Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation

Jeff GILL and George CASELLA

A generalized linear mixed model, ordered probit, is used to estimate levels of stress in presidential political appointees as a means of understanding their surprisingly short tenures. A Bayesian approach is developed, where the random effects are modeled with a Dirichlet process mixture prior, allowing for useful incorporation of prior information, but retaining some vagueness in the form of the prior. Applications of Bayesian models in the social sciences are typically done with ''uninformative'' priors, although some use of informed versions exists. There has been disagreement over this, and our approach may be a step in the direction of satisfying both camps. We give a detailed description of the data, show how to implement the model, and describe some interesting conclusions. The model utilizing a nonparametric prior fits better and reveals more information in the data than standard approaches.

KEY WORDS: Bayesian nonparametrics; Dirichlet process; Gibbs sampler.

1. MOTIVATION

An unanswered question in the executive appointment literature concerns the length of stay of political appointees in the United States government. These positions include not just members of the cabinet, but also chief executives of independent agencies and regulatory commissions, and judges/justices. The nonjudicial titles considered here include Secretary, Undersecretary, Assistant Secretary, Chairperson, and Member of the Board or Commission. Presidential and bureaucratic scholars argue that political appointments represent the single greatest source of presidential influence over the bureaucracy, yet these executives tend to remain on the job for only short periods (2 years on average). So by the time they master the complex responsibilities of public management at a high level, they are inclined to leave government. How is it then possible for political appointees to be a primary source of presidential power while simultaneously serving so briefly?

Early studies generally concluded that presidents achieve little control over bureaucratic outcomes through the appointment process (Fenno 1959; Heclo 1977; Kaufman 1981; Noll 1971). The attributed reasons include strongly organized influence by interested groups (Olson 1965) and close supervision by Congressional committees (Freeman 1965). Scholars of this era used the metaphor of the ''iron triangle'' to describe the three-way, long-term relationship between an administrative agency, the relevant Congressional committee, and regulated industry lobbyists that negotiated policy and budgets in a relatively closed system that tended to exclude the public and the President (Lowi 1969). Other early work observed that Presidents often worry that their appointees may ''go native'' once in place as they align themselves with the goals and culture of their agency at the expense of the President's priorities (Ingraham & Ban 1986; Wilson 1989). Other work of the time focused on a more general macropolitical explanation for agency outcomes that admits influence by Congress as a whole, the President, and the courts (Redford 1969; Rourke 1984; Gormley 1986). The iron triangle analogy is now considered too limited as many policy subsystems have changing participants and may also have substantial visibility with the public (Thurber 1991).

More recently, the relationship between the President and his appointees is described and modeled with *principal-agent theory*. Originally derived from contract theory in the fields of economics and finance, principal-agent theory was employed to explain a more fluid interaction between political principals and agency executives. Principals seek to control their bureaucratic agents because over time the interests of the bureaucracy and those of the principals who created it may diverge. Even if there is no particular policy disagreement, agents are likely to shirk or to produce outputs at a higher cost than is required or to produce a lower level of outputs than is desired by the principal due to information asymmetry (see Niskanen 1971; Miller & Moe 1983).

So if political appointees enjoy some independence from their most important principle and are able to build relationships with other political actors, why would so many leave government early? One obvious, but as-yet unsupported, reason is that the challenges and frustrations of the job are so sufficiently great that they overwhelm loyalty to the President, careerism, or commitment to public service. We seek evidence that the pressures of running large, complex government agencies is a contributing factor in early departures. This is an important question because the experience, professionalism, and resources that political executives have, or do not have, greatly affect agency performance and the delivery of public goods to citizens. Consider the wide range of agency effectiveness in dealing with recent events, from New York City's response to the attacks of 9/11 to FEMA's management of the Hurricane Katrina aftermath.

The remainder of the article is structured as follows. First, we introduce a unique and interesting dataset in political science that helps us understand the behavior of political executives. Next we describe our modeling approach in Section 2.1, including our approach to using nonparametric priors to draw added information from samples. Here we also discuss connections to related work applying item response models in political science. In Section 3 we develop the details of the ordered probit model with nonparametric priors on the random

Jeff Gill is Professor, Center for Applied Statistics, Washington University, One Brookings Drive, Seigle Hall LL-085, St. Louis, MO 63130 (E-mail: *jgill@wustl.edu*). George Casella is Professor, Department of Statistics, University of Florida, 102 Griffin-Floyd Hall, Gainesville, FL 32611 (E-mail: *casella@stat.ufl.edu*). This research is supported by National Science Foundation grants DMS-0631632 and SES-0631588. We thank Larry Bartels, Persi Diaconis, Hani Doss, Andrew Gelman, Sandy Gordon, Jim Hobert, Kosuke Imai, Minjung Kyung, Scotland Leman, Nolan McCarty, Michael Peress, Curt Signorino, and Rick Waterman for helpful comments.

Journal of the American Statistical Association
June 2009, Vol. 104, No. 486, Applications and Case Studies
DOI 10.1198/jasa.2009.0039

effects, and also describe the needed MCMC algorithm for fitting. Section 4 describes the data in great detail, gives specifics of the fit of the model, and discusses the inferences that can be made. There is a concluding discussion, and an Appendix containing some technical details and further description of the coding of the data.

2. DATA AND MODEL BACKGROUND

Useful statistical data on political elites is notoriously hard to come by. Politicians, senior government officials, judges, and lobbyists rarely cooperate with academic survey efforts, particularly when the questions are personal. A notable exception is the dataset provided by Mackenzie and Light (Interuniversity Consortium for Political and Social Research (ICPSR) Study Number 8458, Spring 1987, further described in Section 4), which originates from interviews with *every* presidential appointee to a full-time position requiring Senate confirmation from November 1964 through December 1984 (1,528 nonjudicial individuals altogether). Several steps have been taken to preserve the anonymity of the respondents, including removal of the variables recording: year and month of confirmation, year and month of leaving, birth year, sex, race, geographic location of agency, and the senate committee with jurisdiction over the appointment. Furthermore, Mackenzie and Light provide only a random sample of 532 from the full set of responses to further improve privacy because there is now no assurance that a selected public official is in the dataset. Whereas such steps are necessary to ensure full cooperation, they clearly make the analysts' work more difficult.

The restriction that is the least worrisome here is the reduction of the data size to 532 cases. Unless a model specification is particularly demanding, this sample provides a comfortable number of degrees of freedom. The dates of service for each respondent would have provided useful historical reference, particularly with regard to understanding the relationship between the President at the time and the concurrent Congress. The reduced dataset retains the name of the nominating President, the respondent's self-identified ideology on a five-point scale, and the respondent's self identified party strength on a five-point scale (the last variable turns out to be more useful here). The greatest modeling challenge, however, comes from the omission of agency name or type from the data. These executives collectively run highly diverse agencies and therefore face radically different challenges. In addition to the variability in responsibilities, these agencies perform vastly different functions including: foreign policy, use of military forces, economic regulation, entitlement management, determination of agricultural policy, and so on. Failing to account for this obvious diversity means that some systematic component of the data falls to an error term, exacerbating efforts to find parsimonious models with good fit. Lacking direct covariate information about this effect, we seek here to find help in the data itself by specifying a nonparametric prior that reflects information in the data to help account for underlying structure in the context of the model. Thus this heterogeneity is actually the motivation for our methodological approach and previous attempts to model these data have struggled with the known, but unmeasured, heterogeneity of agency assignment.

Such confidentiality limitations should not obscure the importance of the Mackenzie and Light dataset, which alone provides direct attitudinal measures on the executives' relationship with Congress, the President, and other bureaucratic actors. All other datasets available to academic researchers either provide purely descriptive information, or are small-n qualitative studies. Thus, whereas we would prefer data on more recent appointees, the opportunity to understand the broad views of high-level government executives in the late 20th Century is compelling. This is not merely historical research as the phenomenon of interest, short tenure of service, is basically unchanged to date. Furthermore, there is absolutely no evidence that public leadership matters less in today's bureaucracy than in the period of study (see, in particular, Wolf (2005)).

We should note further that these data are representative of the additional problems that social science analysts generally face. Unlike data that originate from designed experiments, observation of physical phenomena, or microarrays, data that measure interactive human behavior present unique challenges because the subjects are strategic, intelligent, and occasionally even deceptive. Thus covariates are usually correlated, making direct causal inferences more difficult. We recognize this here with our approach and carefully select explanatory variables that represent two general forces on the individuals: personal characteristics that determine preparedness and disposition toward government service, and bureaucratic forces that all of these individuals face when taking high level positions in Washington.

There are several possible outcome variables of interest available from these data, but we use the question on *stress* with the idea that it is a surrogate measure of self-perceived effectiveness and job satisfaction. This question asks to what degree, as a full-time, Senate-confirmed presidential appointee, do the demands of the position induce stress compared with previously held positions, measured as a five-point Likert scale from "not stressful at all" to "very stressful." The claim here is that the stress of high-level public service is a reason for short tenures and we therefore model stress as the outcome variable in a Bayesian generalized linear mixed model, using an ordered probit link, described in Section 3.

