overset{\text{iid}}{\sim} N(\mathbf{v}_i, \mathbf{V}_i)$, where usually $\mathbf{v}_i = \mathbf{0}$ and $\mathbf{V}_i = \mathbf{I}_d$ (an identity matrix of order d), but for legislators we are fixing to an set location (e.g., Kennedy and Helms, so as to normalize the scale of the latent traits), we use the prior $N(\mathbf{x}_i, \nu\mathbf{I}_d)$, where ν is an arbitrarily small, positive quantity (i.e., a "spike prior" at $\mathbf{x}_i$).

The goal is to compute the joint posterior density for all model parameters $\boldsymbol{\beta}_j$ and α_j, $j = 1, \ldots, m$ and $\mathbf{x}_i, i = 1, \ldots, n$. The MCMC algorithm provides a computer-intensive exploration or "random tour" of this joint density, by successively sampling from the conditional densities that together characterize the joint density. Augmenting the MCMC algorithm with the latent y_{ij}^* greatly simplifies the computation of the probit version of the model, letting us exploit standard results on the Bayesian analysis of linear regression models as we show below; we obtain y_{ij}^* by sampling from its predictive density given the current values of the other parameters and the roll call data. Letting t index iterations of the MCMC algorithm, iteration t of the algorithm comprises sampling from the following conditional densities:

1. $g(y^*_{ij}|y_{ij}, \mathbf{x}^*_i, \boldsymbol{\beta}_j \alpha_j)$. At the start of iteration t, we have $\boldsymbol{\beta}^{(t\text{-}1)}_j$, $\alpha^{(t\text{-}1)}_j$ and $\mathbf{x}^{(t\text{-}1)}_i$. We sample $y^{*(t)}_{ij}$ from one of the two following densities, depending on whether we observed a "Yea" (y_{ij} = 1) or a "Nay" (y_{ij} = 0):

$$\begin{aligned} y^*_{ij}|(y_{ij} = 0, x^{(t\text{-}1)}_i, \boldsymbol{\beta}^{(t\text{-}1)}_j, \alpha^{(t\text{-}1)}_j) &\sim N(\mu^{(t\text{-}1)}_{ij}, 1)\, I(y^*_{ij} < 0) \quad \text{(trunc. Normal)} \\ y^*_{ij}|(y_{ij} = 1, x^{(t\text{-}1)}_i, \boldsymbol{\beta}^{(t\text{-}1)}_j, \alpha^{(t\text{-}1)}_j) &\sim N(\mu^{(t\text{-}1)}_{ij}, 1)\, I(y^*_{ij} \geq 0) \quad \text{(trunc. Normal)} \end{aligned}$$

   where $\mu^{(t\text{-}1)}_{ij} = \mathbf{x}^{(t\text{-}1)}_i \boldsymbol{\beta}^{(t\text{-}1)}_j - \alpha^{(t\text{-}1)}_j$ and $I(\cdot)$ is an indicator function. For abstentions and other missing roll calls we sample $y^{*(t)}_{ij}$ from the untruncated Normal density $N(\mu^{(t\text{-}1)}_{ij}, 1)$, effectively generating multiple imputations for these missing data over iterations of the MCMC algorithm.

2. $g(\boldsymbol{\beta}_j, \alpha_j|\mathbf{X}, y^*_{ij})$. For $j = 1, \ldots, m$, sample $\boldsymbol{\beta}^{(t)}_j$ and $\alpha^{(t)}_j$ from the multivariate Normal density with mean vector $\left[\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{T}^{-1}_0\right]^{-1}\left[\mathbf{X}^{*\prime}\mathbf{y}^{*(t)}_{\cdot j} + \mathbf{T}^{-1}_0\boldsymbol{\tau}_0\right]$ and variance-covariance matrix $\left[\mathbf{X}^{*\prime}\mathbf{X}^* + \mathbf{T}^{-1}_0\right]^{-1}$, where $\mathbf{X}^*$ is a n-by-(d + 1) matrix with typical row $\mathbf{x}^*_i = (\mathbf{x}^{(t\text{-}1)}_i, -1)$, $\mathbf{y}^{*(t)}_{\cdot j}$ is a n-by-1 vector of sampled latent utility differentials for the j-th roll call, and recalling that $N(\boldsymbol{\tau}_0, \mathbf{T}_0)$ is the prior for $\boldsymbol{\beta}_j$ and α_j. This amount to running "Bayesian regressions" of $\mathbf{y}^{*(t)}_{\cdot j}$ on $\mathbf{x}^{(t\text{-}1)}_i$ and a negative intercept, and then sampling from the posterior density for the coefficients $\boldsymbol{\beta}_j$ and α_j, for $j = 1, \ldots, m$.

