Identifying Intraparty Voting Blocs in the U.K. House of Commons

Arthur SPIRLING and Kevin QUINN

Legislative voting records are an important source of information about legislator preferences, intraparty cohesiveness, and the divisiveness of various policy issues. Standard methods of analyzing a legislative voting record tend to have serious drawbacks when applied to legislatures, such as the United Kingdom House of Commons, that feature highly disciplined parties, strategic voting, and large amounts of missing data. We present a method (based on a Dirichlet process mixture model) for analyzing such voting records that does not suffer from these same problems. Our method is model-based and thus allows one to make probability statements about quantities of interest. It allows one to estimate the number of voting blocs within a party or any other group of members of parliament (MPs). Finally, it can be used as both a predictive model and an exploratory model. We illustrate these points through an application of the method to the voting records of Labour Party MPs in the 1997–2001 session of the U.K. House of Commons.

KEY WORDS: Dirichlet process mixture models; Political science; Roll call data.

1. INTRODUCTION

Scholars of legislative politics frequently rely on legislative voting records to make inferences about the policy preferences of legislators (Poole and Rosenthal 1985, 1991, 1997; Heckman and Snyder 1997; Clinton, Jackman, and Rivers 2004), the types of political issues that drive political conflict (Poole and Rosenthal 1997; Cowley and Garry 1998), the cohesiveness of parties (Hix 2002; Desposato 2003; Rosenthal and Voeten 2004), and the existence of intraparty factions (Cowley 2002; Rosenthal and Voeten 2004), among other things. In general, the methods used by political scientists to analyze such voting data are best applied to voting records from legislatures, such as the United States Senate and House of Representatives, that feature relatively weak parties composed of fairly autonomous legislators (Spirling and McLean 2007). The reason for this is that the usual models rely on a number of structural assumptions that are typically not even approximately met in parliamentary systems with highly disciplined parties organized around the dichotomy between government and opposition. In what follows, we present a nonparametric model that does perform well (for both exploratory data analysis and predictive inference) when applied to voting data from a parliament with extremely disciplined parties.

We apply this model to parliamentary votes from the United Kingdom House of Commons during the first Blair government (1997–2001), which has been of specific interest to political scientists for several reasons—not least the fact that it represented the first change in party control of British politics for over 18 years. Moreover, as the first "New" Labour government it was seemingly an uneasy alliance: on the one hand was a leadership that had actively abandoned the tenants of socialist policy making and that had received a landslide mandate to rule. On the other was a historically and openly recalcitrant tranche of "Old" Labour legislators, dismissive of the modernizing project in its entirety. There was thus much speculation that explosive divisions and factions would soon emerge to enervate the newly elected prime minister.

Our primary object of interest is the identification and interpretation of blocs of members of parliament (MPs) who have similar expected voting profiles. While much has been written on intraparty factions in the UK (Shaw 1988; Berrington and Hague 1998; Cowley 2002; Spirling and McLean 2007) most of this work relies on relatively crude methods of data analysis. By taking a more sophisticated model-based approach we hope to sharpen existing insights into intraparty divisions within the U.K. Labour Party. In addition, we hope to provide a set of tools that can be usefully applied to parliamentary voting more generally.

This paper is organized as follows. In the next section we briefly discuss voting in the U.K. House of Commons. In Section 3 we discuss previous approaches and their shortcomings when applied to U.K. roll call data. In Section 4 we describe our nonparametric model. Section 5 presents results from our analysis of the 1997–2001 U.K. data, and comments on how this new approach might contribute to recent substantive debates in political science.

2. DIVISIONS IN THE U.K. HOUSE OF COMMONS

The majoritarian nature of the British electoral system ensures that the prime minister is the leader of the largest party by vote share, and will hold a majority of seats in Parliament (Adonis 1990, pp. 21–25). The executive—the cabinet, headed by the prime minister—is fused with the legislature in that it essentially controls the business of the House of Commons. This business refers to the proposing and passing of legislation, and also to the timetabling of the debate on these matters. Formal (and informal) rules of parliament enable influence by the nongovernmental parties throughout the parliamentary session; these are essentially limited to the opportunity to debate and counterpropose legislation.

Arthur Spirling is Assistant Professor, Department of Government, Harvard University, Cambridge, MA 02138 (E-mail: *aspirling@gov. harvard.edu*). Kevin Quinn is Professor of Law, University of California Berkeley, School of Law, Berkeley, CA 94720-7200 (E-mail: *kquinn@law.berkeley.edu*). This work has benefited from helpful comments from panel participants at the 2005 Annual Meeting of the American Political Science Association, the 2005 Political Methodology Summer Meeting, and the 2005 Joint Statistical Meetings as well as seminar participants at Stanford University and Washington University. Quinn's research is supported by National Science Foundation grants SES 03-50613 and BCS 05-27513. Two referees and an Associate Editor provided very thorough comments that improved the content and structure of our paper. We are indebted to the Editor for his helpful advice and assistance.

Journal of the American Statistical Association
Accepted for publication, Applications and Case Studies
DOI: 10.1198/jasa.2009.ap07115

The largest parliamentary party not in government forms the "Official Opposition." They are joined on the opposition benches by all other (smaller) parties which are also outside of the government. Though technically incorrect, we use the term "opposition" as a catch-all for the nongovernment parties here. As its moniker suggests, the opposition's *de facto* role is to oppose the government's legislation. Since it normally lacks a unified political agenda (perhaps other than to defeat the government), the opposition's constituent parts may give dissimilar, even conflicting, reasons for their decision to vote (typically) contrarily to the government's agenda.

Congruent with this adversarial legislative principle, a MP relies on party patronage for career advancement, and their election at the constituency level is almost solely a product of their party label (and that party's current fortunes) rather than their *personal* appeal to voters (Jennings 1969). Almost all voters base their localized ballot box decision on the party labels of those competing: constituent service as a foundation of (re)election—in the U.S. sense of pork-barreling localized benefits—is absent.

By and large, then, roll calls in the U.K. Parliament amount to either supporting or opposing the government's proposals. The parties are highly cohesive in voting: the great majority (approximately 99%) of divisions are whipped, insofar as the respective party leaderships state a party line which MPs are expected (and have incentives) to toe. Rebelling—in the sense of voting against the official party line—scuppers promotion prospects and leads to demotion where applicable. If rebels are particularly reckless, the whip may be withdrawn (see Silk 1987, pp. 46–48). This latter punishment is the equivalent of expulsion from the parliamentary party, and usually prohibits the opportunity to run for that party at the next (and subsequent) general election(s). Contrast this situation with the U.S. Congress where representatives and senators from *different* parties form shifting coalitions from bill to bill.

Given this logic, it should be clear that a government party MP rebels when she votes *against* the explicit wishes of the government, and *with*—in terms of the similarity of the "Aye" or "No" division choice—the opposition parties. In particular, notice that these rebels need not agree *substantively* with any of the opposition parties' positions on the bill: for example, the government party rebels may feel some government bill does not allocate sufficient funds to public spending ("not enough!"), while others (particularly amongst the opposition) hope to defeat the bill on the converse basis that the proposal is profligate ("too much!"). Note that this logic is clearly inconsistent with the assumption of sincere policy-based voting that is at the heart of standard statistical models (Poole and Rosenthal 1985, 1991, 1997; Heckman and Snyder 1997; Clinton, Jackman, and Rivers 2004) of legislative voting. We now clarify this point before developing a different modeling strategy.

3. PREVIOUS LITERATURE AND APPROACHES

Early analysis of roll calls includes the foundational work of Rice (1928) who suggested a straightforward and well-known index, here denoted r, which attempts to capture the unity of a party on any particular vote. That is,

$$r = 100|p - (1 - p)|,$$

where p is the proportion of the party voting "aye" on the roll call. When scholarly interest centers on *aggregate* party voting, the index is very helpful: it allows us, for example, to compare the Labour party over the course of the parliament (mean $r = 96.99$, std dev $= 11.65$) to the main Conservative opposition (mean $r = 96.19$, std dev $= 14.54$). However, it tells us little about the number or membership of *intra*party blocs (we will use "groups" interchangeably in what follows) in the House of Commons—which is the focus here. This critique applies equally to related metrics that consider the number of divisions on which some (arbitrary) proportion of the party votes similarly (see, e.g., Lowell 1901).