Stress is a difficult psychological phenomenon to measure and varying the language of query can dramatically change results (Hurrell, Nelson, and Simmons 1998). A core part of the literature uses the linkage between recent life-events and reported stress through survey research, such as the Holmes & Rahe (1967) Social Adjustment Rating Scale and its later variations. Another traditional approach is based on selecting individuals who have endured a particular trauma with the objective of understanding the characteristics of the resulting stress. These studies are often focused on the linkage between the particular trauma, such as rape (Burgess and Holmstrom 1974, 1979) or the loss of a close family member (Lehman, Wortman, and Williams 1984; Vachon, Lyall, Rogers, Freedman-Letofsky, and Freeman 1980; Vachon et al. 1982), and subsequent stress.

Occupational stress has interested both economists and cognitive psychologists because it appears to be both widespread and deleterious to productivity (Ivancevich, Matteson, Freedman, and Phillips 1990). Yet, the literature has struggled with defining stress, using notions such as: personal anxiety, poor job satisfaction, bad psychological "health," and disequilibrium with the surrounding environment (Hart, Wearing, and Headey 1993; Newton 1989; Pratt and Barling 1988). A psychological definition of stress that is relevant here is that of Lazarus (1991). This definition posits occupational stress as a set of transactions between the individual and potential "stressors" existing in the workplace. Stress, in the form of anxiety or anger, is raised when these stressors present threats that exceed the individual's coping resources. In our case, political executives consider resignation when the frequency and intensity of these events exceeds the psychological rewards provided by high-level government service. We now consider the best way to model stress in a specific political context.

2.1 Modeling Approach

What is the best way to specify prior information that social scientists possess in a statistical modeling context? This is a seldom-addressed question in the social sciences because most researchers developing Bayesian statistical models simply apply flat prior distributions or some other highly uninformed variety. These fields have for the most part ignored the opportunity to include rich troves of prior information into statistical models. On the other end of the spectrum are priors with strong claims about some phenomena of interest where considerable justification is involved to convince readers that this "subjective" information should be included. We argue here for nonparametric priors, which exist in the middle of this spectrum because they retain some researcher intuition but allow the data to drive the analysis.

A strongly informed approach to prior specifications contrasts sharply with nearly all Bayesian work to date in the social sciences (Quinn, Martin, and Whitford 1999; Western 1998; Jackman 2000a; Hill and Kriesi 2001; Smith 1999; Schweinberger and Snijders 2003; Rubin and Schenker 1987), although one recent study (Jackman 2001) demonstrated a particular need for informed (but not elicited) priors. Recent efforts have instead focused on applying simulation tools from Bayesian statistics (i.e., Markov chain Monte Carlo) to solve previously intractable problems. This computationalist perspective mostly avoids the specification of deeply informed priors in favor of diffuse forms (Jackman 2000b; Martin and Quinn 2002). Such priors have useful purposes (particularly in dealing with so-called nuisance parameters), but they do not fully exploit Bayesian capabilities. In essence, the typical approach is Bayesian in that it allows probabilistic discussion of results, but only to a limited extent, so it avoids making decisions about priors. We will, however, use the approach of flat priors to compare with our developed method.

A small number of authors have argued for the use of directly informed priors in the social sciences (Leamer 1972; Western 1996; Bartels 1996; Berk, Western, and Weiss 1995), but we believe that the greater majority of social scientists are more comfortable with vague prior information. Thus, one of our objectives in this article is to explore how to best specify *semi-informed* prior information in a Bayesian statistical model, and to incorporate nonparametric priors to improve the analysis of social and behavioral data (which often have higher measurement error and lower precision than that obtained in other scientific fields). By *improve* we mean using information in the sample, which does not become expressed in a standard likelihood function, to improve the fit of the subsequent model. We will do this by allowing the data to help create a logical grouping in the data.

2.2 Nonparametric Priors

Bayesian nonparametric priors, based on the Dirichlet process, were introduced by Ramsey (1972), Ferguson (1973), and Antoniak (1974), but not fully developed until the advent of better computing resources for estimation. A recent application in economics is given by Hirano (2002). These distributions can be made conditional on additional parameterizations (as done in Escobar and West (1995)), or incorporated into broader estimation challenges (see the application to ecological inference in Imai, Lu, and Strauss (2008)), and thus the models are hierarchical in two senses. Subsequent realizations of the Dirichlet process are discrete (with probability one), even given support over the full real line, and are thus treated like countably infinite mixtures (Sethuraman 1994).

What can nonparametric priors add to the emerging Bayesian paradigm in the social sciences? Consider the question of modeling dichotomous individual choices, Y_i, such as turning out to vote, voting for a specific candidate, joining a social group, discontinuing education, and so on. The most common "regression-style" modeling specification is to assume that an underlying smooth utility curve dictates such preferences, providing the unobserved, but estimated threshold, $\theta \in [0, 1]$. The individual's threshold along this curve then determines the zero or one outcome conditional on an additive right-hand side specification, $\mathbf{X}\boldsymbol{\beta}$. Realistically, we should treat this threshold differently for each individual, but we can apply the reasonable Bayesian approach of assuming that these are different thresholds but still generated from a single distribution G, which is itself conditional on a parameter α, thus $E[nG(\theta|\mathbf{X}\boldsymbol{\beta},\alpha)]$ is the expected number of affirmative outcomes. Suppose there were structures in the data such as unexplained clustering effects, unit heterogeneity, autocorrelation, or missingness that cast doubt on the notion of G as a single model. Note that this can happen in a Bayesian or non-Bayesian setting, the difference being the distributional or deterministic interpretation of θ. The choice of G is unknown by the researcher but determined by custom or intuition. We suggest, instead, a nonparametric Bayesian approach that draws θ from a mixture of appropriate prior distributions conditional on data and parameters.

2.3 Connections

Political Science and Political Methodology have benefited from renewed attention to issues of measurement for political and social data. The discipline has long wrestled with defining *democracy* (Gaubatz 1996), *reciprocity* (Keohane 1986), *war* (Dinstein 2005), and even something as seemingly simple as *preferences* (Laitin and Wildavsky 1988). Interest has increased substantially in improving the measurement of *ideal points*: the most-preferred coordinate for a political decision-maker on some (usually) multidimensional Euclidean space defining policy alternatives. This literature started with a focus on legislators as well as citizens/voters (Aldrich and McKelvey 1977; Palfrey and Poole 1987; Poole and Rosenthal 1985, 1997; Poole, Rosenthal and Koford 1991; Jacoby 1994), and has recently become strongly Bayesian in orientation (Jackman 2001; Martin and Quinn 2002; Clinton, Jackman, and Rivers 2004). The core problem is not unlike that addressed here: the data arrive as sets of decisions (i.e., roll call or votes or judicial decisions) that partly reveal underlying preferences because the choice-set (bills, cases, candidates, etc.) are exogenously determined and discrete. Thus the task is to specify a statistical model that estimates ideal points on some latent continuous measure that accounts for the choices presented as well as the characteristics of the subjects making choices. Item response theory (IRT) models originating from education psychology are an example of such an approach. A common problem in this area is the plethora of parameters that comes with n individuals (ideal points), k bills/cases/questions (proposals), and d dimensions, for a total of $nd + k(d + 1)$. The Bayesian approach works better here because, rather than estimating ideal point parameters and proposal parameters distinctly and serially (Bock and Aitken 1981; Hambleton, Rogers, and Swaminathan 1991), it treats all unknowns identically and readily furnishes posterior distributions via Bayesian stochastic simulation (Baker 2004).

Our problem is both similar to and different from this topic in political methodology. We *know* that there is a latent variable that, to a great extent, affects how the explanatory variables relate to the outcome of stress. We also know that there is information in the data that can inform this unseen measure. In this regard the objective is similar to IRT/ideal-point analysis because a connection is required between the visible and the underlying. However, our objective is to categorize cases (individual public executives) by estimating agency subclusters: data-determined "bins" that are greater in number than the unknown binning by agency, but serve to make the model fit better. That is, these nonparametric subcluster assignments are not the substantive clusters, but could be if further information was available (the actual number of clusters and additional relevant covariates) that allowed the researcher to group the subclusters into clusters. This is actually a very difficult secondary process, which we save for future work.

3. NONPARAMETRIC PRIORS APPLIED TO BAYESIAN MODELS FOR ORDINAL OUTCOMES

We now describe a nonparametric model with Dirichlet process priors for problematic data that may be skewed, multimodal, or possess other characteristics that make standard parametric assumptions unrealistic. Specifically, the model starts with $Y_1, Y_2, \ldots, Y_n$, which are assumed to be drawn from a mixture of distributions denoted $P(\psi)$ where the distribution of ψ is provided by G. The prior on G, $\mathcal{DP}$, is a mixture of Dirichlet processes. Although the model can be defined in great generality, we are most concerned with the case in which the Y_i are manifestations of a latent class variable and can be modeled from an ordered logit or probit model. Whereas the model described in the next paragraph corresponds to the problem at hand, we can easily extend the nonparametric prior approach to a wide range of Bayesian generalized linear models (Dey, Ghosh, and Mallick 2000).