3. $g(\mathbf{x}_i|y^*_{ij}, \boldsymbol{\beta}_j, \alpha_j)$. Re-arranging the latent linear regression yields $w_{ij} = y^*_{ij} + \alpha_j = \mathbf{x}'_i\boldsymbol{\beta}_j + \varepsilon_{ij}$. Collapse these equations over the j subscript, to yield the n regressions $\mathbf{w}_i = \mathbf{B}\mathbf{x}_i + \boldsymbol{\varepsilon}_i$, where $\mathbf{B}$ is the m by d matrix with the j-th row given by $\boldsymbol{\beta}'_j$. That is, we have n regressions, with the ideal points $\mathbf{x}_i$ as parameters to be updated. Again exploiting conjugacy, the update is performed by sampling each $\mathbf{x}^{(t)}_i$ from the d-dimensional Normal density with mean vector $(\mathbf{B}'\mathbf{B} + \mathbf{V}^{-1}_i)^{-1}(\mathbf{B}'\mathbf{w}_i + \mathbf{V}^{-1}_i\mathbf{v}_i)$ and variance-covariance matrix $(\mathbf{B}'\mathbf{B} + \mathbf{V}^{-1}_i)^{-1}$. After updating all $\mathbf{x}_i$ ($i = 1, \ldots, n$), we optionally re-normalize the $\mathbf{x}_i$ to have zero mean and unit variance, dimension-by-dimension (say, when using the same prior on all $\mathbf{x}_i$ that do not provide a unique normalization of the latent traits).

Sampling from these distributions updates all the unknown quantities in the probit model. At the end of iteration t, denote the current values of the parameters of interest as $\boldsymbol{\xi}^{(t)} = (\mathbf{B}^{(t)}, \alpha^{(t)}, \mathbf{X}^{(t)})$. Iterating the MCMC algorithm produces a sequence $\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots$ that comprises a Markov chain, with the joint posterior density for $\boldsymbol{\xi}$ as its limiting distribution. That is, after a large number of iterations of the algorithm, successive samples of $\boldsymbol{\xi}$ are draws from its posterior density. These samples are saved and summarized for inference. Any function of these parameters can also be computed and saved, such as rank orderings of the legislators, pairwise comparisons of legislators, or the separating hyperplanes for particular roll calls.

The MCMC algorithm is initialized as follows. For the ideal points, we first double-center the n-by-m roll call matrix (subtracting out column and row means), compute a n-by-n correlation matrix from the double-centered roll call matrix, and extract its first d eigenvectors; this n-by-d matrix provides our initial values for $\mathbf{x}$, denoted $\mathbf{x}^{(0)}$. Note that these initial values are the estimates we would get from treating the ideal-point estimation problem as a principal-components factor analysis problem, ignoring the fact that the roll call data are not continuous variables (the binary character of the roll call data becomes less problematic as $m \to \infty$, and so for large roll call data sets from contemporary U.S. Congresses this procedure yields excellent start values). We are grateful to Keith Poole for suggesting this procedure, which is also used in NOMINATE.[27] For the bill-specific parameters $\boldsymbol{\beta}_j$ and α_j we obtain start values by running probits of the observed votes $\mathbf{y}_{\cdot j}$ on the $\mathbf{x}^{(0)}$, $j = 1, \ldots, m$.