3.1 Scaling and Projecting

Contemporary analysis of roll calls typically proceeds via an item-response approach similar to that implemented in the education testing literature as the one, two, or three parameter item response theory model (see Rasch 1961; Birnbaum 1968; Lord 1980). In this setup, "test-takers" are the legislators and the "items" are the bills to be voted on: rather than "ability," we seek the representatives' "ideal points" which are located in some multidimensional space. For political scientists, this modeling approach has a particularly tidy behaviorial justification, in that it can be seen as an empirical implication of the spatial theory of voting (Downs 1957; Poole 2005). A standard version of this setup has a legislator with ideal point $x_i \in \mathbb{R}$ considering a choice between the "aye" position ζ_j and "no" (or "status quo") position ψ_j for the jth bill. If legislators have quadratic utility functions over the policy space, the relevant comparison is between the utility garnered from the proposal, $U_i(\zeta_j) = -\|x_i - \zeta_j\|^2 + \eta_{ij}$ versus the utility from the status quo, $U_i(\psi_j) = -\|x_i - \psi_j\|^2 + \nu_{ij}$ where $\|\cdot\|$ is the Euclidean norm and $\boldsymbol{\eta}$ and $\boldsymbol{\nu}$ are error terms. The observed data is the roll call matrix $\mathbf{Y}$ with a typical element y_{ij} being zero (if the ith legislator votes to retain the status quo) or one (if she votes for the new proposal). With distributional assumptions on the stochastic disturbances, estimation may be via (some version of) maximum likelihood (e.g. Poole and Rosenthal 1997) or by way of Markov chain Monte Carlo techniques (e.g. Martin and Quinn 2002; Clinton, Jackman, and Rivers 2004).

Notice that, typically, no *a priori* restrictions are placed on the bills or members included in the analysis and that the researcher does not use any bill-specific ("this vote was about farming subsidies") or legislator-specific knowledge ("this senator is in the Republican party") to obtain estimates of the parameters of interest. In the context of Westminster politics, such scaling procedures will naturally place the government loyalists and the opposition at alternate ends of the ideological space since they have diametrically opposed records: the opposition votes nonsincerely and attempts to defeat the government at any opportunity. Meanwhile the rebels, who vote "with" the opposition some positive proportion of the time, will be placed somewhere in the "middle" (Spirling and McLean 2007). No straightforward postestimation rotation or stretching of that space can resolve the difficulty. Table 1 summarizes this problem, and reports the *relative* positions of five members whose comparative ideological positions and "types" are well known in the Westminster politics literature (see, e.g., Cowley 2002). To represent the loyalists, we report results for Giles

Table 1. Estimation of various MPs' spatial locations with standard routines

Ideological space (Groups)	Left Rebels		Center Govt. Loyalists		Right Opposition
Expected order	Corbyn	Skinner	Prescott	Radice	Hague
Estimated rank [Optimal Classification]	Radice [8.5]	Prescott [304.0]	Skinner [399.0]	Corbyn [419.0]	Hague [535.5]
Estimated position [NOMINATE]	Radice [−0.90]	Prescott [−0.85]	Skinner [−0.60]	Corbyn [−0.56]	Hague [0.90]
Estimated position [Bayesian 2P probit]	Radice [−0.14]	Prescott [−0.14]	Skinner [−0.13]	Corbyn [−0.12]	Hague [0.15]

NOTE: At the top of the table, in the first row, we give the *expected* positions/ordering from left to right: in reality, Corbyn and Skinner are the most left-wing, Prescott and Radice and somewhere in the middle while Hague is to the right of all the other members. The second row [in brackets] gives the *estimated* rank order of the members in question, via "Optimal Classification." The third and fourth rows give the *estimated* ideal points [in brackets] from two popular routines common in political science. Notice that nonsincere voting by the opposition means that "Rebels" from the governing party are placed "in the middle" when we *estimate* their location with standard statistical models: to the right of the government and to the left of the opposition. For example, substantively we know that Radice should have an ideal point closer to Hague, than Corbyn has to Hague—yet we see the opposite in rows 2, 3, and 4.

Radice who was an early reformer of the Labour party known for his fealty to the Blair government under study, and John Prescott, a former leftist who was deputy leader of the government during this period. Jeremy Corbyn and Dennis Skinner are, respectively, the most and the 10th most "rebellious" of the 1997–2001 Labour MPs, in terms of defying their party whip. The last member for consideration is William Hague, leader of the Conservative party and of the opposition, and no doubt to the right ideologically of the House of Commons as a whole. The table reports the estimated parliamentarian positions from three popular routines common in political science: a nonparametric cutting procedure known as "Optimal Classification" (Poole 2000), a scaling model known as "NOMINATE" (Poole and Rosenthal 1997) and a two-parameter probit model estimated via MCMC with diffuse priors (Clinton, Jackman, and Rivers 2004). The data consist of the votes of every Labour and Conservative MP on every division. The results clearly clash with our strong substantive priors; we see the left-wing rebels to the *right* of the loyalists, and closer to the Conservative opposition than the government apparatchiks.

There are, of course, simpler alternatives to those suggested above, and MacRae (1970) discusses several. The researcher could, for example, obtain the Euclidean (or other) distance between rows of the roll-call matrix $\mathbf{Y}$, and then apply, say, principal component analysis to the resulting dissimilarities. A practical concern here is how to deal with missing observations in $\mathbf{Y}$, but, no matter how that is solved, the problem noted above reemerges. Consider the partition of the distance matrix that includes just the rows dealing with our example MPs, as displayed in Table 2. Again, we see that Hague, the opposition leader, is "closer" to Corbyn (the left-wing Labour rebel) than he is to Prescott (the government loyalist). Unsurprisingly, neither principal components analysis nor factor analysis does much better.

3.2 Clustering

Explicitly searching for groups within parties is a sensible alternative to "scaling." Part of the appeal stems from the fact that we need not assume independence among observations when interpreting the clusters. Rather than asserting that *individuals* have latent traits (and that their errors are independent of one another) cluster approaches are congruent with a conglomeration of members who may be actively influencing one another to vote certain ways.

The methodological options are many. Of the *partition* approaches, K-means (and its derivatives) is perhaps best known (MacQueen 1967; Hartigan 1975). Standard drawbacks include the fact that missingness is not handled by the routine and the number of clusters must be *a priori* specified (much as the number of dimensions must be decided for an item response approach). This latter point means that searching for an "optimal" choice of K is not generally possible. Instead, one can conceive of the data as the result of a mixture distribution, the "optimal" number of components of which may be estimated (Fraley and Raftery 2002). Typically though, these techniques assume that the mixture is composed of multivariate normals, which is an odd modeling choice when the voting responses in question are binary. This criticism similarly applies to model-based hierarchical clustering algorithms. In any case though, investigating the substantive nature of the clusters is problematic. This is primarily because we do not have immediate access to the criteria—in terms of the divisions—around which the clusters form. Thus, we cannot know from the estimation *itself* whether a cluster corresponds to, say, "prolife members" or "antideath penalty liberals" or "antitax libertarians" etc. Hartigan (2000) considers a "partition model" wherein each bloc of legislators

Table 2. Distance matrix (partition of whole matrix) for MPs in running example

	Corbyn	Skinner	Prescott	Radice	Hague
Corbyn	0.00	5.63	10.75	10.01	$\boxed{33.74}$
Skinner	5.63	0.00	8.83	8.20	34.17
Prescott	10.75	8.83	0.00	3.73	$\boxed{34.35}$
Radice	10.01	8.20	3.73	0.00	34.54
Hague	33.74	34.17	34.35	34.54	0.00

NOTE: Notice that Corbyn is "closer" to Hague than Prescott is to Hague. This does not accord with our priors about ideological position.

votes similarly to one another on some set of bills; this approach allows for possible cross-party voting that would seem incompatible with projecting the roll matrix to one or two dimensions. Though estimated via Markov chain Monte Carlo, Hartigan's method does not facilitate explicit comparison of models via a posterior over the number of blocs in the data, which might have a philosophical appeal from a substantive political science stand point. Moreover, the approach is reasonably computationally intensive: for 100 senators, and 101 votes, a billion partition pairs must be checked. In our whole data set (i.e., Labour and main opposition members), there are 591 legislators, and 1278 bills—suggesting a somewhat larger estimation problem.

