

# Bayesian Inference for Heterogeneous Event Counts

ANDREW D. MARTIN

*Washington University*

*This article presents an integrated set of Bayesian tools one can use to model heterogeneous event counts. While models for event count cross sections are now widely used, little has been written about how to model counts when contextual factors introduce heterogeneity. The author begins with a discussion of Bayesian cross-sectional count models and discusses an alternative model for counts with overdispersion. To illustrate the Bayesian framework, the author fits the model to the number of women's rights cosponsorships for each member of the 83rd to 102nd House of Representatives. The model is generalized to allow for contextual heterogeneity. The hierarchical model allows one to explicitly model contextual factors and test alternative contextual explanations, even with a small number of contextual units. The author compares the estimates from this model with traditional approaches and discusses software one can use to easily implement these Bayesian models with little start-up cost.*

**Keywords:** *event count; Markov chain Monte Carlo; hierarchical Bayes; multilevel models; BUGS*

## INTRODUCTION

Heterogeneity is what makes society and, for that matter, statistics interesting. However, until recently, little attention has been paid to modeling count data observed in heterogeneous clusters (a notable exception is Sampson, Raudenbush, and Earls 1997). Explanations

---

AUTHOR'S NOTE: This research is supported by the National Science Foundation Law and Social Sciences and Methodology, Measurement, and Statistics Sections, Grant SES-0135855 to Washington University. I would like to thank Dave Peterson, Kevin Quinn, and Kyle Saunders for their helpful comments. I am particularly indebted to Christina Wolbrecht for sharing her valuable data set with me. All replication materials, including BUGS code (in plain text and as WinBUGS ODC files) to estimate the models and the data employed, are available at the replication Web site: <http://adm.wustl.edu/replicate.html>.

throughout the social sciences predict that identical individuals will behave quite differently in different contexts. For example, rational choice accounts of behavior posit that individuals will act in accordance with an equilibrium that is in part a function of the other actors in the game. In short, individuals with the *same* preferences and the *same* characteristics are predicted to behave *differently in different* contexts. In this article, I review and introduce a type of statistical model one can use to model counts of observed behavior that allow the researcher to model and test different explanations of heterogeneity.

For many decades, social scientists have been interested in modeling the number of times a particular event occurs. Yet, these models have almost universally been used for homogeneous event counts, that is, individual cross sections when all observations are assumed to be exchangeable.<sup>1</sup> In many applications, this assumption is not tenable. For example, suppose that I am interested in modeling the number of times local school boards issued bonds for infrastructure projects in the 1990s and have collected data from every school board in every state. If I were to pool every observation into a single collection, I would be implicitly assuming that upon knowing the relevant characteristics of each school district, contextual state-by-state factors are unimportant. This assumption would be inappropriate in this example because state laws and funding mechanisms vary widely. The hierarchical model I introduce here relaxes this assumption; not only would it allow one to control for important contextual factors that might affect bond issues, but it also allows one to test competing contextual explanations.

The purpose of this article is to develop a general modeling strategy one can use to draw inferences from heterogeneous event counts. For practical reasons detailed below, I adopt a Bayesian approach and use Markov chain Monte Carlo (MCMC) methods for estimation. I begin by discussing Bayesian inference for cross-sectional event counts in Section 2. This section includes the standard Poisson regression model, as well as a model that allows for overdispersion. The section also includes an introduction to the software used throughout this article. In Section 3, I present results from a cross-sectional model of the number of women's rights cosponsorships for each member of the 83rd to 102nd House of Representatives. I then turn my attention in Section 4 to heterogeneous counts and offer a hierarchical Poisson

regression model suitable for this type of data. Section 5 contains results from the hierarchical model fit to the cosponsorship data. In this section, I also compare the findings to those from alternative modeling strategies. The final section concludes with a discussion of other types of models one can use for heterogeneous event counts.

### *BAYESIAN INFERENCE FOR CROSS-SECTIONAL EVENT COUNTS*

Before I turn my attention to heterogeneous event counts, it is important to cover Bayesian inference for homogeneous counts. These models are most commonly applied to single cross sections of data and require the statistical assumption that each observation is conditionally independent and thus exchangeable. These models have been used extensively in social science for a variety of applications. For a good introduction to event count models, I refer the reader to Cameron and Trivedi (1998). In the remainder of this section, I discuss cross-sectional event count models that allow for equidispersion and overdispersion.

#### *POISSON REGRESSION*

The canonical model for a cross section of event count data is the Poisson regression model. This model is based on the Poisson distribution, which is a discrete distribution defined on the nonnegative integers and can be derived from distributions of waiting times (Cameron and Trivedi 1998:6-8). The Poisson regression model is used to model a set of counts  $y_i \in 0, 1, 2, \dots, \infty = \mathbb{Z}^+$  on the nonnegative integers for a set of observations  $i = 1, \dots, N$ .