With any Markov chain Monte Carlo approach, diagnosing convergence of the Markov chain is critical. Our experience is that the MCMC algorithm performs reasonably well for the roll call problem, moving away from start values to the neighborhood of a posterior mode quite quickly. For simple unidimensional fits, we usually let the sampler run for anywhere between 50,000 to 500,000 iterations, and then thin the output (storing the output of every 100-th to every 1,000-th iteration) so to produce a reasonable number of approximately independent samples from the posterior for inference (say, between 250 and 1,000 samples).

Figure 8 shows trace plots of the MCMC algorithm for single parameters (the ideal points of Kennedy, Collins and Helms, in the top three panels), and for the joint density of selected pairs of legislative ideal points in the lower panels. In the upper panels, the MCMC algorithm is approximately a random walk without drift in the parameter space, consistent with the sampler behaving well, randomly traversing the posterior density; the grey line indicates the posterior mean based on the post "burn-in" samples, and the dotted lines indicate 95% confidence intervals. The snaking solid lines are smoothed or moving averages, and a cumulative mean; note that after sufficient iterations, the cumulative mean of the MCMC algorithm becomes indistinguishable from the posterior mean, and the running mean slowly undulates about the posterior mean, indicating lack of drift (consistent with the sampler having converged on the posterior density). Half a million iterations were computed; so to as produce approximately independent samples from the posterior density, only every 1000-th iteration is retained for making inferences and have AR(1) parameters averaging about 0.06 (maximum AR(1) = .23, for Boxer). In the lower panels, each joint posterior mean is indicated with an open circle, and joint 95% confidence intervals are indicated with ellipses, the latter computed assuming the joint posterior densities can be approximated with bivariate normal densities.

A.2. Computing

For small roll call data sets, the free, general-purpose MCMC package WinBUGS (Spiegelhalter et al. 1997) can be used to implement our approach: only a few lines of WinBUGS commands are needed (an example appears on Jackman's website, http://jackman.stanford.edu/mcmc). Given the computational burden of analyzing larger roll call data sets, we use a `C` program, authored by

[27] In unpublished work, Poole has shown that if voting is perfect in one dimension, then the ranks of the first eigenvector of the correlation matrix of the double-centered roll call matrix recovers the rank ordering of the legislators' ideal points.