All told, studying *intra*party blocs is suitably approached as a clustering problem. Desiderata include a model-based method that allows for a binomial data-generating process. We would like to make uncertainty statements over cluster membership and know on what basis the blocs are divided. Moreover, we would prefer an approach that is able to deal with missing observations in a reasonable manner.

4. A NONPARAMETRIC MODEL FOR DIVISIONS DATA

Before describing the model's formal components, it is important to be clear about the *philosophical* task at hand. We wish to provide a sensible *summary* of the data that has internal validity—that is, we wish to uncover patterns in the roll calls that are helpful to substantive researchers. We are less concerned about the external validity of our estimates; otherwise put, we are not attempting to provide a general empirical picture of Westminster politics for all times and situations.

Let $i = 1, \ldots, I$ index MPs and $j = 1, \ldots, J$ index divisions. Our goal is to model the $I \times J$ matrix of observed votes by the I MPs on the J divisions. We let $\mathbf{Y}$ denote this matrix, $\mathbf{y}_i$ denote the vector of votes specific to MP i, and y_{ij} the observed vote of MP i on division j. By convention, we code $y_{ij} = 1$ if MP i voted "Aye" on division j, $y_{ij} = 0$ if MP i voted "No" on division j, and y_{ij} is coded as missing if a formal vote was not recorded for MP i on division j. All data analyzed below are from Firth and Spirling (2005), who have written a package for the `R` statistical language and environment—called `tapiR`—that downloads roll-call records and formats them in a helpful way for analysts.

4.1 The Likelihood

Given the binary nature of the observed vote matrix, a natural choice for the likelihood of MP i's vector of observed votes $\mathbf{y}_i$ is

$$L(\mathbf{y}_i|\boldsymbol{\theta}_i) = \prod_{j=1}^{J} \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}}, \tag{1}$$

where $0 \le \theta_{ij} \le 1$ gives the probability that MP i votes "Aye" on division j for $j = 1, \ldots, J$. We assume that missing values of y_{ij} do not enter into the product above. Assuming the votes of MPs i and i' are independent given $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i'}$ for all i and i' we can write the likelihood of the entire observed vote matrix as

$$L(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}}. \tag{2}$$

As the reader can easily see, the likelihood has as many free parameters as there are observed data points and is thus primarily useful as a starting point for a hierarchical model rather than a final probability model in its own right. The next subsection details the prior used in our model and shows how this prior greatly reduces the effective number of parameters to be estimated.

4.2 The Prior

Our modeling strategy borrows heavily from recent work on Bayesian nonparametrics (Escobar and West 1995, 1998; Neal 2000; Dahl 2003; see also Blackwell and MacQueen 1973; Ferguson 1973; and Antoniak 1974). In particular, we assume that each $\boldsymbol{\theta}_i$ is *a priori* drawn from a distribution G:

$$[\boldsymbol{\theta}_i|G] \sim G.$$

Our model can be considered nonparametric in that we do not specify the parametric form of G. Instead, we assume that G is generated by a Dirichlet process with parameters α and $\boldsymbol{\lambda}$ which we write as

$$[G|\alpha, \boldsymbol{\lambda}] \sim \mathcal{DP}(\alpha G_0(\cdot|\boldsymbol{\lambda})).$$

Here G_0 is the centering distribution and represents, in a certain sense, one's best guess as to the shape of G. The $\alpha > 0$ is a concentration parameter that determines how close realizations of G are to the centering distribution G_0. As α gets larger, realizations of G tend to more closely resemble G_0. Finally, $\boldsymbol{\lambda}$ represents a vector of hyperparameters that determine the shape of G_0. For the work in this paper we assume that G_0 is the distribution with density

$$g_0(\boldsymbol{\theta}_i|\boldsymbol{\lambda}) = \prod_{j=1}^{J} \left\{ \frac{\Gamma(\lambda_{1j} + \lambda_{0j})}{\Gamma(\lambda_{1j})\Gamma(\lambda_{0j})} \theta_{ij}^{\lambda_{1j}-1} (1 - \theta_{ij})^{\lambda_{0j}-1} \right\}.$$

In words, G_0 is the product of J beta distributions.

One property of the Dirichlet process that is of use in our application is that realizations of G will be discrete with probability one (Ferguson 1983). This property of the Dirichlet process is useful in our application because it allows us to view our model as a countably infinite mixture model and to perform inference not only about which MPs are likely to cluster together in voting blocs (i.e., have the same value of $\boldsymbol{\theta}$) but also about the number of distinct voting blocs. In what follows, we let K denote the number of discrete support points of G. We note that the value of the concentration parameter α induces a prior distribution over K—as α gets larger more mass is assigned to larger values of K.

To complete the model, prior distributions for α and $\boldsymbol{\lambda}$ need to be chosen. While α could be fixed at a user-specified value we prefer to give it a proper prior distribution and estimate it. In what follows we assume that α follows a gamma distribution with shape a_0 and inverse scale b_0. As noted above our prior beliefs about α imply prior beliefs about the number K of clusters. Herein, we set $a_0 = 4$ and $b_0 = 0.75$. Via simulation it was found that when α is equal to its prior 5th percentile (1.82) 10 clusters contain 95% of 400 observations. When α is set equal to its prior median (4.90) 20 clusters contain 95% of 400 observations. Finally, when α is set equal to its prior 95th percentile 33 clusters will contain 95% of 400 observations. These numbers seem reasonable to us given what we know about the amount of party discipline in the British Labour Party during this time period.

We complete the prior by assuming $\lambda_{0j} = \lambda_{1j} = 0.1$ for all $j = 1, \ldots, J$. This is consistent with a prior belief that the expected probability of a randomly chosen MP voting "aye" on a randomly chosen division j (θ_{ij}) is the same as the expected probability of that MP voting "no" on that division ($1 - \theta_{ij}$) and that these probabilities are not close to 0.5. Specifically, our assumption implies that under the centering distribution $\Pr(\theta_{ij} > 0.95) = \Pr(\theta_{ij} \leq 0.05) = 0.378$. This prior decision is consistent with our goal of finding voting blocs that are relatively homogenous in terms of observed votes. Allowing for more prior mass near 0.5 would allow for more within-bloc heterogeneity which we explicitly want to avoid.

It is important to be clear about the nature and stringency of the independence assumptions—across votes and across MPs—being made in our model. As per Equation (1), we require *conditional* independence of MP i's voting decisions *given* MP i's vector of vote probabilities ($\boldsymbol{\theta}_i$). Since $\boldsymbol{\theta}_i$ has as many elements as there are voting decisions, and we have chosen to set the prior so that most of the elements of $\boldsymbol{\theta}_i$ are very close to either 0 or 1, this independence assumption is relatively innocuous. Relatedly, we assume that $\mathbf{y}_i$ is independent of $\mathbf{y}_{i'}$ given $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{i'}$ [see Equation (2)]. While this assumption is unlikely to be correct, we don't see this as particularly problematic if our goal is to summarize an existing vote matrix. Note that if $\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i'}$ MPs i and i' are in the same cluster, thus a very strong form of dependence is actually allowed in our setup. The conditional independence assumptions we make would be more worrisome if we hoped to use our model to examine counterfactuals of the sort, "how would MP i vote if MPs i' and i'' both vote 'aye'?" We are not interested in such questions here.

5. RESULTS

Our primary focus is on identifying intraparty voting blocs in the British Labour Party during the first Blair government (1997–2001) along with the divisions that separated these major groupings of MPs. In what follows, we analyze the Labour MPs behavior without reference to the behavior of MPs from rival parties. With two exceptions, no Labour MP has a voting record that even remotely resembles the voting record of any Conservative MP. The exceptions are Shaun Woodward and Peter Temple-Morris who switched parties during the 1997–2001 session of Parliament.