We further observe a row vector of explanatory variables  $x_i'$  of dimensionality  $(1 \times J)$ . The Poisson regression model is thus

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i), \\ \lambda_i &= \exp(x_i'\beta). \end{aligned} \tag{1}$$

We are interested in estimating the parameter vector  $\beta$ , which has dimensionality  $(J \times 1)$ . King (1988) estimates this model using maximum likelihood. Indeed, equation (1) defines the likelihood, which is simply the product of Poisson densities across all observations  $i$ .

The approach I take here is to estimate this model from a Bayesian perspective. Fundamentally, Bayesians assume that the data are fixed and that parameters have a distribution. Inference is conducted by computing the posterior probability density of the parameters conditioned on the data observed. The alternative frequentist (or classical) approach, on the other hand, assumes that parameters are fixed and unknown. One thus finds the parameters most likely to have generated the observed data by maximizing the likelihood. These approaches are inextricably related through Bayes theorem. For the Poisson regression model, I am interested in

$$f(\beta|y) \propto f(y|\beta) f(\beta).$$

The distribution  $f(\beta|y)$  is called the posterior density and is the probability density of  $\beta$  conditioned on observing the data. The posterior summarizes our a posteriori beliefs about the parameter vector  $\beta$  after having observed the data. From it, one can derive many probability statements, including those about the probability a particular element  $\beta_j$  differs from zero.

In many realistic applications, the posterior density  $f(\beta|y)$  does not take a closed form. However, there is a set of algorithms one can use to simulate from the posterior density. In so doing, one can make posterior inferences to any level of precision, depending on the computation time (which, thankfully, has gotten very inexpensive). The algorithms employed here are MCMC techniques, introduced by Geman and Geman (1984), Tanner and Wong (1987), and Gelfand and Smith (1990). For general introductions to MCMC methods, see Gelman, Carlin, Stern, and Rubin (1995) and Gill (2002).

As with any Bayesian model, it is necessary for the practitioner to formally posit her or his prior beliefs. To complete the Poisson regression model, I assume Normal independent priors for each element of the  $\beta$  vector  $\beta_j; \beta_j \sim \mathcal{N}(\mu_0, \sigma_0^2)$ .<sup>2</sup> In most applications, one uses noninformative priors—that is, priors that contribute little information about the parameters of interest (but there are cases when using informative priors is quite valuable) (Western and Jackman 1994). In the models presented here, I assume relatively noninformative priors. Thus, for each  $\beta_j$  in Poisson regression, I assume  $\beta_j \sim \mathcal{N}(0, 10^4)$ , which is an extremely flat distribution that contributes little information.<sup>3</sup>

Estimating statistical models using MCMC algorithms requires two steps that are oftentimes quite difficult. Most common MCMC techniques, including the Gibbs sampler and the Metropolis-Hastings algorithm, require the practitioner to derive the full conditional distributions. In most cases, this algebra is far from trivial and beyond the interest of most social scientists. The other difficulty is programming a computer to sample from the posterior density. To date, most commonly used statistical packages lack facility for MCMC estimation. Thus, programming in a high-level language such as Gauss, R, or S-Plus or a lower level language such as C++ or FORTRAN is oftentimes required to estimate these models. Thankfully, there is a freely available software package whose sole purpose is the estimation of Bayesian hierarchical models. This software is called BUGS, which stands for Bayesian inference Using Gibbs Sampling (Spiegelhalter, Thomas, Best, and Gilks 2000).<sup>4</sup>

BUGS eliminates many of the start-up costs of performing Bayesian inference and requires the practitioner to only write down the probability model and specify the priors. BUGS then chooses an appropriate updating algorithm to simulate from the posterior distribution. I refer the reader to the appendix for a brief description of BUGS and code for the models discussed in this article.

#### *POISSON REGRESSION WITH OVERDISPERSION*

The Poisson regression model has been used extensively, but there is a well-known deficiency of the model: The expected value (mean) of the distribution equals its variance (Cameron and Trivedi 1998:9-10). There are many observed counts, however, in which positive or negative contagion may occur and cause observed counts to be overdispersed (having variance greater than the mean) or underdispersed (having variance less than the mean). It is important to note that heterogeneity at the individual level will cause counts to be under- or overdispersed (see King 1989b). Thus, models that allow for under- or overdispersion allow the researcher to account for (but not model) unobserved heterogeneity within a single cross section. By neglecting overdispersion and fitting a model that requires equidispersion, standard errors will be inconsistent and inefficient.