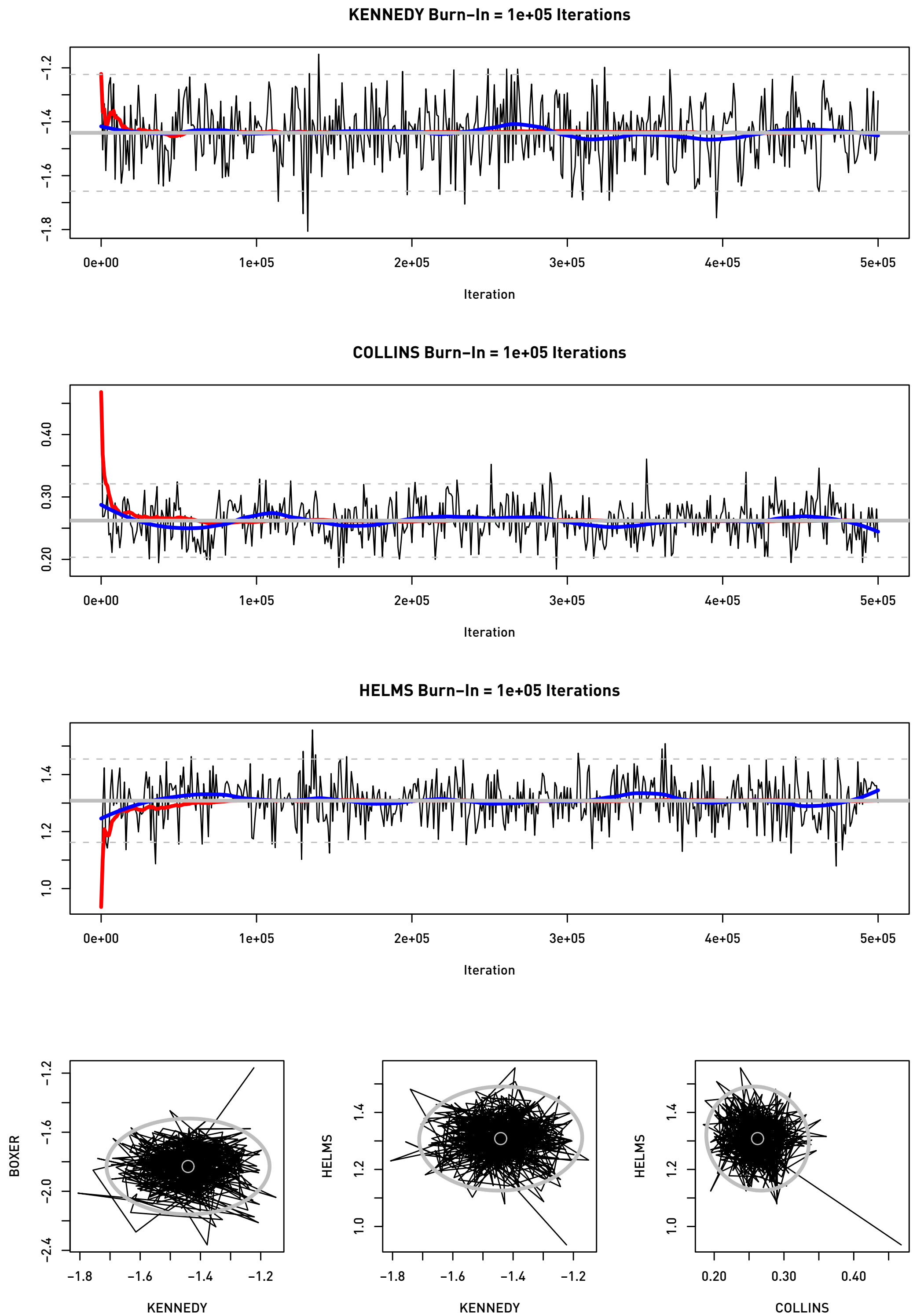

FIGURE 8. Iterative History of the MCMC algorithm, 106th U.S. Senate. The red line is a cumulative mean, the blue line is a moving average, and the grey line is the posterior mean, based on the post-burn-in iterations.

Jackman, implementing the MCMC algorithm for the probit model discussed above. We use WinBUGS for the logit version of the model. We use R for preparing roll call data sets for analysis, inspecting the output of the MCMC algorithm, and for producing the graphs and tables in the body of the paper. All code is available upon request.

Computing time is clearly a relevant consideration for simulation-based methods. Our experience is that compute time increases in nmT, where n is the number of legislators and m is the number of bills, and so nm is the number of individual voting decisions being modeled (assuming no abstentions or missing data), and T is the number of MCMC iterations desired. The computational cost of moving to higher dimensions is surprisingly small. The unidimensional 106th U.S. Senate example involved modeling n=102 ideal points, m=596 pairs of bill (or item) parameters, with 58,156 non-missing individual voting decisions. Half a million iterations of the MCMC algorithm required 3.2 hours on a 2.4GhZ Pentium machine, or about 2,600 iterations per minute; re-doing this analysis with a four dimensional model (using the interest group key votes) generated about 2,100 iterations per minute. For a one-dimensional fit to the 106th House (n = 440, m=1,073, with 444,326 non-missing individual voting decisions), a 150,000 iteration run took 7.3 hours, or about 335 iterations per minute, on the same hardware. For the one-dimensional fit to the U.S. Supreme Court (n=9, m=208, with 1,860 non-missing individual voting decisions), a 500,000 run took 8 minutes, or about 63,300 iterations per minute. These extremely long runs are usually not necessary, but we were being especially cautious about ensuring that the MCMC algorithm had converged on and thoroughly explored the posterior density.