5.1 Labour Party Voting Blocs 1997–2001

After dropping unanimous votes and MPs who never voted (along with the party-switches Shaun Woodward and Peter Temple-Morris) we have $I = 424$ Labour MPs and $J = 198$ nonunanimous (within the Labour MPs) divisions during the 1997–2001 Parliament. In Figure 1 we summarize the data. The thin grey line in the main body of the plot records the proportion of the party voting in the minority for any given bill in the dataset (which are arranged chronologically on the x-axis). Notice that this proportion has reasonable variance over time, although its moving average—described by the broken loess line—is relatively low. The thick black line gives the *cumulative* proportion of members who have voted with the minority

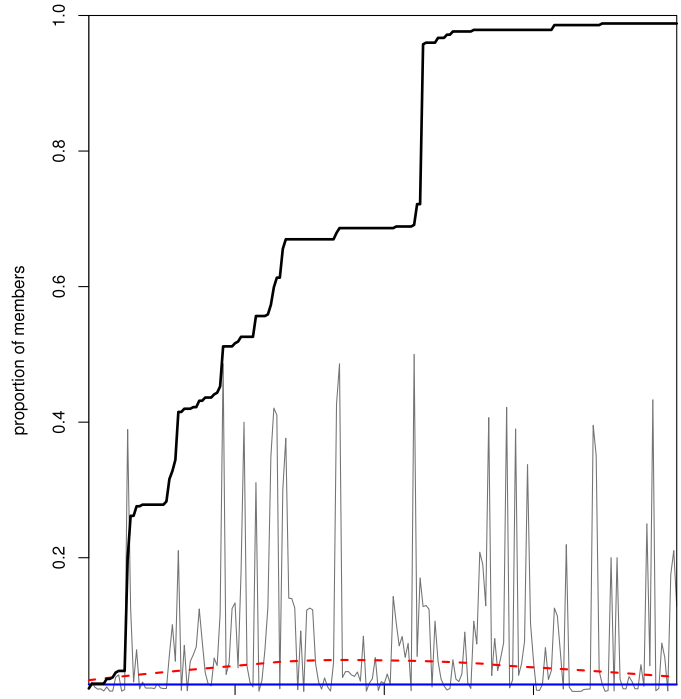

Figure 1. Voting summary for Labour MPs 1997–2001. Thin grey line is proportion of MPs voting in minority on any given vote (broken line is loess moving average of this proportion). Thick black line is cumulative proportion of MPs voting in minority over time. The intercept of the x-axis with the y-axis is the mean proportion (essentially zero) of MPs voting with the minority of their party over the entire universe of bills for the period—which includes some 1279 divisions. A color version of this figure is available in the electronic version of this article.

of their party as each bill occurs. Note that, of 424 members, almost every single one has deviated from the majority view at least once by the time the parliament ends. Finally, the x-axis itself is the *mean* proportion of members voting differently to the majority over *all* (1279) bills in the parliament. This is very close to zero, demonstrating the generally strong party discipline in the Commons. All told, it is not immediately clear whether there are multiple distinct subgroups of Labour MPs who tend to vote similarly. Our modeling goal is to determine this.

The priors used to generate the results discussed below are that $\alpha \sim \mathcal{G}\text{amma}(4, 0.75)$ and $\lambda_{0j} = \lambda_{1j} = 0.1$ for $j = 1, \ldots, J$. As noted above, the prior for α was chosen based on our prior beliefs about the number of groups and that fact that we believe that there will be very few occasions when it is sensible to think the probability of an MP voting in either direction is close to 0.5.

The model was fitted to the Labour data using Markov chain Monte Carlo (MCMC). Details of the MCMC algorithm employed are provided in the Appendix. The chain was run for 510,000 iterations and the first 10,000 iterations were discarded as burn-in. Because of memory constraints every 100th draw after burn-in was stored for a total of 5,000 stored draws.

Figure 2 shows the marginal posterior density of α. As can be seen here, there is information in the data about the likely value of α. In particular, the data suggest that α is toward the lower end of the range we deemed *a priori* likely.

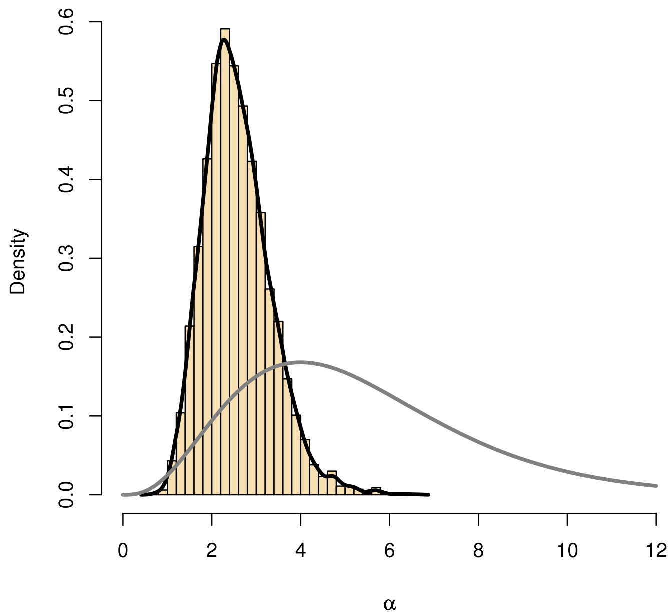

Figure 2. Marginal posterior density of α in fit to Labour 1997–2001 data. The dark line is a kernel density estimate of the marginal posterior and the light line is the $\mathcal{G}amma(4, 0.75)$ prior density. A color version of this figure is available in the electronic version of this article.

The α parameter is really of interest primarily because of its effect on the number of clusters (voting blocs) that are supported by the data. Looking directly at the posterior distribution over the number of clusters K is a more intuitive way to get some sense of the extent of clustering. This distribution is presented in Table 3.

As can be seen from Table 3, there is strong evidence that there are either 11 or 12 distinct voting blocs within the Labour party during the first Blair government. The posterior probability of their being either 11 or 12 blocs is 0.882.

While the number of voting blocs is of some interest we are primarily concerned with the size and membership of each of the blocs. Because the number of clusters is not constant and the cluster labels are completely arbitrary (and change over the course of the MCMC sampling) it is not meaningful to look at the number and identities of MPs in a cluster with a particular label. Instead, we focus on the probability that any two Labour MPs are in the same cluster for all pairs of Labour MPs; that is, $\Pr(\boldsymbol{\theta}_i = \boldsymbol{\theta}_{i'})$, $i = 1, \ldots, I$, $i' = 1, \ldots, I$. These probabilities do not depend on the cluster labels or the number of clusters and are thus meaningful quantities. Taken together, these probabilities define an $I \times I$ matrix $\mathbf{P}$.

It is worth emphasizing that $\mathbf{P}$ contains qualitatively different information than that usually examined by scholars of Westminster voting. In particular, previous work focuses almost exclusively on voting in the Commons in terms of "rebellion" by backbenchers. In those accounts, behavior is binary: either MPs obey the whip, or they do not. Exemplars in this vein include comprehensive tomes by Norton (1975, 1980) and detailed accounts by Cowley (2002, 2005) where the goal is to describe particular episodes of dissent, those involved and the effect on policy making. In this undertaking though, notice that we are not constrained by previously defined categorization of either MPs ("rebellious," "loyalist," etc.) nor divisions ("controversial," "whipped," etc.). Indeed, several of our clusters below arise from bills that few scholars would necessarily know *a priori* were useful for categorization of actors. The result is thus a more nuanced picture of behavior and structure in the (governing) Labour party. Some scholars, for example, Dunleavy (1993), have looked at *intra*party dimensions of conflict beyond a "left" versus "right" or "leadership" versus "backbenchers" dichotomy though such accounts do not often use data to test hypotheses *per se*. Thus our approach and findings provide a new testing ground for such accounts.