When faced with overdispersion, the traditional approach is to mix a Gamma distribution with the Poisson, yielding a model with

overdispersion. By the Greenwood and Yule (1920) compounding rule, this results in the negative binomial regression model, which can also be estimated using maximum likelihood. For underdispersion, one must truncate the counts, which yields what King (1989b) terms the *continuous parameter binomial model*. When the existence of overdispersion or underdispersion is not known with certainty, King offers a generalized event count (GEC) model that allows the researcher to estimate the amount of dispersion without assuming a priori what type exists. In practice, most social science event counts tend to exhibit equidispersion or overdispersion (but for some examples of underdispersion, see King 1989b).

In this article, I will focus only on models that allow for overdispersion. However, instead of mixing the Poisson with the Gamma distribution, I offer an alternative model in which I mix the Poisson with the Normal distribution.<sup>5</sup> This model was offered by Breslow (1984). Here we have the same dependent variable  $y_i$ , the same number of observations  $N$ , and the same row vector of covariates  $x_i'$ . The only difference is that now I include a random effect, which has zero mean (thus not contributing to the count) but adds to the variance of  $y_i$ . I will call these effects  $\eta_i$ . The Poisson regression model with overdispersion can now be written as

$$\begin{aligned} y_i &\sim \text{Poisson}(\lambda_i), \\ \lambda_i &= \exp(x_i'\beta + \eta_i), \\ \eta_i &\sim \mathcal{N}(0, \tau^{-1}). \end{aligned} \tag{2}$$

$\tau$  is a scalar that captures the precision (inverse variance) of the random effect, thus indicating how much (or little) overdispersion there is. Again, to estimate this model in the Bayesian framework, I assume independent priors for each  $\beta_j$  :  $\beta_j \sim \mathcal{N}(\mu_0, \sigma_0^2)$ . I also assume a Gamma prior for the precision parameter  $\tau$  :  $\tau \sim \mathcal{G}(\nu_0, \delta_0)$ . For the models presented here, I assume noninformative priors for  $\tau$  :  $\tau \sim \mathcal{G}(0.001, 0.001)$ , which has its mass immediately to the right of zero and is decreasing on  $\mathbb{R}^+$ . This model is easily estimable in BUGS; see the appendix for the code employed. The payoff is that this model allows us to estimate the *magnitude* of the overdispersion, which could carry substantive importance in its own right or have profound effects on out-of-sample forecasting.

*CASE STUDY: PREFLOOR BEHAVIOR IN THE HOUSE*

To illustrate Bayesian inference for cross-sectional event count models, I fit models to data regarding cosponsorship on women's rights legislation in the 83rd to 102nd House of Representatives (1953-1992). This time period was particularly important due to the consideration of the Equal Rights Amendment. See Wolbrecht (2000) for a rich account of the politics of women's rights during this time period.

*WOMEN'S RIGHTS COSPONSORSHIPS*

In this case study, I employ data collected and previously analyzed by Wolbrecht (2000), who provides a detailed description of the data collection and coding scheme. The dependent variable  $y_i$  is the number of pro-women's rights bills cosponsored by each member in each Congress from the 83rd to 102nd sessions. Thus, the member of Congress in each term is the unit of analysis. The data set contains  $N = 8,808$  observations. I am interested in seeing what explains these counts of Congressional behavior.

The legislative studies literature suggests that cosponsorship is primarily a preference-based activity geared toward position taking (Campbell 1982) and intralegislative politics (Schiller 1995; Kessler and Krehbiel 1996; Martin and Wolbrecht 2000). While the implications of these explanations for the timing of cosponsorships differ, the implication after the intralegislative game has played out is clear: Members tend to cosponsor legislation close to their preferred policy position. Thus, I include a measure of general left-right ideology as an explanatory variable: Liberal members should cosponsor more women's rights legislation than conservative members. I include each member's D-Nominate first-dimension score as a measure of left-right tendency (Poole and Rosenthal 1997).<sup>6</sup> It is also likely that partisanship may affect women's rights cosponsorships. Indeed, majority party leaders may compel their members to cosponsor legislation (Cox and McCubbins 1993) as a tool for highlighting bill importance to constituents. I thus include party as an explanatory variable that takes a value 1 for Democrats and 0 for Republicans. Wolbrecht (2000) argues that prior to the passage of the Equal Rights Amendment (ERA) in the early 1970s, the Republican Party was more

supportive of women's rights than the Democratic Party. Thus, for these data, I expect Republicans to be *more* supportive of the ERA after controlling for policy preferences. Finally, I hypothesize that gender may affect the number of women's rights cosponsorships. Simply put, because of their greater awareness of and interest in women's rights issues, I expect women to be more likely to cosponsor women's rights legislation than men. I code the gender variable 1 for women and 0 for men.