References

Bailey, Michael and Douglas Rivers. 1997. "Ideal Point Estimation: A Survey." Paper presented at the Annual Meetings of the Midwest Political Science Association, Chicago.

Baker, Frank B. 1992. *Item Response Theory*. New York: Marcel Dekker.

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

Canes-Wrone, Brandice, David W. Brady and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96:127--140.

Clinton, Joshua D. and Adam Meirowitz. 2003. "Integrating Voting Theory and Roll Call Analysis: A Re-examination of the Compromise of 1790." Typescript. Department of Politics, Princeton University".

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

Cox, Gary and Keith T. Poole. 2002. "On Measuring Partisanship in Roll-Call Voting: The U.S. House of Representatives, 1877-1999." *American Journal of Political Science* 46:477--489.

Davis, Otto A., Melvin J. Hinich and Peter C. Ordeshook. 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64:426--48.

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.

Howell, William, Scott Adler, Charles Cameron and Charles Riemann. 2000. "Divided Government and the Legislative Productivity of Congress, 1945-94." *Legislative Studies Quarterly* 25:285--312.

Jenkins, Jeffrey A. 1999. "Examining the Bonding Effects of Party: A Comparative Analysis of Roll-Call Voting in the U.S. and Confederate Houses." *American Journal of Political Science* 43:1144--1165.

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

Krehbiel, Keith. 2003. "The Coefficient of Party Influence." *Political Analysis* 11:95--103.

Lancaster, Tony. 2000. "The Incidental Parameter Problem since 1948." *Journal of Econometrics* 95:391--413.

Londregan, John. 2000a. "Estimating Legislator's Preferred Points." *Political Analysis* 8:35--56.

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

Martin, Andrew and Kevin Quinn. 2001. "Re-Examining the Effect of Public Opinion on Judicial Behavior." Typescript. Department of Political Science, Washington University, St Louis.

McCarty, Nolan, Keith T. Poole and Howard Rosenthal. 2001. "The Hunt for Party Discipline in Congress." *American Political Science Review* 95:673--687.

Metropolis, N. and S. Ulam. 1949. "The Monte Carlo Method." *Journal of the American Statistical Association* 44:335--341.

Neath, Andrew A. and Francisco J. Samaniego. 1997. "On the Efficacy of Bayesian Inference for Nonidentifiable Models." *The American Statistician* 51:225--32.

Nokken, Timothy P. 2000. "Dynamics of Congressional Loyalty: Party Defection and Roll Call Behavior." *Legislative Studies Quarterly* 25:417--44.

Poole, Keith T. 2000. "A Non-Parametric Unfolding of Binary Choice Data." *Political Analysis* 8:211--37.

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

Rivers, Douglas. 2003. "Identification of Multidimensional Item-Response Models." Typescript. Department of Political Science, Stanford University.

Schickler, Eric. 2000. "Institutional Change in the House of Representatives, 1867-1998: A test of Partisan and Ideological Power Balance Models." *American Political Science Review* 94:269--288.

Snyder, Jr. James M. and Tim Groseclose. 2000. "Estimating Party Influence in Congressional Roll-Call Voting." *American Journal of Political Science* 44:187--205.

Spaeth, Harold J. 2001. "United States Supreme Court Judicial Database: 1953-1997 Terms." ICPSR Study 9422.

Spiegelhalter, David J., Andrew Thomas, Nicky Best and Wally R. Gilks. 1997. *BUGS 0.6: Bayesian inference using Gibbs sampling*. Cambridge, UK: MRC Biostatistics Unit.

Tierney, Luke. 1996. "Introduction to general state-space Markov chain theory." In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson and D. J. Spiegelhalter. London: Chapman and Hall pp. 59--74.

Voeten, Eric. 2000. "Clashes in the Assembly." *International Organization* 54:185--215.

Western, Bruce and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88:412--23.

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