The standard ordered probit model assumes first that there is a multinomial selection process where we observe iid Y_i according to

$$Y_i \sim \text{Multinomial}(1, (p_1, p_2, \ldots, p_C)), \quad i = 1, \ldots n \tag{1}$$

where $\sum_j p_j = 1$, and $Y_i = (y_{i1}, \ldots, y_{iC})$ is a $C \times 1$ vector with a 1 in one position and 0 elsewhere. It may be the case that clustering of these Y_i values calls into question the independence assumption here, and this is an issue that we address with our model. The p_j are ordered by a probit model on $\Re$ for the random U_i.

$$p_j = P(\theta_{j-1} \le U_i \le \theta_j) \tag{2}$$

where these "cutpoints" or "thresholds" between categories have the property that $-\infty = \theta_0 < \theta_1 < \cdots < \theta_C = \infty$, and

$$U_i \sim N(X_i\beta + \psi_i, \sigma^2) \tag{3}$$

where X_i are covariates associated with the *ith* observation, β is the coefficient vector, and ψ denotes a random effect to account for subject-specific deviation from the underlying model. Consider as a comparative illustration a standard Bayesian ordered probit model without the random effect term and putting flat, uninformed priors on all unknown parameters. The results are given in Table 1 where we observe that the fit is not totally satisfactory. Six of the 11 estimates for the effects of explanatory variables are reliable at conventional levels (a 95% Highest Posterior Density (HPD) interval bounded away from zero), and the set of five estimates of the covariates of greatest theoretical interest, those involving bureaucratic politics with underlings, are particularly disappointing. Furthermore, simple manipulation of the prior specifications show substantial sensitivity moving away from diffuse uniform distributions. We therefore seek to improve on this analysis with Dirichlet process priors, which pay more attention to the data. T1

Whereas the random effect is indexed by i, it retains the usual hierarchical interpretation because individuals will share ψ values by being grouped together in a subcluster: the ψ_i values are not necessarily unique. Now the data are

Table 1. Simple Bayesian posterior for survey of political executives

Posterior	Mean	95% HPD Interval
`Government Experience`	0.121	[−0.068: 0.310]
`Ideology`	0.077	[−0.020: 0.174]
`Committee Relationship`	−0.178	[−0.288: −0.069]
`Career.Exec-Compet`	−0.175	[−0.326: −0.023]
`Career.Exec-Liaison/Bur`	0.105	[−0.011: 0.221]
`Career.Exec-Liaison/Cong`	−0.029	[−0.119: 0.060]
`Career.Exec-Day2day`	−0.153	[−0.290: −0.017]
`Career.Exec-Diff`	0.114	[−0.012: 0.241]
`Confirmation Preparation`	−0.315	[−0.570: −0.061]
`Hours/Week`	0.446	[0.359: 0.534]
`President Orientation`	−0.338	[−0.593: −0.082]
Cutpoints: (None) (Little)	−0.792	[−1.627: 0.042]
(Little) (Some)	−0.270	[−1.097: 0.557]
(Some) (Significant)	0.361	[−0.465: 1.186]
(Significant) (Extreme)	1.530	[0.696: 2.365]

hierarchical with respect to these latent subclusters, which is the structure in the Dirichlet process mixing prior. These subclusters represent characteristics of agency type, and provide improved model fit by recovering an estimate of the key missing explanatory variable of agency assignment for the political executive. Therefore we address the challenge of nonparametrically recovering some information about the key missing covariate of federal agency assignment per individual that confidentiality restrictions forbid us to include directly.

The U_i in (3) are, in fact, unobservable continuous random variables, and we could specify the model without them, that is, from both (2) and (3),

$$p_j = \Phi\left(\frac{\theta_j - X_i\beta - \psi_i}{\sigma}\right) - \Phi\left(\frac{\theta_{j-1} - X_i\beta - \psi_i}{\sigma}\right), \quad (4)$$

where Φ is the standard normal cumulative distribution function (CDF). However, using the U_i will sometimes result in easier calculations with the Gibbs sampler, as shown later, so we use this latent variable model. The setup is equivalent to that seen in standard econometric texts (e.g., Greene 2003, Chap. 21), except for the inclusion of ψ.

3.1 Background on Mixtures and Dirichlet Forms

A standard mixture distribution for continuous y has the form

$$f(y|g) = \int f(y|\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad (5)$$

where $g(\boldsymbol{\theta})$ could have a parametric form conditional on parameters, $g(\boldsymbol{\theta}|\xi)$, or a nonparametric form, G, suggested by the data. Ferguson (1973) and Antoniak (1974) used the Dirichlet process prior as a way to create the nonparametric G. First we start with a base distribution, G_0, which forms the expected value of the distributions, and a precision parameter m, reflecting the concentration around G_0 for the other distributions. Ferguson and Antoniak show that for *any* finite partition of the parameter space, which they label $(B_1, \ldots, B_K)$, the *Dirichlet process prior* constitutes a random probability measure through a Dirichlet distribution that uses the parameter vector $(mG_0(B_1), \ldots, mG_0(B_K))$ to produce a vector of probabilities $(G(B_1), \ldots, G(B_K))$, where $G \in \mathcal{P}$. Thus, using G we can write the mixture (5) in the hierarchical form

$$\begin{aligned} Y_i|\boldsymbol{\theta}_i &\sim f_i(y_i|\boldsymbol{\theta}_i) \\ \boldsymbol{\theta}_i|G &\sim G(\boldsymbol{\theta}) \\ G|m, G_0 &\sim \mathcal{DP}(mG_0). \end{aligned}$$

The final task to complete the Bayesian model is to assign a prior value for m and parameters for G_0.

3.2 Dirichlet Mixture Specification for the Ordered Choice Model

There is usually no compelling reason to force a particular structure or distribution on the random effect ψ in (3) (Doss 2007). Therefore, we make our model semiparametric and assume that ψ is an observation from the described Dirichlet process,

$$\psi_i \sim G \qquad G \sim \mathcal{DP}_{mG_{\mu,\tau^2}} \quad (6)$$

where G_{μ,τ^2} is the *base measure* with parameters μ, τ^2, and m is the *concentration parameter*. Thus, ψ is modeled to come from a distribution that sits in a neighborhood of G_{μ,τ^2}, with the size of the neighborhood being controlled by m.

The model specified by (1)–(6) is, in fact, a classical semiparametric random effects model, and with further Bayesian modeling of the parameters lends itself to a Gibbs sampler. In particular, Neal (2000) looks into convergence properties of these models and, as a first step in fitting, we use a variation of one of Neal's samplers adapted to the ordered probit model, which has not been considered before. In fact, the presence of the Dirichlet term makes the use of the Gibbs sampler somewhat complicated in nonconjugate situations, which is what we have here. However, the modeling of the probit through the latent variables U_i partially alleviates such problems (Albert and Chib 1993).

To complete the Bayesian modeling of (1)-(6) we add the following priors:

$$\beta \sim N(\beta_0, \sigma^2_\beta) \quad \mu \sim N(0, d\tau^2) \quad \frac{1}{\tau^2} \sim \text{Gamma}(a, b). \quad (7)$$

We note that, in practice, we set $\sigma^2_\beta = \infty$, resulting in a flat prior on β, as a flat prior at this point in the model simply results in least-squares-type estimation of the covariates. This decision appears to have little influence on the resulting posteriors and allows the ψ specification sufficient latitude to draw nonparametric information from the data. The parameters in the priors on μ and τ^2 are chosen to make the priors sufficiently diffuse to allow the random effect to do its work.

The choice of prior mean zero for ψ does not lose generality, as the $X_i\beta$ term in (4) locates the distribution. Generation of the parameters in the Gibbs sampler is somewhat involved, and is described in detail in Appendix A.

4. THE PRESIDENTIAL APPOINTEE DATA

As noted before, the data we use come from interviews with every presidential appointee to a full-time position requiring Senate confirmation from November 1964 through December 1984 (collected by Mackenzie and Light, ICPSR Study Number 8458, Spring 1987). The biography component of the data has 1,528 cases, and the survey component, which we use, has 532 cases. Details on coding of the variables from the Mackenzie and Light study are described in the appended **Data Notes** section (Appendix B) and more details are given in the analysis by Gill and Waterman (2004). The survey queries various aspects of the Senate confirmation process, acclimation to running an agency or program, and relationships with other functions of government. A Senate confirmed administrative executive reports hierarchically to the President but budgetarily (and often on policy matters as well) ultimately to Congress.

4.1 Data Details

We focus on *stress* as a surrogate measure of self-perceived effectiveness and job satisfaction, where `Stress`, is measured as a five-point Likert scale from ''not stressful at all'' to ''very stressful.'' The claim here is that the stress of public service at this level is a reason for short tenures and we therefore model stress as the outcome variable because length of service per case is embargoed from the dataset as a possible identifier (Gill and Waterman 2004). Thus the stress variable is treated as a self-assessed catch-all measure of effectiveness in the bureaucratic and political environment, degree of satisfied or unsatisfied demands from Congress and the White House, and the overall utility received from senior government service. Political executives serve only 2 years on average and therefore do not generally integrate themselves into the culture of their agency (Heclo 1988). Therefore it is reasonable to equate *stress*, *personal sense of accomplishment*, and *bureaucratic effectiveness* for the respondents.