In Table 1, I present results from the Poisson regressions of women's rights cosponsorships on the explanatory variables noted above. In this table, I report the posterior mean and standard deviation for each coefficient. Each of these can be interpreted like the familiar parameter point estimates and standard errors in likelihood inference. I also report the 95 percent credible intervals (CI), which summarize the central 95 percent of the posterior densities.<sup>7</sup> For both the equidispersion and overdispersion models, preferences are statistically significant; the more liberal the member of Congress, the more women's rights bills that member is likely to cosponsor. Gender is also statistically significant in both models: Women tend to cosponsor at a higher rate than men. In both models, party is also significant; after controlling for preferences and gender, Republicans tend to cosponsor more women's rights bills. There is profound overdispersion in this model, which implies that many members "over-cosponsor."<sup>8</sup> The importance of this finding of overdispersion cannot be overstated. The results suggest that there exists heterogeneity that is not explicitly accounted for by the model. Substantively, this implies that in different contexts (likely in different Congresses), members with the same issue preferences, the same gender, and the same party are behaving quite differently. To model and thus test for contextual explanations of these differences, it is necessary to employ a model that allows for this behavioral heterogeneity. I turn to such a model in Section 4.

#### *POSTERIOR PREDICTIVE DENSITIES*

When performing any statistical analysis, it is always useful to take coefficient estimates and transform them in a manner that is substantively meaningful. Perhaps the most effective way to do this is to use the parameter estimates to predict out-of-sample behavior. In



**TABLE 1: Summary of the Posterior Densities From Poisson Regressions of Cosponsorships on All Women’s Rights Legislation During the 83rd to 102nd House of Representatives**

Parameter	Equidispersion Model				Overdispersion Model			
	Posterior		95 Percent CI		Posterior		95 Percent CI	
	Mean	Standard Deviation	Lower	Upper	Mean	Standard Deviation	Lower	Upper
$\beta_1$ -Constant	0.368	0.020	0.328	0.408	-0.277	0.038	-0.348	-0.201
$\beta_2$ -Preferences	-3.129	0.042	-3.211	-3.048	-3.193	0.086	-3.363	-3.024
$\beta_3$ -Party	-0.892	0.032	-0.955	-0.830	-0.986	0.059	-1.108	-0.867
$\beta_4$ -Gender	1.193	0.025	1.142	1.242	1.410	0.077	1.258	1.558
$\tau^{-1}$ (Variance)					1.432	0.047	1.343	1.526
$\tau^1$ (Precision)					0.699	0.023	0.656	0.744

NOTE: Burn-in iterations = 1,000; Markov chain Monte Carlo (MCMC) iterations = 10,000;  $n = 8, 808$ . The equidispersion model is written in equation (1), and the overdispersion model is written in equation (2). The 95 percent CI is the 95 percent credible interval, which summarizes the central 95 percent of the posterior density. Noninformative priors for the  $\beta_j$  and precision  $\tau$  are assumed:  $\beta_j \sim \mathcal{N}(0, 10^4)$  and  $\tau \sim \mathcal{G}(0.001, 0.001)$ .

the Bayesian framework, one is interested in the posterior predictive density (Rubin 1984). For some out-of-sample observations  $\tilde{y}$ , the posterior predictive distribution for a Poisson regression model is simply

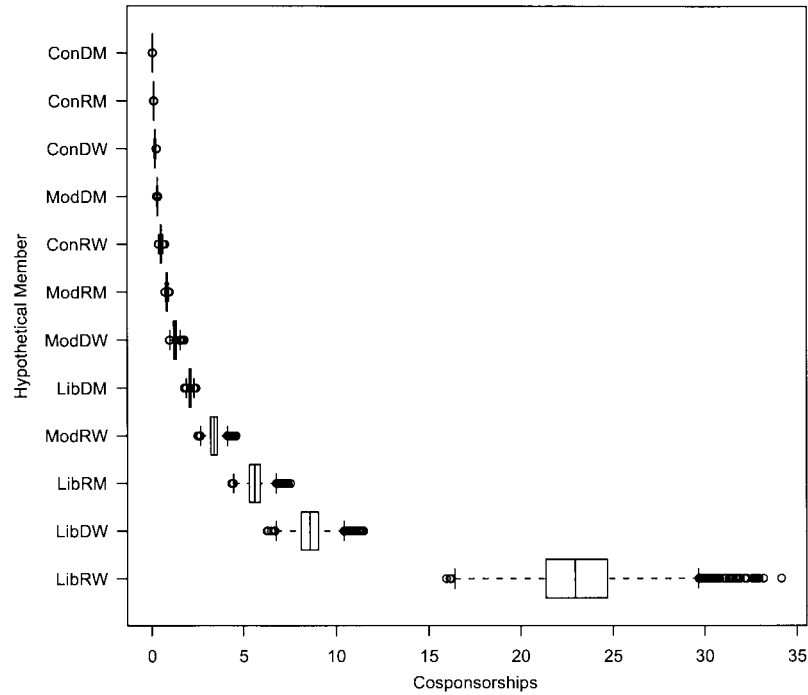
$$f(\tilde{y}|y) = \int f(\tilde{y}, \beta|y)d\beta = \int f(\tilde{y}|\beta)f(\beta|y)d\beta.$$