Since $\mathbf{P}$ is too large to summarize directly, we sort the matrix to reveal the underlying groups. We do this as follows. First, we create a dissimilarity matrix $\mathbf{D}$ based on Euclidean distance between the columns of $\mathbf{P}$ (we experimented with several metrics, and all give approximately similar results). Seriation with a minimized Hamiltonian path is then used to sort the rows and columns of $\mathbf{P}$. This creates a two-dimensional representation of the MPs' expected voting profiles in which MPs with similar voting profiles are located adjacent to each other (see Hahsler, Buchta, and Hornik 2008, for details). The results are plotted in Figure 3, and the names refer to individuals we discuss below. Darker (square) areas are groups that are tightly bound together in terms of their cluster membership. Notice that the plot is symmetric, but that there is no *substantive* sense in which we have a continuum of groups from, say, "left" to "right" in the

Table 3. Posterior distribution of number of clusters (Labour 1997–2001)

	$k = 10$	$k = 11$	$k = 12$	$k = 13$	$k = 14$
$\Pr(K = k \mid \mathbf{Y})$	0.0478	0.4402	0.4418	0.0684	0.0018

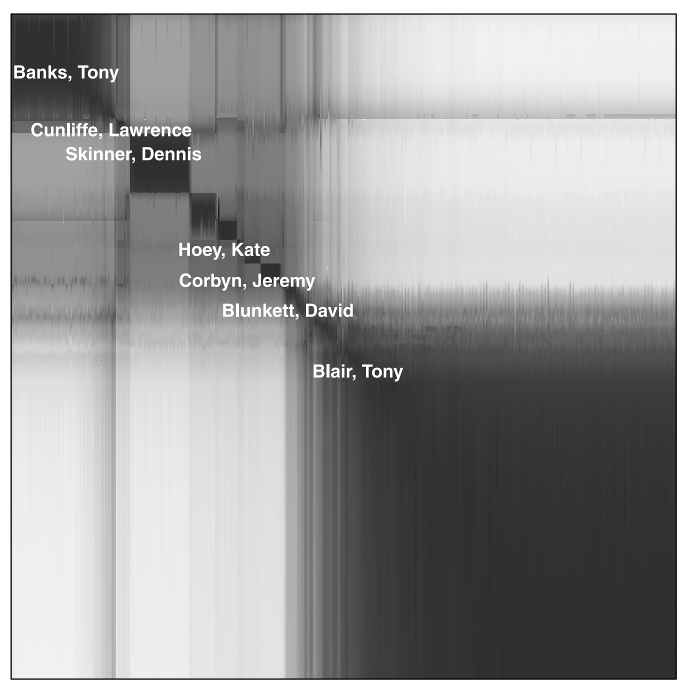

Figure 3. Seriation representation of expected voting groups (Labour MPs 1997–2001). Loyalists are bottom right dark square. MPs in Bloc 1 are in the upper left corner (Tony Banks). MPs in Bloc 12 are in the lower right corner (Tony Blair). Numbering of the blocs proceeds down the main diagonal from 1 to 12.

party; this logic applies similarly to our use of *nominal* numerical labels for the blocs in what follows.

5.1.1 Core Loyalists. The first bloc are the core Labour government loyalists who consistently support the government on almost all divisions. It is very likely to contain Prime Minister Tony Blair, Chancellor Gordon Brown, and Foreign Secretary Jack Straw. Indeed, with just three exceptions, every other member of the Cabinet that served between 1997 and 2001 inhabits this group. Intriguingly, of these three ministerial deviants, two were fired in 1998 after just a year of service to the prime minister. This group also includes the majority of the members joining the Commons for the first time in 1997. Bloc 11 is made up of similarly loyal MPs such as Giles Radice, an early reformer in the "New Labour" mold, along with David Blunkett, the prime minister's first choice for secretary of state for education. The primary difference between this group and the core loyalists is over the choice of speaker. In the House of Commons, this is a nonpartisan position, in which the governing party often has a preferred candidate but whose election it does not whip. In this particular case, these MPs were less supportive of the purely loyalist choice of Michael Martin (the eventual winner). All told, the loyalists constitute about 60% of all Labour members.

5.1.2 Leftists. Bloc 9 are the "hardcore" rebels in the parliamentary party, who took a general disliking to the government's policy plans. This group includes "usual suspects" like Diane Abbott, Tony Benn, Bernie Grant, and Jeremy Corbyn. All of these members are part of the left-wing "Socialist Campaign Group" that criticizes New Labour relentlessly from the backbenches. They disapproved of government attempts to, *interalia*, cut benefits to lone parents and disabled people, to privatize the national air control system, and to reform trial-by-jury procedures. Other groups in the data are similar to this leftish caucus, but are exorcized by particular subsets of issues: the welfare socialists of Bloc 4 (which includes former coal miner Dennis Skinner) were keen to avoid changes to the benefit system; Lawrence Cunliffe (who forms Group 2) was unimpressed with plans to cut funding for university students; Bloc 3, which includes the fired cabinet minister David Clark, wanted the Freedom of Information Act (2000) to go much further.

5.1.3 Mavericks. Several members, and several groups, are not best described as either loyal or rebellious. Instead, they take unusual lines on certain (typically free) votes, that suggest they are ideologically different to the bulk of their colleagues. For example, members of Blocs 1 (which includes sometime minister Tony Banks) and 10 generally felt that the City of London Ward Election bill—which dealt with possible democratization of an ancient arrangement for municipal governance in the capital—should go further in giving citizens voting rights. By contrast, Blocs 6 and 8 were concerned that legislation dealing with firearms would inappropriately punish law-abiding sporting shooters. Intriguingly, Bloc 6, which includes Kate Hoey (now the chair of a major countryside interest group), was unusually keen on fox hunting, a position associated with rural and aristocratic conservatives. For a party that stresses equal rights and opportunities, it is odd that members such as Denzil Davies (forming Bloc 7) and those in Bloc 5 would be skeptical of government-backed plans to reduce the age of homosexual consent from 18 to 16 (thus bringing it on par with that of heterosexual consent). Nonetheless, these individuals made their views known in the voting lobbies of the Commons.

5.1.4 Judging the Extent of Rebellious Behavior. To gain some additional insight into the nature and degree of rebellious voting behavior in the Labour party during this time period we calculate an average voting profile for each of the voting blocs. We let $\bar{\mathbf{y}}_k$ denote the average voting profile for bloc k. The jth element of $\bar{\mathbf{y}}_k$ is simply taken to be the sample average $\bar{y}_{k,j} = I^{-1} \sum_{i \in B_k} y_{i,j}$, where B_k denotes the kth voting block as derived from the seriation of $\mathbf{P}$ above. With these average voting profiles in hand we can make a number of comparisons between groups to examine the extent of their divergence in voting behavior. Our measure of voting similarity between bloc k and bloc k' is

$$s_{i,i'} = 1 - J^{-1} \sum_{j=1}^{J} \mathbb{I}[|\bar{y}_{k,j} - \bar{y}_{k',j}| > 0.5]$$

which is 1 minus the fraction of absolute differences in average votes that are greater than 0.5. A value of 1 indicates that blocs k and k' have identical voting profiles, while a value of 0 indicates that all elements of the voting profiles differ by more than 0.5.

Figure 4 presents comparisons of all blocs to the Labour Loyalist group (group 12) and all blocs to Bloc 9 (the Socialist Campaign Group rebels). Here we see that even the blocs most similar to the Labour loyalists differ substantially from the loyalists on about 5% of votes. On the other hand, four of the eleven nonloyalists blocs differ substantially from the loyalists on between 20 and 50% of the votes. It is also the case

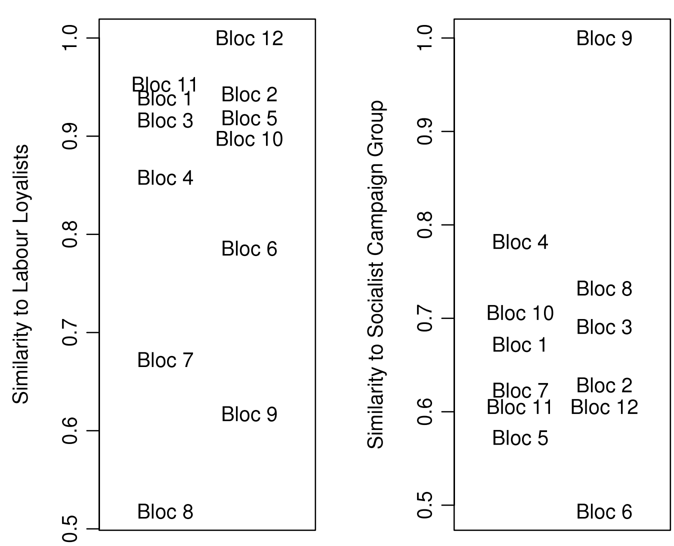

Figure 4. Degree of similarity to Labour loyalists and Socialist Campaign Group rebels. The y-axis in the left panel depicts $1 - J^{-1} \sum_{j=1}^{J} \mathbb{I}[|\bar{y}_{12,j} - \bar{y}_{k,j}| > 0.5]$ and the y-axis in the right panel depicts $1 - J^{-1} \sum_{j=1}^{J} \mathbb{I}[|\bar{y}_{9,j} - \bar{y}_{k,j}| > 0.5]$. Note that that even after excluding Bloc 8 (which only consists of two MPs) that similarity to the Labour loyalists (Bloc 12) is not monotonically related to similarity to the Socialist Campaign Group (Bloc 9). This is consistent with there being multiple types of "rebellious" voting in the House of Commons. Note that the horizontal location of each point has been shifted purely for display reasons.