The executives obviously come to their appointed positions with varying governmental experience that bears upon their probability of success at the bureaucratic level. So we include a dichotomous explanatory variable indicating whether they come from previous government employment at the state or federal level: `Government Experience`. This job experience variable is also a surrogate for political information; presumably the more government experience an appointee has the more *political* knowledge she has. Consequently, greater levels of political experience/information should be related to lower levels of stress.

One major consideration in how executives perceive their role in government service and their relationship with other governmental institutions is political ideology (Keiser and Soss 1998; Mazmanian and Sabatier 1980; Ringquist 1995). For all but a 2-year period (the Senate from 1981–1983), the Democrats controlled both houses of Congress. This means that Republican appointees may have a more contentious relationship with pertinent oversight committees. So we include the variable `Ideology` measured at five points from very liberal (1) to very conservative (5). Furthermore, there is a common notion that the candidate's relationship with the confirming Senate committee can affect subsequent interactions between their agency and Congress in general. We therefore include a variable that describes the executives' view on their relationship with this committee before and during the confirmation hearings. The variable `Committee Relationship` is measured along a five-point scale from hostile to friendly (ascending).

One variable measures whether the candidate received pre-hearing preparation from the White House by any of: the personnel office, the Counsel to the President, the general staff (or even within the relevant agency). This variable, `Confirmation Preparation`, is dichotomously measured in the obvious direction. If the White House is actively engaged in preparing the candidate then it may be a signal that they perceive the administration as the primary principal for this agent. This also is another way of measuring political information that differs from `Government Experience` that instead focuses on cooperation with a particular political principal.

Four variables directly address the efficacy and relevancy of our generalized principal-agent configurations. The first is `Career.Exec-Compet`, which is the appointee's assessment of the overall competence of his or her career executive direct reportees (measured from low to high). Certainly one would expect the impact of this assessment to affect both the working relationship and the sense of effectiveness between political and career executives. The variable `President Orientation` indicates whether the executive identified the need for training on learning how the president's priorities are expressed through agency policy-making. Each executive was also asked to what extent senior career employees (i.e., direct reports and those perhaps one level below) were helpful in dealing with other parts of the bureaucracy (`Career.Exec-Liaison/Bur`) and with Congress (`Career.Exec-Liaison/Cong`). Both of these measures deal with the relationship of bureaucratic expertise to outside actors. The first of these measures is a measure of comfort for subordinate's sharing or using bureaucratic expertise with other agencies of government, and hence allows us to examine what happens to information in a multiple agent model. The second measure examines the relationship of bureaucratic expertise in relation to a principal, in this case Congress.

We also include two variables to test the efficacy of the hierarchical relationship *within* an agency. One question measures the extent to which bureaucratic subordinates were helpful in "day-to-day" contact (i.e., regarding routine tasks requiring some knowledge about agency procedures). The variable `Career.Exec-Day2day` measures their helpfulness and is measured on a five-point scale from "not helpful" to "very helpful." Secondly, we use a question measured in the same way that asks to what extent these direct reports were helpful with difficult technical issues. This will obviously matter more in an agency such as the Nuclear Regulatory Commission (NRC) than in a less-technical agency and is thus a test of whether information distinctions matter.

Lastly we include a variable that provides self-imposed workload, which differs dramatically across executives. This is related to the required effort to manage career agents, the time required to prepare for and respond to principals (Congress and the President), and the effects of technical and organizational complexity. This variable is included to control for the impact of such internalized challenges on the level of stress.

4.2 Model Specifics

From the Bayesian hierarchical model with Dirichlet process priors as specified in Section 3, with concentration parameter $m = 500$ and number of outcome categories $C = 5$, we generated 50,000 MCMC samples using the algorithm in Section A.2 and recorded the final 20,000 iterates of the chain. Strong evidence of convergence of this Markov chain is provided by standard diagnostics (Brooks and Gelman 1998; Gelman and Rubin 1992; Geweke 1992; Heidelberger & Welch 1981, as well as graphical methods).

4.3 The Results

Table 2 gives posterior means and 95% HPD intervals for each of the coefficients, including the cutpoints, where the sign of the coefficients is oriented such that positive numbers point toward increasing stress and negative numbers point toward decreasing stress. This model shows better overall fit than the simple approach even though it also has six of the posterior distributions with 95% HPD intervals bounded away from zero. These results are quite interesting. More conservative appointees have higher expected stress than their more liberal T2

Table 2. Dirichlet process posterior for survey of political executives

Posterior	Mean	95% HPD Interval
Government Experience	0.120	[−0.086: 0.141]
Ideology	0.076	[−0.031: 0.087]
Committee Relationship	−0.181	[−0.302: −0.168]
Career.Exec-Compet	−0.176	[−0.343: −0.158]
Career.Exec-Liaison/Bur	0.105	[−0.024: 0.118]
Career.Exec-Liaison/Cong	−0.029	[−0.130: 0.019]
Career.Exec-Day2day	−0.154	[−0.304: −0.139]
Career.Exec-Diff	0.115	[−0.024: 0.129]
Confirmation Preparation	−0.316	[−0.598: −0.286]
Hours/Week	0.447	[0.351: 0.457]
President Orientation	−0.338	[−0.621: −0.309]
Cutpoints: (None) (Little)	−1.488	[−1.958: −1.598]
(Little) (Some)	−0.959	[−1.410: −1.078]
(Some) (Significant)	−0.325	[−0.786: 0.454]
(Significant) (Extreme)	0.844	[0.411: 0.730]

counterparts, although the 95% interval barely covers zero (6% of the posterior density is below zero on the axis). We have already speculated that this might be the case because the data cover an era in which the Congress was generally controlled by the Democrats. Furthermore, Democratic House caucus rules mean that the chairmen (no chair-women during this time) were the most senior member of the committee. Given the short tenure of our objects of study on their jobs, they are almost certainly swamped in terms of experience by the individual holding the gavel (Joyce 1990; Cohen 1986). This is supported by noting that better relationships with the relevant oversight committee is associated with lower levels of stress, from the model results.

It turns out that working longer is associated with increased expected stress. This is clearly a difficult causal relationship to untangle because the effects are almost certainly endogenous: higher stress induces longer work-hours and longer work-hours lead to higher stress, or there is an antecedent intervening variable influencing both. Relatedly, but more surprising, having previous experience in government increases expected stress. We have no way of knowing from the confidential side of the data (which is not distributed, as mentioned), but it may be that these are individuals picked for higher pressure assignments: greater stakes, more partisan policy decisions, important national security or economic decision-making, and so on. In fact, it is logical for presidents to pick experienced government hands for sensitive posts and we see plenty of evidence that presidents do this by nominating former members of Congress and former cabinet members from previous administrations. We have to be cautious here because about 10% of the posterior density is on the negative side of the x-axis.

Notice that those that undergo confirmation preparation before testifying have lower expected stress on the job. We suspect that this increases the information these executives have about detailed presidential priorities and preferences, as communicated in the process of preparing them to answer questions from Senators. There is plenty of anecdotal and journalistic evidence to suggest that potential cabinet members vary considerably in how they do during this process. Contrast John Roberts' flawless performance during his confirmation hearing for Chief Justice to John G. Tower's disastrous attempt to become confirmed as George H.W. Bush's first Secretary of Defense. This finding is supported by the strong negative on stress effect found for executives who believe that there should be a formal orientation process whereby the president's goals and policy preferences are clearly laid out. Taken together, these findings suggest to future presidents that agency executives function better in administrations that clearly articulate their direction and prepare future agency heads as full-fledged members of the policy team.

The five bureaucratic posteriors provide both confirmatory and surprising findings. The implication that increased competence of direct underlings in the agency decreases expected stress is not particularly earth-shaking, but it makes intuitive sense. Secondly, the more these lower tier managers work with members of Congress and their staff (presumably mostly with the staff), the lower the expected stress of the senior agency executive. This makes sense given the often formal and difficult visits that senior government executives have with members and committees on the hill. On the other side of the coin, there is some evidence that the more these subordinates deal directly with other (perhaps competing) agencies the greater the expected stress with 96% of the density on the positive side of zero. This implies that agency heads working in policy environments that are less clearly delineated, thus requiring more cross-agency interaction, have higher stress. Additionally, the more helpful underlings are on a day-to-day basis, the less expected stress. Whereas this result is quite intuitive, it also implies that in agencies where the senior executives need more interaction from middle managers and technocrats, there is more stress to be reduced. Similarly, the finding that great help with difficult technical analysis from underlings increases expected stress (over 96% of the density on the positive side) is likely strongly related to the technical challenges confronting the agency in general (only 4% of the posterior density below zero).