This average of  $f(\tilde{y}|\beta)$  over the posterior distribution  $f(\beta|y)$  can be evaluated using MCMC techniques. In fact, the distribution of any function of  $\beta$  can be estimated using MCMC methods.

To wit, I have computed the posterior predictive densities for hypothetical members of Congress. I am specifically interested in liberal (Lib), moderate (Mod), and conservative (Con) members of both parties (D or R) and both genders (M or W). I define a typical liberal as someone 2 standard deviations below the sample mean on the D-Nominate scale ( $-0.627$ ), a moderate equal to the sample mean ( $-0.027$ ), and a conservative as someone 2 standard deviations above the sample mean ( $0.573$ ). I plot these densities in Figure 1. Many important patterns can be discerned from this figure. First, liberal women, whether Democrats or Republicans, tend to cosponsor women's rights legislation most. The posterior predictive mean is approximately 23 cosponsorships per Congress for liberal Republican women and approximately 9 cosponsorships for liberal Democratic women. After liberal women, liberal and moderate Republican men cosponsor women's rights legislation next most often, followed by liberal Democratic men. For all other groups, the number of posterior predictive cosponsorships is negligible. It seems that preferences, party, and gender seem to be driving cosponsorship behavior in the 83rd to 102nd Houses. Yet, the results are not wholly satisfying. There is strong evidence of overdispersion, which suggests heterogeneity. To model that heterogeneity, I now turn to a multilevel event count model.

#### *BAYESIAN INFERENCE FOR MULTIPLE CROSS SECTIONS: HIERARCHICAL POISSON REGRESSION*

So far, I have presented results from commonly used models that can be estimated in standard software packages using maximum



**Figure 1: Boxplots of Posterior Predictive Densities for Hypothetical Members of the House**

likelihood. Besides philosophical claims about the importance of subjective probability, what is the advantage of Bayesian inference? In this section, I argue that the advantage of Bayesian approaches generally, and hierarchical or multilevel models specifically, is that they allow extremely flexible modeling of heterogeneous data. Many of these models are computationally impossible to estimate (or very nearly so) in classical settings. In many cases, asymptotic results are unreliable due to the small number of clusters or the small number of observations within each cluster. With a continuous-response variable, a hierarchical linear model can be used. Frequentist estimation for this model, assuming a reasonably large number of clusters, is tractable (Raudenbush and Bryk 2002; Jones and Steenbergen 2002). But for a count response variable, the frequentist estimation technology is not nearly as well developed (see Raudenbush and Bryk

2002, chaps. 10, 14). Also, the number of clusters typically has to be large (Breslow and Clayton 1993). Bayesian inference using MCMC, on the other hand, provides a computationally efficient way to estimate hierarchical models for heterogeneous data, even with a small number of clusters.

Assume now that we observe data from  $K$  different contexts (or clusters). Within each cluster, we have the following observations:  $i = 1, \dots, N_k$ . Note that there may be a different number of observations  $N_k$  in each cluster. The dependent variable  $y_{i,k} \in 0, 1, 2, \dots, \infty = \mathbb{Z}^+$  is an event count variable as discussed above. What sort of model should be applied to this data? One approach would be to pool all of the observations into one sample, which is the model discussed in the previous section. This would yield

$$\begin{aligned} y_{i,k} &\sim \text{Poisson}(\lambda_{i,k}), \\ \lambda_{i,k} &= \exp(x'_{i,k}\beta). \end{aligned}$$

which is the Poisson regression model discussed above. For this model, however, there is an implicit assumption that an individual  $i$  with the same vector of characteristics  $x'_{i,k}$  would behave the same in each decision context  $k$ . That is, pooling the data requires one to assume behavioral homogeneity. If individuals are hypothesized to behave differently in different contexts, pooling data in this manner is inappropriate. In some applications, this assumption is justifiable, yet what happens if contextual factors affect behavior?

An alternative model would be to estimate independent models for each context  $k$ . So, one would estimate

$$\begin{aligned} y_{i,k} &\sim \text{Poisson}(\lambda_{i,k}), \\ \lambda_{i,k} &= \exp(x'_{i,k}\beta_k). \end{aligned}$$

which is simply  $k$  separate Poisson regressions. This is a viable strategy but throws away the baby with the bathwater. If contextual factors are important, one would ideally *model* those contextual effects as opposed to ignoring them by estimating individual models. By dividing the data in this fashion, drawing inferences about contextual factors is impossible.