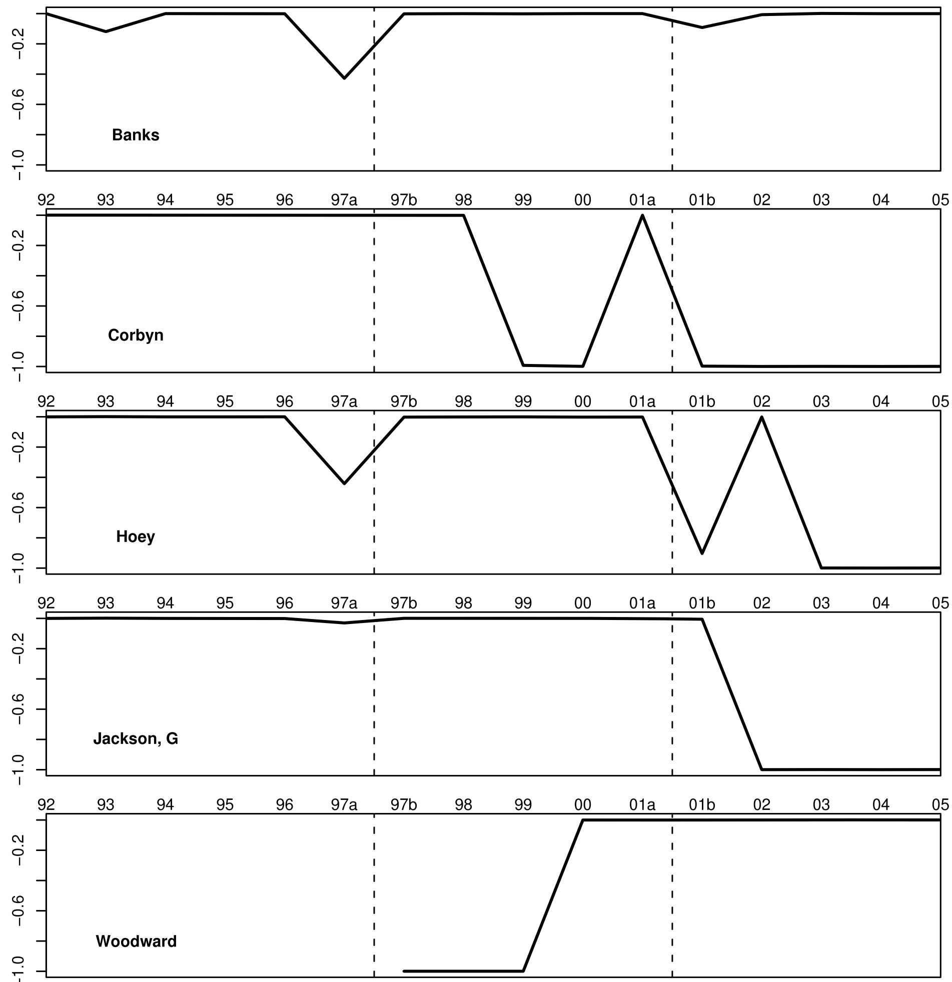

Figure 5. Probability profile, in terms of "loyalist" cluster membership, for different "types" of MPs over time in the House of Commons. Broken lines are election dates: "a" denotes the period in parliament before the election in a given year and "b" the period after.

that blocs that appear to be roughly equally similar to the loyalist bloc (say Blocs 3 and 5) have very different similarities to the Socialist Campaign Group bloc. This highlights the point that the groups are being formed not just on the basis of their similarity to the Labour loyalists but on their overall similarity to other voting groups. Finally, note that ranking blocs by their similarity to Bloc 12 does not produce the inverse similarity ranking of those same blocs to Group 9. This suggests that the nature of voting within the Labour party cannot be accurately captured by a single dimension measuring loyalty to the government.

5.1.5 Difference versus Dissent, 1992–2005. Our approach here is a complement to, rather than a substitute for, earlier work that considers "dissent" as the key behavior of interest (e.g. Kam 2001; Garner and Letki 2005). We can build on accounts such as Benedetto and Hix (2007), for example, who argue that rebellion is most likely for backbenchers who are "rejected" for promotion to higher office, or "ejected" from the front benches, having been fired for some reason or other. While the **P** matrix of bloc probabilities will certainly take the instances of rebellion into account, it will also be conditioned on instances of "difference" in less whipped votes. Moreover, the **P** matrix can be calculated for different subsets of the rolls calls—which allows a time component to be introduced to the analysis. To see how this might work, consider Figure 5 wherein each subplot shows the difference between five MPs and the median MP's closeness to the loyalist bloc for the Labour party in any given year. Nick Ainger, a long serving member from the core loyalist group provides the baseline for the loyalist comparison. Here we have extended the analysis, *mutandis mutatis*, to 1992–2005 which covers the period before Blair became leader of the Labour party and just before he stepped down (and have included party switchers). The plots may be interpreted straightforwardly: when the lines rise, the MP is becoming more likely a part of the loyalist cluster; when they fall, she is more likely to be part of a different (nonloyalist) cluster. The broken lines represent election dates. The first row is Tony Banks, a veteran left-winger who is, in fact, remarkable similar to the median MP in terms of his probability of being within the loyalist cluster over time. This is despite his promotion and then firing as sports minister in 1999. By contrast, Jeremy Corbyn, in the second plot, is very average in his behavior until just after the 1997 landslide election. He then moves in a more rebellious direction than the median Labour MP, before starting to vote with the government again just before the 2001 election. By the time of the Iraq war, in 2003, he is has seemingly abandoned the loyalists for good. In the third row, Kate Hoey is an unusual case: a loyalist and latterly a minister until 2001, she is then fired and moves from a zenith of loyalism to a grouping entirely separate from them. The loyalists do not seem to lose the Vauxhall MP for good, however, and she returns to the fold briefly in 2002. As noted above, Hoey is not necessarily in the same camp as Corbyn ideologically: for example, at least part of her distance from the average Labour member is due to her outspoken support in the Commons for fox hunting. A less nuanced story emerges for Glenda Jackson (fourth row), who entered parliament in 1992 and was gradually promoted to government roles with more responsibility. Perhaps unsurprisingly, she is no more or less loyalist than the median MP during this time. Upon leaving front bench politics in 2000 (in an effort to run for Mayor of London) she remained relatively loyal, but then moved far from the median just after the 2001 election and towards the left-wing of the Labour party. The last subplot is that of Shaun Woodward, a loyal *Conservative* MP (and thus a member of the opposition) who switched parties in 1999 and instantly became loyal to his new masters in the Labour party.

These plots are hardly conclusive, but they do begin to illustrate the extra depth researchers can obtain by modeling voting records in terms of *probabilistic clustering*. Rather than trying to laboriously code every division as whipped or unwhipped, and then to "count up" the instances of deviant behavior to be used as a dependent (or independent) variable, our approach allows analysts to directly and unavoidably compare and contrast MPs with their colleagues. This facilitates a more helpful *relative* positioning and classification label that has some of the strengths of scaling without the misleading inferences that arise with "off-the-shelf" methods. Moreover, unlike the foregoing literature on Westminster politics, we can straightforwardly obtain the relevant divisions that "probably" contribute to the profiles in the plot, without resorting to arranging every member on a crude "rebel"/"loyalist" continuum.