How do these results differ from the earlier analysis with uniform priors? The posterior means differ little since the Dirichlet process was applied to the random effect. Thus the primary improvement is in overall model fit from the binning process. Where we see the greatest difference is in the width and placement of the posterior HPD intervals. For 9 of the 11 coefficient posteriors, the distributions are noticeably narrower around the mean with the new model, dramatically so in the case of Confirmation Preparation. However, whereas the outcome category widths are similar, there is a different latent scale for each model estimation. One way to account for this is to look at relative effects put on a common scale. In Table 3 we provide the 0.025 and 0.975 quantile posterior effect $(\tilde{\beta})$ at the mean of the data $(\bar{\mathbf{X}})$ compared with the 0.025 and 0.975 quantile posterior effect at a reversion level: $\bar{\mathbf{X}}_j\tilde{\beta}_j - \mathbf{X}_{j,r}\tilde{\beta}_j$ for each of $j = 1$: 11 covariates. Arbitrarily we pick the reversion level as the minimum X_j, which is either zero or one (see Appendix A for data details). So what this shows is the ability of each explanatory variable to ''push'' the outcome through ordinal categories as it moves from the minimum to the mean (analyzing the analogous effect on the other side of the mean could also be easily done). Picking quantiles of the posteriors is done because the difference in models is not at the mean values, but across the full posterior width. Scale comparability is achieved by dividing these values by the model's average threshold difference estimated on the latent variable scale, $d\bar{\theta}$ for uniform priors and Dirichlet process priors.

The important difference in effects found in Table 3 are all found in the lower quantile (except that `Hours/Week` has a notable difference also at the 0.975 level), suggesting an asymmetry of effects. This is interesting because more of the responses were at the high end of the five-point scale: (51, 54, 96, 200, 131), and thus suggests that the most important effect is moving upward from a low level to the mean is `Hours/Week`. Also, in two cases, `Career.Exec-Liaison/Bur` and `Career.Exec-Day2day`, the outcome effect is in opposite directions between the two models over this data range at the 0.025 quantiles. Therefore the additional information in the data reflected in the nonparametric priors process gives a strikingly different result independent of the similar posterior means comparisons. Lastly, note that collectively the biggest differences are with the bureaucracy-oriented explanatory variables, suggesting that the new model is successful because the binning process is picking up agency heterogeneity, which we know to exist but cannot directly control for due to data restrictions.

Additionally, we recorded the number of occupied subclusters and the average size of these subclusters as the Gibbs sampler was running. Figure 1 shows histograms of these values for the last 2,000 iterations. It appears that there is considerable heterogeneity amongst cases reflected in typical number of occupied bins. There are also relatively few cases in each bin, averaging a little over two across runs. These results are highly robust to prior parameter values suggesting a strong preference by the data.

Finally, we issue a reminder of the difficulties in making strong causal claims with observational social science data (Morgan and Winship 2007, chap. 1). It is likely that the bureaucratic variables are intertwined in their effect on individual stress, whereas the individual characteristics stand more on their own. For instance, the overall competence of reporting career executives, as well as their ability to effectively handle day-to-day management and difficult technical issues cannot possibly be independent. Furthermore, in terms of intragovernmental relations, the degree to which these subordinates liaison with Congress and liaison with other bureaucracies must be related. Whereas we have stayed away from counterfactual claims here (Pearl 2000), our assertions are inferential in the regression sense with claims that changes in explanatory variables are associated with *expected* effects on the outcome variable. Most of these marginal posteriors in Table 2 are noticeably distinct from zero and would survive standard one-sided significance tests, if we were inclined to use such devices. Notice also that the coefficient posterior effects corresponding to `Career.Exec-Liaison/Bur` and `Career.Exec-Liaison/Cong` are large and *in the opposite directions*, meaning that they survive any watering-down that may occur from endogeneity or conflicting explanation.

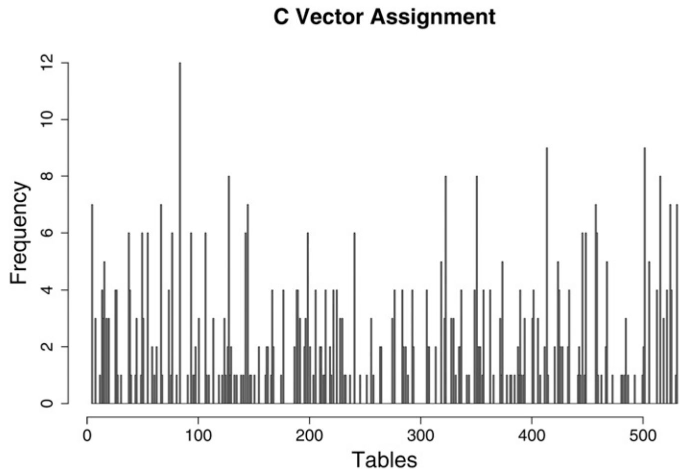

Figure 1. Subclustering summary across Gibbs sampling.

5. DISCUSSION

Priors are really about theories in the absence of undeniable facts. Priors thus give researchers a way to insert incomplete information into a model specification without pretending it fits into one of two extremes: completely unknown (parameters in classical inference), or completely known (observed data). Thus theories hold a place between absolutes, and are easily recognized with Bayesian approaches. Theories are also clarifications and distillations. In *Discourse on Metaphysics* (sec. IV), Leibnitz says that a theory has to be more simple than the

Table 3. Posterior model quantile comparison, $(\bar{\mathbf{X}}_j\tilde{\beta}_j - \mathbf{X}_{j,r}\tilde{\beta}_j)/d\bar{\theta}$

	Uniform Priors		Dirichlet Process Priors	
Marginal Posterior Quantiles	**0.025**	**0.975**	**0.025**	**0.025**
`Government Experience`	−0.070	0.233	−0.058	0.231
`Ideology`	−0.287	0.779	−0.252	0.774
`Committee Relationship`	−1.739	−0.051	−1.685	−0.050
`Career.Exec-Compet`	−1.894	0.287	−3.345	0.285
`Career.Exec-Liaison/Bur`	−0.342	1.262	1.882	1.253
`Career.Exec-Liaison/Cong`	−0.601	0.408	−1.568	0.405
`Career.Exec-Day2day`	−1.660	0.278	0.078	0.276
`Career.Exec-Diff`	−0.382	1.448	−0.952	1.438
`Confirmation Preparation`	−0.150	−0.003	−0.073	−0.003
`Hours/Week`	1.448	2.719	−0.787	2.700
`President Orientation`	−0.150	−0.008	−0.145	−0.008

data it explains to be useful. In fact, today we usually call that *data reduction*. The question is what kind of theories is a researcher willing to support with prior specifications: strongly informed substantive knowledge that has to be rigorously defended, or very weak but not totally uninformative statements? The latter tend to be well-known forms, such as a normal with large variance, uniforms, and more recently Gelman's (2006) half-Cauchy prior distribution case for variance terms.

Our notion is that the nonparameteric form of Dirichlet mixtures represents a new and useful paradigm for semi-informed prior information that reflects both information from observations and researcher intuition, where neither dominates. This is especially important in the social sciences, where the envelopment of nonparametric prior families helps resolve the historical distrust of overtly subjective prior specifications.

Ultimately the value of our approach lies in its ability to model difficult data and produce results that existing alternative methods fail to discover. We have presented evidence here that not only are we able to account for unobserved, but important structure, we are also able to substantially improve on previous attempts to model these data (i.e., Gill and Waterman 2004). This not only demonstrates the flexibility of the Bayesian paradigm, but also the utility of nonparametric model components in the context of analyzing social science data.

APPENDIX A: GENERATION OF PARAMETERS

It is easiest to treat the parameters of the Gibbs sampler in two blocks, one consisting of the Dirichlet parameters, and the other block consisting of the remaining parameters.

A.1 Generation of the Dirichlet Parameters

Because realizations of the $\mathcal{DP}$ are almost surely discrete (even though the generating mechanism is continuous), the model for ψ is a countably infinite mixture (Ferguson 1973; Berry and Christensen 1979; Lo 1984). Blackwell and MacQueen (1973) discovered that if $\psi_1, \ldots, \psi_n$ are iid from $G \sim \mathcal{DP}$, then the marginal distribution of $\psi_1, \ldots, \psi_n$ (marginalized over the prior parameters) is equal in distribution to the first n steps of a Pólya process. Therefore reference can be made to a finite rather than infinite dimension, and Dirichlet process posterior calculations involve a single parameter over this space.