The common ground between these approaches is the Bayesian hierarchical model. This modeling strategy assumes that there are some commonalities between each context  $k$  but that there may



parameters. Thus, one only needs to specify the priors on the  $\alpha_{j,1}$  to  $\alpha_{j,p}$  and the  $\xi_j$  parameters for all  $j$ . In this application, I assume standard priors:  $\alpha_{\cdot,\cdot} \sim \mathcal{N}(\mu_{0,\cdot,\cdot}, \sigma_{0,\cdot,\cdot}^2)$ , where  $\cdot$  denotes all allowable values, and for the precisions  $\xi_j \sim \mathcal{G}(v_{0,j}, \delta_{0,j})$ .

The advantage of this model over the pooled or independent models is clear. One can look at the posterior density of the parameters  $\beta_k$  to see how the characteristics  $x'_{i,k}$  affect behavior *within* each cluster  $k$ . Also, one can look at the parameters  $\alpha_{\cdot,\cdot}$  to see how contextual factors affect behavior at the first level. In short, this modeling strategy allows one to *model* and *test* for various causes of behavioral heterogeneity. This approach will also provide more efficient estimates of the first-level effects due to “borrowing of strength” across clusters.

Others have applied Bayesian hierarchical event count models to various sorts of data. Albert (1992) fits a random-effects Poisson model to homerun data to detect contextual factors that affect various seasons of 12 baseball players. Christiansen and Morris (1997) develop a hierarchical Poisson model with a different parameterization than that above and provide software one can use in S-Plus to simulate from the posterior density. Chib, Greenberg, and Winkelmann (1998) posit an MCMC algorithm one can use to estimate from random-effects panel event count model using a unique parameterization and apply the model to epilepsy and patent data. Berry, Reese, and Larkey (1999) develop an interesting set of hierarchical models that can be used to compare players in different eras of sports. For their hockey data, they offer a hierarchical Poisson model that includes an elaborate clustering scheme and a random curve function to allow for aging. While these models are well traveled in applied statistics, they have been underused in social science applications.

#### *MODELING HETEROGENEITY IN WOMEN'S RIGHTS COSPONSORSHIPS IN THE HOUSE, 1953-1992*

The results from pooling all of the data from the 83rd to the 102nd House of Representatives and treating them as if they were a single cross section demonstrated that there was a great deal of unobserved heterogeneity in the data. Substantively, this is not difficult to swallow. Clearly, contextual factors, such as the national political

agenda, congressional rules, and chamber characteristics, would induce profoundly different behavior from a member of Congress with the identical policy preference of the same party and gender. In this section, I fit a hierarchical Poisson regression to model behavioral heterogeneity in this women's rights sponsorship data. Not only am I interested in explaining behavior within a particular Congress, but I am also interested in how the macropolitical factors discussed below affect congressional behavior.

This data set contains the number of women's rights bills cosponsored by every member in each of these Congresses.<sup>10</sup> Returning to the notation used in equation (3), here we observe behavior for  $K = 20$  Congresses (or clusters). Within each House, there are  $i = 1, \dots, N_k$  members. With member replacement within each Congress and the brief change of the size of the House in the late 1950s,  $N_k$  is approximately 440 in each Congress, although the actual numbers are slightly different for each Congress. We thus observe counts  $y_{i,k} \in \mathbb{Z}^+$ , one for each member  $i$  in each Congress  $k$ . I use the same explanatory variables  $x_{i,k}$ : policy preferences measured with first-dimension D-Nominate scores, a party dummy, and a gender dummy.

To serve as a baseline, I have estimated a set of independent Poisson regressions—one for each Congress  $k$ . Due to the large number of estimated parameters and limited space, I do not summarize the posterior distributions here.<sup>11</sup> There are some things to be taken from the results. First, the constants across each decision context vary substantially: The posterior mean takes its lowest value of  $-2.971$  in the 83rd House and its maximum value of  $1.738$  in the 99th House. Substantively, exponentiating these constants gives us the expected number of bills cosponsored for a moderate Republican man. For the 83rd House, we would expect this individual to cosponsor 0.01 bills, but in the 99th House, this member would be expected to cosponsor 5.7 bills. Clearly, the assumption of homogeneity is not met. Furthermore, we note that preferences significantly affect behavior in all but the 86th House. Yet, the magnitude of the relationship varies throughout the time period, again implying behavioral heterogeneity. The same story holds for the party and gender dummies: The  $\beta_{k,3}$  and  $\beta_{k,4}$  coefficients vary across clusters. The weakness of this baseline model is that while heterogeneity is apparent, we can neither model it nor test for contextual effects. If the goal of social science is to evaluate

competing social explanations, it is necessary to employ statistical models that allow one to test contextual explanations. I thus turn my attention to a hierarchical Poisson regression model, which does just this.