5.2 Model Fit

To get some sense as to how our model fits the data we employ a posterior predictive check (Gelman, Meng, and Stern 1996), that is similar in spirit to those discussed by Gelman (2004) and Buja (2004). Specifically, we generate 19 new roll-call matrices from the appropriate posterior predictive distribution. We then display image plots of these matrices in Figure 6 *along with the observed roll call matrix*. The observed roll-call matrix is put in a randomly chosen position. If it is the case that the model does a poor job of capturing observed patterns in the data then the observed roll-call matrix should be easily spotted in the 5×4 "lineup" in Figure 6. We submit that the observed data are not easily detected and thus this check provides no evidence of poor model fit.

6. DISCUSSION

The modeling strategy discussed here has much to recommend it as a purely exploratory method. It is model-based and thus allows for statements of uncertainty about all quantities of interest. Further, model-based approaches do not suffer from the need to make essentially *ad hoc* choices about a distance metric and/or how to handle missing data. However, unlike standard model-based clustering methods (Banfield and Raftery 1993) we do not fix the number of clusters *a priori*. Instead, we estimate the number of clusters and allow the uncertainty about this number to propagate naturally to the first-level model parameters of most direct interest (the vote probabilities).

From a practical perspective, our method works well on real data from the U.K. House of Commons. It allowed us to find meaningful groupings of MPs within the Labour Party. These voting blocs make substantive sense and accord well with more in-depth qualitative analyzes of the House of Commons during this period (Cowley 2002). Similarly, we demonstrated how our approach can be used to identify key divisions and conflictual issues. It is worth noting that this was accomplished with nothing more than the observed vote matrices. For this reason, our method may be of some use as a tool for more qualitative researchers whose goal is a detailed examination of key pieces of legislation and key factions within the parliamentary parties. Rather than sifting through the entirety of a parliamentary voting record, researchers can use our approach to identify key votes and groupings of MPs that are worthy of more detailed analysis.

APPENDIX: THE MCMC ALGORITHM

We begin by introducing some additional notation. Let $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ denote the unique values of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I$ (the support points of G). Let $c_i \in \{1, \ldots, K\}$ denote the cluster membership of MP i, $i = 1, \ldots, I$. Specifically, $c_i = k \iff \boldsymbol{\theta}_i = \boldsymbol{\phi}_k$.

One scan of the MCMC sampling scheme used to fit the models discussed in this paper is the following:

1. (a) Draw $[\mathbf{c}|\mathbf{Y}, \boldsymbol{\lambda}, \alpha]$
   or
   (b) Draw $[c_i|\mathbf{Y}, c_{-i}, \boldsymbol{\lambda}, \alpha]$ for $i = 1, \ldots, I$
2. Draw $[\boldsymbol{\phi}_k|\mathbf{Y}, \mathbf{c}, \boldsymbol{\lambda}, \alpha]$ for $k = 1, \ldots, K$
3. Draw $[\boldsymbol{\lambda}_j|\mathbf{Y}, \boldsymbol{\phi}, \mathbf{c}, \alpha]$ for $j = 1, \ldots, J$
4. Draw $[\alpha|\mathbf{Y}, \boldsymbol{\phi}, \mathbf{c}, \boldsymbol{\lambda}]$.

Step 1(a) above is accomplished using the merge-split sampler of Dahl (2003) while step 1(b) uses algorithm 3 of Neal (2000). Every 3rd scan attempted a merge or split while the other scans used Neal's algorithm 3 to sample $\mathbf{c}$. It should be noted that the conjugacy of the beta centering distribution to the Bernoulli sampling density makes this step quite straightforward. All that is required is the ability to evaluate a beta-binomial density and the willingness to take care of some rather tedious bookkeeping.

Given $\mathbf{Y}$, $\mathbf{c}$, $\boldsymbol{\lambda}$, and α, ϕ_{kj} follows a $\mathcal{B}\text{eta}(n_{1kj} + \lambda_{1j}, n_{0kj} + \lambda_{0j})$ distribution where n_{1kj} denotes the number of "Aye" votes on division j by MPs $i = 1, \ldots, I$ for which $c_i = k$ and n_{0kj} denotes the number of "No" votes on division j by MPs $i = 1, \ldots, I$ for which $c_i = k$.

Realizing that $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ are an iid sample from $G_0(\cdot|\boldsymbol{\lambda})$ (Escobar and West 1998, p. 12) we can write the full conditional in step 3 as

$$p(\boldsymbol{\lambda}_j) \prod_{k=1}^{K} g_0(\boldsymbol{\phi}_{kj}|\boldsymbol{\lambda}_j)$$

where $p(\boldsymbol{\lambda}_j)$ is the (possibly degenerate) prior density for $\boldsymbol{\lambda}_j$. We use the univariate slice sampling algorithm of Neal (2003) to sample from this full conditional.

Finally, given the gamma prior we adopt for α, it is possible to use the data augmentation approach of Escobar and West (1995) (see also Escobar and West 1998) to sample α. This is the approach used in this paper.

[Received February 2007. Revised June 2009.]

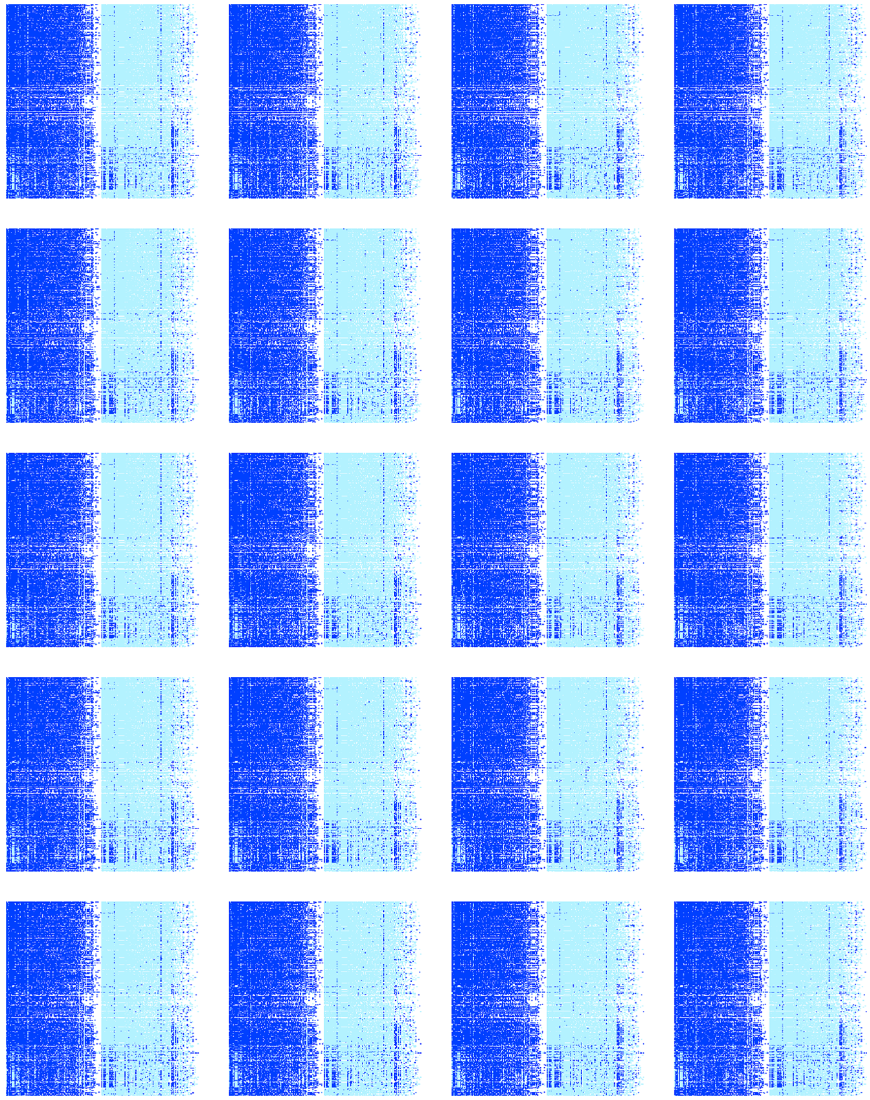

Figure 6. Posterior predictive replications along with observed data. The rows have been organized by group membership and the columns have been sorted first by observed majority “No” versus majority “Aye” and then by the number of nonmissing observed votes within each of these two groupings. Dark pixels represent “No” votes, light pixels “Aye” votes, and white pixels missing values. The observed data are in fourth row from top and second column from left. A color version of this figure is available in the electronic version of this article.