Following Neal (2000), and using the Blackwell and MacQueen (1973) result, we write the Dirichlet process (6) as

$$\begin{aligned} c_i &\sim \text{Discrete}(q_1, \ldots, q_K) \\ \psi_{c_i} &\sim g(\psi) = N(\mu, \tau^2) \\ \mathbf{q} &\sim \text{Dirichlet}(m/k, \ldots, m/k), \quad k \le n \end{aligned} \tag{A.1}$$

where the c_i serve only to group the ψ_i, resulting in a common value of $\psi_i = \psi_j$ if $c_i = c_j$. (It is important to remember that this binning process is there to provide better model fit without regard to the parsimony of the number of groups. Therefore, these are not clusters in the typical substantive sense, but are instead subclusters that could be aggregated later into agency type with a follow-on procedure. Such follow-on procedures tend to be elaborate and this is an area of current attention.) For instance, if $c_2 = c_5$ then y_2 and y_5 share a common value of ψ. To reflect this fact, instead of writing the distribution of U_i as in (3), we write $U_i \sim N(X_i\beta + \psi_{c_i}, \sigma^2)$. We fix the value of σ at one so that the model is identified, and this also sets the scale of the latent utility dimension to a standard and convenient interpretation. Now a Gibbs sampler iterates between Dirichlet parameters $\{U_i.C_I, \psi_{C_I}\}$ and model parameters $\{\beta, (\mu, \tau^2), m, \theta\}$ enabling us to define the process in two stages.

In the Gibbs sampler, the c_i are generated conditionally as follows for $\ell \le n$. First, we define:

$$\begin{aligned} \mathbf{c} &= (c_1, c_2, \ldots, c_n) \\ \mathbf{c}_{-i} &= (c_1, c_2, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n) \\ n_{-i,\ell} &= \#(c_j = \ell), j \ne i, \end{aligned}$$

then draw from:

$$\begin{aligned} f(y_i|\psi_i) &= \prod_{j=1}^{c} \left[\Phi\left(\frac{\theta_j - X_i\beta - \psi_i}{\sigma^2}\right) - \Phi\left(\frac{\theta_{j-1} - X_i\beta - \psi_i}{\sigma^2}\right) \right]^{y_{ij}} \\ &= p_j = P(\theta_{j-1} \le U_i \le \theta_j) \qquad \text{from (2)}, \end{aligned}$$

and, for $i = 1, \ldots, n$

$$P(c_i = \ell \,|\mathbf{c}_{-i}) \propto \begin{cases} \dfrac{n_{-i,\ell}}{n-1+m} f(y_i|\psi_i) & \text{if } n_{-i,\ell} > 0 \\ \dfrac{m}{n-1+m} H_i & \text{if } n_{-i,\ell} = 0 \end{cases}$$

where

$$H_i = \int_{-\infty}^{\infty} \int_{\theta_{j-1}}^{\theta_j} f(u|\psi) g(\psi)\, du\, d\psi.$$

This is the Pólya process weighted according to our selection criteria. Efficient implementation of the Gibbs sampler requires easy calculation of H_i, and thus many authors have stressed the need for a conjugate setup. Here, because of the ordered probit model, the setup is not conjugate. But the introduction of the latent U_i still allows quick calculation of H_i because it is straightforward to show

$$\begin{aligned} H_i &= \int_{-\infty}^{\infty} \int_{\theta_{j-1}}^{\theta_j} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u - X_i\beta - \psi)^2/2\sigma^2} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\psi^2/2\tau^2}\, du\, d\psi \\ &= \int_{\theta_{j-1}}^{\theta_j} \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} e^{-(u - X_i\beta)^2/2(\sigma^2+\tau^2)} du \\ &= \Phi\left(\frac{\theta_j - X_i\beta}{\sqrt{\sigma^2+\tau^2}}\right) - \Phi\left(\frac{\theta_{j-1} - X_i\beta}{\sqrt{\sigma^2+\tau^2}}\right). \end{aligned}$$

Note that if a new value of c_i is not equal to a previous value, there will be no ψ associated with that group. Thus, in such a case, we must immediately generate a value of ψ from its posterior based on Y_i (see (A.2)). Once the set $\mathbf{c}$ has been updated we can then update $(\psi_1, \ldots, \psi_n)$, which is done with a common value for all $c_i = c$. Let $I_c = \{i\colon c_i = c\}$. To update ψ_c use the posterior distribution

$$\psi_c \sim N\left(\frac{n_c\tau^2 T}{\sigma^2 + n_c\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n_c\tau^2}\right), \tag{A.2}$$

where $n_c = \#(c_i = c)$ and $T = (1/n_c)\sum_{i \in I_c}(U_i - X_i\beta)$.

A.2 Generating the Model Parameters

Once the Dirichlet parameters have been generated, the remainder of the Gibbs sampler is relatively straightforward. We describe the steps now.

1. The posterior distribution of β is

$$\beta \sim N\big(A^{-1}b,\ A^{-1}\big), \tag{A.3}$$

where

$$A = \frac{1}{\sigma^2}X'X + \frac{1}{\sigma_\beta^2}I, \qquad b = \frac{1}{\sigma^2}X'(\mathbf{U}-\psi) + \frac{1}{\sigma_\beta^2}\beta_0,$$

and $\mathbf{U} = (U_1, \dots, U_n)$, $\psi = (\psi_1, \dots, \psi_n)$.

2. The posterior distribution of μ and τ^2 is calculated as follows. When updating μ and τ^2 we *only use the distinct values* of $(\psi_1, \dots, \psi_n)$. Denote these distinct values by $(\psi_{c_1}, \dots, \psi_{c_r})$, $r \le n$. Then

$$\begin{aligned} \mu \mid \psi, \tau^2 &\sim N\left(\frac{d\tau^2}{rd\tau^2+1}\sum_{i=1}^{r}\psi_{c_i}, \frac{d\tau^2}{rd\tau^2+1}\right) \\ \frac{1}{\tau^2} &\sim \text{Gamma}\left(\frac{r+1}{2}+a, \frac{1}{2}\sum_{i=1}^{r}(\psi_{c_i}-\mu)^2 + \frac{1}{b}\right), \end{aligned} \tag{A.4}$$

3. The distribution of θ is updated from $\prod_{i=1}^{n}\prod_{j=1}^{C} I(\theta_{j-1} \le U_i \le \theta_j)^{yij}$ and is given by

$$\theta \sim \text{Uniform}\left(\max_{i:y_{ij}=1} U_i,\ \max_{i:y_{i,j+1}=1} U_i\right), \tag{A.5}$$

4. Lastly, U_i is updated from

$$U_i \sim N\big(X_i\beta + \psi_{c_i}, \sigma^2\big) \quad \text{truncated to } (\theta_{j-1}, \theta_j). \tag{A.6}$$

because the conditional posterior of U_i is a truncated normal.

Now consider generating values for m, as the concentration parameter of the Dirichlet process, m, can also be put into the Gibbs sampler. In the joint distribution for the model, m only appears in the distribution of $\mathbf{c}$. Therefore, the conditional posterior distribution of $M = m$ given $\mathbf{c}$ is, by Bayes's rule,

$$P(M = m \mid \mathbf{c}) = \frac{P(\mathbf{c} \mid M = m)P(M = m)}{\sum_{m'} P(\mathbf{c} \mid M = m')P(M = m')}, \tag{A.7}$$

where $P(M = m)$ is the prior specification. It can be shown, (Appendix A.3) that

$$P(M = m \mid \mathbf{c}) = \frac{\frac{\Gamma(m)m^{c*}}{\Gamma(n+m)}P(M = m)}{\sum_{m'}\frac{\Gamma(m')m'^{c*}}{\Gamma(n+m')}P(M = m')}, \tag{A.8}$$

where $c*$ is the number of distinct c_i. Thus, we have a simple conditional form to draw from where the prior on m is a discrete set and G has root density $\mathcal{N}(\mu,\tau^2)$. In practice, we establish a finite set of positive integer values from 1–200 for m, and place a uniform prior over this support. In experimental work fixing m at specific values ranging up to 20,000, we found that the model described here is relatively insensitive to this assignment, because the Dirichlet process is applied to the random effect rather than directly to the covariate coefficients. The treatment of this parameter continues to be of great interest in the emerging Dirichlet process priors literature.

A.3 Distribution of m

To establish (A.8), note that from (A.1)

$$P(\mathbf{c} \mid M = m) = \frac{\Gamma(m)}{\Gamma(m/K)^K}\int q_{c_1}\cdots q_{c_n}\prod_{j=1}^{K} q_j^{\,m/K-1}\,d\mathbf{q}.$$

For each j, let $n_j = \#(c_i = j)$, and note that $\sum_j n_j = n$. Then

$$\begin{aligned} P(\mathbf{c} \mid M = m) &= \frac{\Gamma(m)}{\Gamma(m/K)^K}\int\prod_{j=1}^{K} q_j^{\,n_j+m/K-1}\,d\mathbf{q} \\ &= \frac{\Gamma(m)}{\Gamma(n+m)}\prod_{\substack{j=1\\ n_j>0}}^{n}\frac{\Gamma(n_j+m/K)}{\Gamma(m/K)}, \end{aligned} \tag{A.9}$$

where in the last line the upper limit on the product is changed to n, as no more than n of the c_i can have $n_j > 0$. Now we substitute this result back into (A.7) for m and m', to get:

$$P(M = m \mid \mathbf{c}) = \frac{\frac{\Gamma(m)}{\Gamma(n+m)}P(M = m)}{\sum_{m'}\frac{\Gamma(m')}{\Gamma(n+m')}\prod_{n_j>0}\frac{\Gamma(n_j+m'/K)\Gamma(m/K)}{\Gamma(m'/K)\Gamma(n_j+m/K)}P(M = m')}.$$

This can be simplified by first noting that $\lim_{K\to\infty}\Gamma(m/K)/\Gamma(m'/K) = m'/m$ so that

$$\lim_{K\to\infty}\prod_{n_j>0}\frac{\Gamma(n_j+m'/K)\Gamma(m/K)}{\Gamma(m'/K)\Gamma(n_j+m/K)} = \left(\frac{m'}{m}\right)^{\#(n_j>0)} = \left(\frac{m'}{m}\right)^{c^*},$$

where $c*$ is the number of distinct c_i. Thus

$$P(M = m \mid \mathbf{c}) = \frac{\frac{\Gamma(m)m^{c^*}}{\Gamma(n+m)}P(M = m)}{\sum_{m'}\frac{\Gamma(m')m'^{c^*}}{\Gamma(n+m')}P(M = m')}, \tag{A.10}$$

which provides us with a simple conditional form to draw from. This part of our algorithm is supplied to researchers as an option through an `R` package at the CRAN webpage.