#### *HIERARCHICAL POISSON REGRESSION ESTIMATES*

The hierarchical model presented earlier allows one to model heterogeneity with suitable contextual variables. Clearly, members of Congress have behaved quite differently in different Congresses. What explains this heterogeneity? I hypothesize that four contextual factors are likely to affect cosponsorship behavior on women's rights. First, I hypothesize that the national political agenda will affect cosponsorship; once women's rights generally, and the ERA specifically, are firmly a part of the national agenda, the number of cosponsorships should increase. This is derived from electoral considerations. Once women's rights became supported by a majority, more members, regardless of preferences, party, and gender, would tend to cosponsor legislation. As argued by Wolbrecht (2000), the women's rights agenda was not firmly entrenched on the national political agenda until the 92nd Congress, the Congress that passed the ERA.

Second, congressional rules are also likely to affect congressional behavior. Throughout this time period, cosponsorship rules were liberalized. Until 1967, cosponsorship was not allowed in the House. Thus, members introduced identical legislation, using their name as the primary sponsor. In 1967, the rule was changed to allow for multiple cosponsorship, allowing up to 25 members to cosponsor each bill. This did not catch on for many years, and most members continued to introduce individual bills. This changed in 1978, when the rules were again changed to allow for unlimited cosponsorship. The second rule change was widely accepted by the membership and began to be used immediately. Thus, after the rules change in the 95th House, I expect cosponsorship behavior for all members to increase.

These first two contextual factors are hypothesized to explain the baseline level of cosponsorship for all members. In terms of the hierarchical model, they should thus affect the constant. For the 92nd Congress and thereafter, the constant  $\beta_{1,k}$  should be bigger than before



due to the change in the women's rights agenda. Similarly, after the rules change in the 95th House, we would expect the baseline level of cosponsorships to be higher, all things being equal. I thus include two dummy variables at the second level of the hierarchy, one that takes the value 1 for all Houses after the 91st House and one that takes the value 1 after the 95th House. I use these variables to model the distribution of the constant as

$$\beta_{k,1} = \alpha_{1,1} + \alpha_{1,2}(\text{Post 91st Dummy}_k) + \alpha_{1,3}(\text{Post 95th Dummy}_k) + \varepsilon_k \quad \varepsilon_k \sim \mathcal{N}(0, \xi_1^{-1}). \quad (4)$$

The hypothesis is that  $\alpha_{1,2}$  should be positive, indicating an increased level of women's rights cosponsorships for all members.  $\alpha_{1,3}$  should also be positive since it captures the agenda effect and the effect of the rules change.

I hypothesize that two additional contextual factors affect women's rights cosponsorships. First, I expect that chamber heterogeneity should affect the importance of preferences. I measure chamber heterogeneity by computing the standard deviation of the D-Nominate first-dimension scores for the chamber. The more heterogeneous the chamber, the more the magnitude of the preference  $\beta_{2,k}$  should decrease and vice versa. As preference heterogeneity increases in the chamber, the agenda for all members becomes larger. In addition, the probability of effecting policy change on any issue is more difficult in a heterogeneous chamber. It is known that more cosponsorship is a meaningful signal to policy makers (Young and Wilson 1997; Kessler and Krehbiel 1996). For women's rights specifically, liberal members desire policy change from the status quo, but with more issues on the agenda, I hypothesize that they are less likely to cosponsor. This implies that the marginal impact of preferences on women's rights cosponsorships is decreasing in chamber heterogeneity.

Second, differences in party structure are likely to affect cosponsorship behavior. As the distance between each party's median member (again measured using first-dimension D-Nominate scores as the absolute value of the median Democrat less the median Republican) increases, policy change will be easier for the majority party because of polarization. In this time period, the Democratic Party

was the majority, which also happened to be more liberal than the Republicans. Thus, with greater polarization between the parties, liberals (who happened to be overwhelmingly Democratic) would be more likely to cosponsor women's rights bills to influence their fellow members, and conservatives would be less likely. I thus hypothesize that the marginal impact of preferences is increasing in my measure of party difference. The specification for the  $\beta_{k,2}$  parameters is thus

$$\begin{aligned} \beta_{k,2} = & \alpha_{2,1} + \alpha_{2,2}(\text{Heterogeneity}_k) \\ & + \alpha_{2,3}(\text{PartyDiff}_k) + \varepsilon_k \quad \varepsilon_k \sim \mathcal{N}(0, \xi_2^{-1}). \end{aligned} \quad (5)$$

On the basis of the arguments above, I expect  $\alpha_{2,2}$  to be negative and  $\alpha_{2,3}$  to be positive. I assume that the remaining  $\beta_{k,3}$  and  $\beta_{k,4}$  on party and gender, respectively, are drawn from a common distribution with means  $\alpha_{3,1}$  and  $\alpha_{4,1}$  and variances  $\xi_3^{-1}$  and  $\xi_4^{-1}$ , respectively.