REFERENCES

Adonis, A. (1990), *Parliament Today*, Manchester, NY: Manchester University Press. [1]

Antoniak, C. E. (1974), “Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems,” *The Annals of Statistics*, 2, 1152–1174. [4]

Banfield, J. D., and Raftery, A. E. (1993), “Model-Based Gaussian and Non-Gaussian Clustering,” *Biometrics*, 49, 803–821. [9]

Benedetto, G., and Hix, S. (2007), “The Rejected, the Ejected and the Dejected: Explaining Government Rebels in the 2001–2005 British House of Commons,” *Comparative Political Studies*, 40 (7), 755–781. [8]

Berrington, H., and Hague, R. (1998), “Europe, Thatcherism, and Traditionalism: Opinion, Rebellion and the Maastricht Treaty in the Backbench Conservative Party, 1992–94,” *West European Politics*, 21 (1), 44–71. [1]

Birnbaum, A. (1968), “Some Latent Trait Models and Their Use in Inferring an Examinee’s Ability,” in *Statistical Theories of Mental Test Scores*, eds. F. Lord and M. Novick, Reading, MA: Addison-Wesley. [2]

Blackwell, D., and MacQueen, J. B. (1973), “Ferguson Distributions via Pólya Urn Schemes,” *The Annals of Statistics*, 1, 253–255. [4]

Buja, A. (2004), Discussion of “Exploratory Data Analysis for Complex Models,” by A. Gelman, *Journal of Computational and Graphical Statistics*, 13 (4), 780–784. [9]

Clinton, J., Jackman, S., and Rivers, D. (2004), “The Statistical Analysis of Roll Call Data,” *American Political Science Review*, 98 (2), 355–370. [1-3]

Cowley, P. (2002), *Revolts and Rebellions: Parliamentary Voting Under Blair*, London: Politico’s Publishing. [1,2,6,9]

——— (2005), *The Rebels: How Blair Mislaid His Majority*, London: Politico’s Publishing. [6]

Cowley, P., and Garry, J. (1998), “The British Conservative Party and Europe: The Choosing of John Major,” *British Journal of Political Science*, 28, 473–499. [1]

Dahl, D. B. (2003), “An Improved Merge-Split Sampler for Conjugate Dirichlet Process Mixture Models,” Technical Report 1086, University of Wisconsin, Madison, Dept. of Statistics. [4,9]

Desposato, S. (2003), “Comparing Group and Subgroup Cohesion Scores: A Nonparametric Method With an Application to Brazil,” *Political Analysis*, 11, 275–288. [1]

Downs, A. (1957), *An Economic Theory of Democracy*, New York: Harper. [2]

Dunleavy, P. (1993), “The Political Parties,” in *Developments in British Politics*, Vol. 5, eds. P. Dunleavy, A. Gamble, I. Holliday, and G. Peele, Palgrave Macmillan. [6]

Escobar, M. D., and West, M. (1995), “Bayesian Density Estimation and Inference Using Mixtures,” *Journal of the American Statistical Association*, 90, 577–588. [4,9]

——— (1998), “Computing Nonparametric Hiearchical Models,” in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. D. Dey, P. Müller, and D. Sinha, New York: Springer. [4,9]

Ferguson, T. S. (1973), “A Bayesian Analysis of Some Nonparametric Problems,” *The Annals of Statistics*, 1, 209–230. [4]

——— (1983), "Bayesian Density Estimation by Mixtures of Normal Distributions," in *Recent Advances in Statistics*, eds. H. Rizvi and J. Rustagi, New York: Academic Press. [4]

Firth, D., and Spirling, A. (2005), *tapiR: Tools for Accessing UK Parliamentary Information in R*, R package version 0.7-2, available at *http://www.warwick.ac.uk/go/tapir*, *http://www.warwick.ac.uk/go/dfirth*. [4]

Fraley, C., and Raftery, A. E. (2002), "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, 97, 611–631. [3]

Garner, C., and Letki, N. (2005), "Party Structure and Backbench Dissent in the Canadian and British Parliaments," *Canadian Journal of Political Science*, 38, 463–482. [8]

Gelman, A. (2004), "Exploratory Data Analysis for Complex Models" (with discussion), *Journal of Computational and Graphical Statistics*, 13 (4), 755–779. [9]

Gelman, A., Meng, X.-L., and Stern, H. (1996), "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies" (with discussion), *Statistica Sinica*, 6, 733–807. [9]

Hahsler, M., Buchta, C., and Hornik, K. (2008), *seriation: Infrastructure for Seriation*, R package version 1.0-0, available at *http://R-Forge.R-project.org/projects/seriation/*. [6]

Hartigan, J. (1975), *Clustering Algorithms*, New York: Wiley. [3]

——— (2000), "Bloc Voting in the United States Senate," *Journal of Classification*, 17, 29–49. [3,4]

Heckman, J. J., and Snyder, J. M. (1997), "Linear Probability Models of the Demand for Attributes With an Empirical Application to Estimating the Preferences of Legislators," *RAND Journal of Economics*, 28, S142–S189. [1,2]

Hix, S. (2002), "Parliamentary Behavior With Two Principals Preferences: Parties, and Voting in the European Parliament," *American Journal of Political Science*, 46 (3), 688–698. [1]

Jennings, I. (1969), *Parliament* (2nd ed.), London: Cambridge University Press. [2]

Kam, C. (2001), "Do Ideological Preferences Explain Parliamentary Behavior? Evidence From Britain and Canada," *Journal of Legislative Studies*, 7, 98–126. [8]

Lord, F. M. (1980), *Applications of Item Response Theory to Practical Testing Problems*, Mahwah, NJ: Lawrence Erlbaum Associates. [2]

Lowell, L. A. (1901), "The Influence of Party Upon Legislation in England and America," *Annual Report of the American Historical Association*, 1, 319–541. [2]

MacQueen, J. (1967), "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, Berkeley, CA: University of California Press. [3]

MacRae, D. (1970), *Issues and Parties in Legislative Voting: Methods of Statistical Analysis*, New York: Harper & Row. [3]

Martin, A., and Quinn, K. (2002), "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999," *Political Analysis*, 10 (2), 134–153. [2]

Neal, R. M. (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265. [4,9]

——— (2003), "Slice Sampling" (with discussion), *The Annals of Statistics*, 31, 158–182. [9]

Norton, P. (1975), *Dissension in the House of Commons: Intra-Party Dissent in the House of Commons' Division Lobbies, 1945–1974*, London: Macmillan. [6]

——— (1980), *Dissension in the House of Commons, 1974–1979*, Oxford: Clarendon Press. [6]

Poole, K. (2005), *Spatial Models of Parliamentary Voting*, Cambridge: Cambridge University Press. [2]

Poole, K. and Rosenthal, H. (1997), *Congress: A Political Economic History*, New York: Oxford University Press. [1-3]

Poole, K. T. (2000), "Non-Parametric Unfolding of Binary Choice Data," *Political Analysis*, 8, 211–237. [3]

Poole, K. T., and Rosenthal, H. (1985), "A Spatial Model of Legislative Roll Call Analysis," *American Journal of Political Science*, 29 (2), 357–384. [1,2]

——— (1991), "Patterns of Congresional Voting," *American Journal of Political Science*, 35, 228–278. [1,2]

Rasch, G. (1961), *Probabilistic Models for Some Intelligence and Attainment Tests*, Copenhagen: Danish Institute for Educational Research. [2]

Rice, S. (1928), *Quantitative Methods in Politics*, New York: Knopf. [2]

Rosenthal, H., and Voeten, E. (2004), "Analyzing Roll Calls With Perfect Spatial Voting: France 1946–1958," *American Journal of Political Science*, 48 (3), 620–632. [1]

Shaw, E. (1988), *Discipline and Discord in the Labour Party: The Politics of Managerial Control in the Labour Party, 1951–87*, New York: Palgrave Macmillan. [1]

Silk, P. (1987), *How Parliament Works*, London: Longman. [2]

Spirling, A., and McLean, I. (2007), "UK OC OK? Interpreting Optimal Classification Scores for the United Kingdom House of Commons," *Political Analysis*, 15 (1), 85–96. [1,2]