APPENDIX B: DATA NOTES

This addendum summarizes the data format and coding decisions applied to the dataset: *PRESIDENTIAL APPOINTEES, 1964–1984 (ICPSR 8458)*, Principal Investigators: G. Calvin Mackenzie and Paul Light (Spring 1987). The categorical sums are given for each outcome exclusive of missing values. Missing data values, 3.5% of the total, are addressed here with *multiple imputation* (Little and Rubin 1983, 1987; Rubin 1987) using the `mice` (multiple imputation by chained equations) package in the `R` statistical environment. This package is a straightforward implementation of multiple imputation (e.g., Schafer 1997, chap. 4). See King, Honaker, Joseph, and Scheve (2001) for a review of missing data issues in the social sciences. Whereas we could have handled the missing data values in the context of the Gibbs sampler by drawing from their full conditional distribution, this produced a substantially slower mixing Markov chain.

1. `Stress`. This question was worded: "Thinking about your most recent service as a full-time, Senate-confirmed presidential appointee, how would you describe the impact of the demands of your work on your private life and your family? Compared with other employment experiences you have had, to what extent did your work as a presidential appointee create stress in your personal life or in relations with your family?"

The responses were coded from (1) "not stressful at all" through (5) "very stressful," with categorical totals: $n_1 = 51$, $n_2 = 54$, $n_3 = 95$, $n_4 = 198$, $n_5 = 130$.

2. `Government Experience`. Coded (1) if previous employer was United States, state, or local government, $n = 246$, and (0) otherwise, $n = 285$.

3. `Ideology`. This variable ascends across five points according to: (1) "very liberal," $n = 19$, (2) "liberal," $n = 113$, (3) "moderate," $n = 182$, (4) "conservative," $n = 179$, (5) "very conservative," $n = 37$.

4. `Committee Relationship`. This question was worded "How would you describe your interactions with committee members prior to and during your confirmation hearings?" and measured on a five-point scale from (1) "hostile" to (5) "friendly." The categorical totals were: $n_1 = 4$, $n_2 = 11$, $n_3 = 40$, $n_4 = 131$, $n_5 = 337$.

5. The survey contains a bank of questions asking the appointee to rate subordinate senior career executives in his or her department. These are all measured on a five-point scale with the following individual wordings:

(a) `Career.Exec-Compet`. Evaluation of *competence*, from (1) "low competence" to (5) "high competence," with the following totals: $n_1 = 3$, $n_2 = 18$, $n_3 = 64$, $n_4 = 240$, $n_5 = 199$.

(b) `Career.Exec-Liaison/Bur`. Helpful as a liaison with the federal bureaucracy, from (1) "not helpful" to (5) "very helpful" with totals: $n_1 = 14$, $n_2 = 24$, $n_3 = 89$, $n_4 = 208$, $n_5 = 191$.

(c) `Career.Exec-Liaison/Cong`. Helpful as a liaison with Congress, (1) "not helpful" to (5) "very helpful" with totals: $n_1 = 43$, $n_2 = 103$, $n_3 = 155$, $n_4 = 111$, $n_5 = 109$.

(d) `Career.Exec-Day2day`. Helpful in handling day-to-day management tasks, (1) "not helpful" to (5) "very helpful" with totals: $n_1 = 4$, $n_2 = 22$, $n_3 = 79$, $n_4 = 225$, $n_5 = 195$.

(e) `Career.Exec-Diff`. Helpful with technical analysis of difficult issues, (1) "not helpful" to (5) "very helpful" with totals: $n_1 = 7$, $n_2 = 22$, $n_3 = 80$, $n_4 = 161$, $n_5 = 257$.

6. `Confirmation Preparation`. This question asked whether various elements of the White House or host agency helped prepare the nominee for Senate committee testimony. We dichotomize it according to: no help from the White House or agency (0), or some form of preparation help (1). In total 88 appointees received direct help and 444 did not.

7. `Hours/Week`. The question asks "Including the time you spent working at the office, at home, and in other locations, how many hours per week on average did you spend working on your job during your most recent service as a full-time, Senate-confirmed presidential appointee? " Responses are coded according to: (1) less that 40 hr per week, $n = 0$, (2) 40–50 hr per week, $n = 27$, (3) 51–60 hr per week, $n = 115$, (4) 61–70 hr per week, n – 187, (5) 71–80 hr per week, $n = 140$, (6) 81–90 hr per week, $n = 39$, (7) more than 90 hr per week, $n = 21$.

8. `President Orientation`. This is a dichotomous response, a positive indication means that the respondent placed "learning the President's policy objectives" as "the most important component of an orientation program for new presidential appointees, one especially designed to serve the needs of appointees new to the federal government." A total of 82 ranked this first, and 425 picked another topic or none at all.

[Received January 2007. Revised April 2008.]

REFERENCES

Albert, J., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association,* 88, 669–679.

Aldrich, J. H., and McKelvey, R. D. (1977), "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections," *The American Political Science Review,* 71, lll–130.

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems," *Annals of Statistics,* 2, 1152–1174.

Baker, F. (2004), *Item Response Theory: Parameter Estimation Techniques*, New York: Chapman & Hall/CRC Press.

Bartels, L. (1996), "Pooling Disparate Observations," *American Journal of Political Science,* 40, 905–942.

Berk, R. A., Western, B., and Weiss, R. E. (1995), "Statistical Inference for Apparent Populations," *Sociological Methodology,* 25, 421–458.

Berry, D. A., and Christensen, R. (1979), "Empirical Bayes Estimation of a Binomial Parameter Via Mixtures of Dirichlet Processes," *Annals of Statistics,* 7, 558–568.

Blackwell, D., and MacQueen, J. B. (1973), "Ferguson Distributions Via Pólya Urn Schemes," *Annals of Statistics,* 1, 353–355.

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

Brooks, S. P., and Roberts, G. O. (1998), "Convergence Assessment Techniques for Markov Chain Monte Carlo," in *Statistics and Computing* 8, 319–335. Oxford: Oxford University Press, 169–193.

Burgess, A. W., and Holmstrom, L. L. (1974), "Rape Trauma Syndrome," *The American Journal of Psychiatry,* 131, 981–986.

——— (1979), "Adaptive Strategies and Recovery From Rape," *The American Journal of Psychiatry,* 136, 1278–1282.

Clinton, J., Jackman, S., and Rivers, D. (2004), "The Statistical Analysis of Roll Call Data," *The American Political Science Review,* 98, 355–370.

Cohen, J. E. (1986), "On the Tenure of Appointive Political Executives: The American Cabinet, 1952-1984," *American Journal of Political Science,* 30, 507–516.

Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000), *Generalized Linear Models: A Bayesian Perspective*, New York: Marcel Dekker.

Dinstein, Y. (2005), *War, Aggression and Self-Defence*, Cambridge: Cambridge University Press.

Doss, H. (2007), "Estimation of Bayes Factors for Nonparametric Bayes Problems via Radon-Nikodym Derivatives," Technical Report, University of Florida, Department of Political Science.

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association,* 90, 577–588.

Fenno, R. F. (1959), *The President's Cabinet: An Analysis in the Period from Wilson to Eisenhower.* New York: Vintage Books.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *Annals of Statistics,* 1, 209–230.

Freeman, J. L. (1969), *The Political Process: Executive Bureau-Legislative Committee Relations*. New York: Random House.

Gaubatz, K. T. (1996), ""Democratic States and Commitment in International Relations." *International Organization* 50, 109-39," *Annals of Statistics,* 1, 209–230.

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models," *Bayesian Analysis (Online),* 1, 515–533.

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science,* 7, 457–511.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4*, eds, J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger, Oxford: Oxford University Press.

Gill, J., and Waterman, R. W. (2004), "Solidary and Functional Costs: Explaining the Presidential Appointment Contradiction," *Journal of Public Administration: Research and Theory,* 14, 547–569.

Gormley, W. T. (1986), "Regulatory Issue Networks in a Federal System," *Policy,* 18, 595–620.

Greene, W. H. (2003), *Econometric Analysis*, (5th ed), Saddle River, NJ: Prentice Hall.

Hambleton, R. K., Rogers, H. J., and Swaminathan, H. (1991), *Fundamentals of Item Response Theory*, Newbury Park, CA: Sage Publications Inc.

Hart, P. M., Wearing, A. J., and Headey, B. (1993), "Assessing Police Work Experiences: Development of the Police Daily Hassle and Uplifts Scales," *Journal of Criminal Justice,* 21, 553–572.

Heclo, H. (1977), *A Government of Strangers: Executive Politics in Washington*, Washington: Brookings Institution Press.

——— (1988), "The In-and-Outer System: A Critical Assessment," *Political Science Quarterly,* 103, 37–56.

Heidelberger, P., and Welch, P. D. (1981), "Adaptive Spectral Methods for Simulation Output Analysis," *IBM Journal of Research and Development,* 25, 860–876.

Hill, J. L., and Kriesi, H. (2001), "Classification by Opinion-Changing Behavior: A Mixture Model Approach," *Political Analysis,* 9, 301–324.

Hirano, K. (2002), "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica,* 70, 781–799.

Holmes, T. H., and Rahe, R. H. (1967), "The Social Readjustment Rating Scale," *Journal of Psychosomatic Research,* 11, 213–218.

Hurrell, J. J., Nelson, D. L., and Simmons, B. L. (1998), "Measuring Job Stressors and Strains: Where We Have Been, Where We Are, and Where We Need to Go," *Journal of Occupational Health Psychology,* 3, 368–389.

Imai, K., Lu, Y., and Strauss, A. (2008), "Bayesian and Likelihood Inference for 2×2 Ecological Tables: An Incomplete Data Approach," *Political Analysis,* 16, 41–69.

Ingraham, P. W., and Ban,, C. R. (1986), "Models of Public Management: Are They Useful to Federal Managers in the 1980s," *Public Administration Review,* 46, 152–160.

Ivancevich, J. M., Matteson, M. X., Freedman, S. M., and Phillips, J. S. (1990), "Worksite Stress Management Interventions," *The American Psychologist,* 45, 252–261.

Jackman, S. (2000a), "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo," *American Journal of Political Science,* 44, 375–404.

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

——— (2001), "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking," *Political Analysis,* 9, 227–241.

Jacoby, W. G. (1994), "Public Attitudes Toward Government Spending," *American Journal of Political Science,* 38, 336–361.

Joyce, P. G. (1990), "An Analysis of the Factors Affecting the Employment Tenure of Federal Political Executives," *Administration & Society,* 22, 127–145.

Kaufman, H. (1981), *The Administrative Behavior of Federal Bureau Chiefs*. Washington, DC: Brookings Institution.

Keiser, L. R., and Soss, J. (1998), "With Good Cause: Bureaucratic Discretion and the Politics of Child Support Enforcement," *American Journal of Political Science,* 42, 1133–1156.

Keohane, R. O. (1986), "Reciprocity in International Relations," *International Organization,* 40, 1–27.

King, G., Honaker, J., Joseph, A., and Scheve, K. (2001), "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *The American Political Science Review,* 95, 49–69.

Laitin, D. D., and Wildavsky, A. (1988), "Political Culture and Political Preferences," *The American Political Science Review,* 82, 589–597.

Lazarus, R. S. (1991), "Psychological Stress in the Workplace, " in *Handbook On Job Stress,* ed. P. L. Perrewé, corte Madera, CA: Select Press, pp. 1–13.

Leamer, E. E. (1972), "A Class of Informative Priors and Distributed Lag Analysis," *Econometrica,* 40, 1059–1081.

Lehman, D. R., Wortman, C. B., and Williams, A. F. (1984), "Long-Term Effects of Losing a Spouse or Child in a Motor Vehicle Crash," *Journal of Personality and Social Psychology,* 52, 218–231.

Leibniz, G. W. (1686), *Discourse on Metaphysics and the Monadology* (trans. George R. Montgomery). Prometheus Books (1992).

Little, R. J. A., and Rubin, D. B. (1983), "On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-Data Likelihood," *The American Statistician,* 37, 218–220.

——— (1987), *Statistical Analysis with Missing Data*, New York: John Wiley & Sons.

Lo, A. Y. (1984), "On A Class of Bayesian Nonparametric Estimates: I. Density Estimates," *Annals of Statistics,* 12, 351–357.

Lowi, T. (1969), *The End of Liberalism: The Second Republic of the United States,* New York: W. W. Norton.

Martin, A. D., and Quinn, K. M. (2002), "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999," *Political Analysis,* 10, 134–153.

Mazmanian, D. A., and Sabatier, P. A. (1980), "A Multivariate Model of Public Policy-Making," *American Journal of Political Science,* 24, 439–468.

Miller, G. J., and Moe, T. M. (1983), "Bureaucrats, Legislators, and the Size of Government," *The American Political Science Review,* 77, 292–322.

Morgan, S. L., and Winship, C. (2007), *Counterfactuals and Causal Inference*, Cambridge: Cambridge University Press.

Neal, R. (2000), "The Short-Cut Metropolis Method," Technical Report No. 0606, University of Toronto, Department of Statistics.

Newton, T. J. (1989), "Occupational Stress and Coping With Stress: A Critique," *Human Relations,* 42, 441–461.

Niskanen, W. (1971), "*Bureaucracy and Representative Government*, Chicago: Aldine-Atherton Press.

Noll, R. G. (1971), "The Behavior of Regulatory Agencies," *Review of Social Economy,* 29, 15–19.

Olson, M. (1965), *The Logic of Collective Action,* Cambridge: Harvard University Press.

Palfrey, T. R., and Poole, K. T. (1987), "The Relationship Between Information, Ideology, and Voting Behavior," *American Journal of Political Science,* 31, 511–530.

Pearl, J. (2000), "*Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

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

——— (1997), *Congress: A Political-economic History of Roll Call Voting.* Oxford: Oxford University Press.

——— (1991), "Patterns of Congressional Voting," *American Journal of Political Science,* 35, 228–278.

——— (1965), *Congress: A Political-Economic History of Roll Call Voting.* Oxford: Oxford University Press.

Pratt, L. I., and Barling, J. (1988), "Differing Between Daily Events, Acute and Chronic Stressors: A Framework and Its Implications," in *Occupational Stress: Issues and Developments in Research*, eds. J. J. Hurrell, L. R. Murphy, S. L. Sauter, and C. L. Copper, New York: Taylor & Francis, pp. 41–53.

Quinn, K. M., Martin, A., and Whitford, A. B. (1999), "Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models," *American Journal of Political Science,* 43, 1231–1247.

Ramsey, F. L. (1972), "A Bayesian Approach to Bioassay," *Biometrics,* 28, 841–858.

Redford, E. S. (1969), *Democracy in the Administrative State*, New York: Oxford University Press.

Ringquist, E. J. (1995), "Political Control and Policy Impact in EPA's Office of Water Quality," *American Journal of Political Science,* 39, 336–363.

Rourke, F. E. (1984), *Bureaucracy, Politics, and Public Policy*, Boston: Little, Brown & Co.

Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons.

Rubin, D. B., and Schenker, N. (1987), "Logit-Based Interval Estimation for Binomial Data Using the Jeffreys Prior," *Sociological Methodology,* 17, 131–144.

Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, New York: Chapman & Hall/CRC Press.

Schweinberger, M., and Snijders, T. A. B. (2003), "Settings in Social Networks: A Measurement Model," *Sociological Methodology,* 33, 307–341.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica,* 4, 639–650.

Smith, A. (1999), "Testing Theories of Strategic Choice: The Example of Crisis Escalation," *American Journal of Political Science,* 43, 1254–1283.

Thurber, J. (1991), "Dynamics of Policy Subsystems in American Politics," in *Interest Group Politics,* eds. A. J. Cigler and B. A. Loomis, Washington, DC: CQ Press, pp. 302–348.

Vachon, M. L. S., Lyall, W. A. L., Rogers, J., Freedman-Letofsky, K., and Freeman, S. J. J. (1980), "Controlled Study of Self-help Intervention for Widows," *The American Journal of Psychiatry,* 137, 1380–1384.

Vachon, M. L S., Rogers, J., Lyall, W. A. L., Lancee, W. J., Sheldon, A. R., and Freeman, S. J. J. (1982), "Predictors and Correlates of Adaptation to Conjugal Bereavement," *The American Journal of Psychiatry,* 139, 998–1002.

Western, B. (1996), "Vague Theory and Model Uncertainty in Macrosociology," *Sociological Methodology,* 26, 165–192.

——— (1998), "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach," *American Journal of Political Science,* 42, 1233–1259.

Wilson, J. Q. (1989), "*Bureaucracy: What Government Agencies Do and Why They Do It*, New York: Basic Books.

Wolf, J. (2005), "Public Administration's Multiple Institutionalized Frameworks," *Public Organization Review,* 5, 183–200.