I present the results from the hierarchical Poisson regression in Table 2. For the sake of space, I do not present the posteriors from the first-level parameters in this table.<sup>12</sup> The more interesting findings are in the hyperparameters  $\alpha_{\cdot}$  that capture the contextual effects. While the CI for  $\alpha_{1,2}$  contains zero, 93.8 percent of the posterior density sample falls above zero. Thus, I can state with 93.8 percent certainty that this coefficient is positive, which substantively implies that the macropolitical change in the women's rights agenda did in fact contribute to a greater number of cosponsorships for all members, consistent with the argument of Wolbrecht (2000).  $\alpha_{1,3}$  is also positive as hypothesized and significantly different from zero, which indicates that the rules change contributed above and beyond the agenda change in affecting women's rights cosponsorships. Preference heterogeneity explains variance in the  $\beta_{k,2}$  parameters. As hypothesized, as preference heterogeneity increases, the marginal impact of preferences increases. There is no substantively significant finding for the party difference variable, however. The  $\alpha_{3,1}$  and  $\alpha_{4,1}$  parameters summarize the grand mean of the  $\beta_{k,3}$  and  $\beta_{k,4}$  parameters across the models. The conditional error variances of the hyperparameter regressions are interesting. Indeed, given only  $K = 20$  clusters, precise estimation of the hyperstructure is unlikely. However, these rather small variances indicate a good fit.<sup>13</sup>

**TABLE 2: Summary of the Posterior Density of a Hierarchical Poisson Regression of Cosponsorships on Women’s Rights Legislation for the 83rd to the 102nd House of Representatives**

Parameter	Posterior	Posterior	95 Percent CI	
	Mean	Standard Deviation	Lower	Upper
$\alpha_{1,1}$ -Constant	-0.993	0.286	-1.564	-0.428
$\alpha_{1,2}$ -Post-91st	0.777	0.513	-0.241	1.805
$\alpha_{1,3}$ -Post-95th	1.088	0.534	0.022	2.144
$\alpha_{2,1}$ -Constant	23.05	7.07	8.63	36.53
$\alpha_{2,2}$ -Hetero.	-139.4	41.58	-217.9	-53.5
$\alpha_{2,3}$ -PartyDiff	26.26	12.43	0.53	49.72
$\alpha_{3,1}$ -Constant	-0.806	0.077	-0.961	-0.655
$\alpha_{4,1}$ -Constant	0.984	0.077	0.829	1.135
$\xi_1^{-1}$ (Variance)	0.705	0.288	0.340	1.431
$\xi_2^{-1}$ (Variance)	0.820	0.367	0.362	1.762
$\xi_3^{-1}$ (Variance)	0.079	0.041	0.027	0.180
$\xi_4^{-1}$ (Variance)	0.077	0.050	0.020	0.206

NOTE: Burn-in iterations = 10,000; Markov chain Monte Carlo (MCMC) iterations = 25,000;  $n = 8, 808$ . The model is written in equation (3). This table only contains the Level 2 parameters. The 95 percent CI is the 95 percent credible interval, which summarizes the central 95 percent of the posterior density. Noninformative priors for the  $\alpha_{\cdot, \cdot}$  and  $\xi_j$  are assumed:  $\alpha_{\cdot, \cdot} \sim \mathcal{N}(0, 10^4)$  and  $\xi_j \sim \mathcal{G}(0.001, 0.001)$ .

*POSTERIOR PREDICTIVE DENSITIES*

While the table of coefficients is important in judging the significance and magnitude of the explanatory variables at the first and second levels of the hierarchies, the question remains as to what the substantive impact of these variables have on women’s rights cosponsorships. If it were the case that the number of cosponsorships varied between 4.9 to 5.1 for a typical member across decision contexts, we could conclude that the contextual factors had little substantive impact. To assess the substantive fit of the model, I have examined some interesting posterior predictive densities for two typical members of Congress: a liberal female Republican and a moderate male Democrat (using the same operationalizations as Figure 1). For each Congress, I plot a boxplot of the number of predicted cosponsorships. Figure 2 contains the summaries for the liberal Republican woman—the individual most likely to cosponsor women’s rights legislation. For the 83rd to the 91st Congresses, this

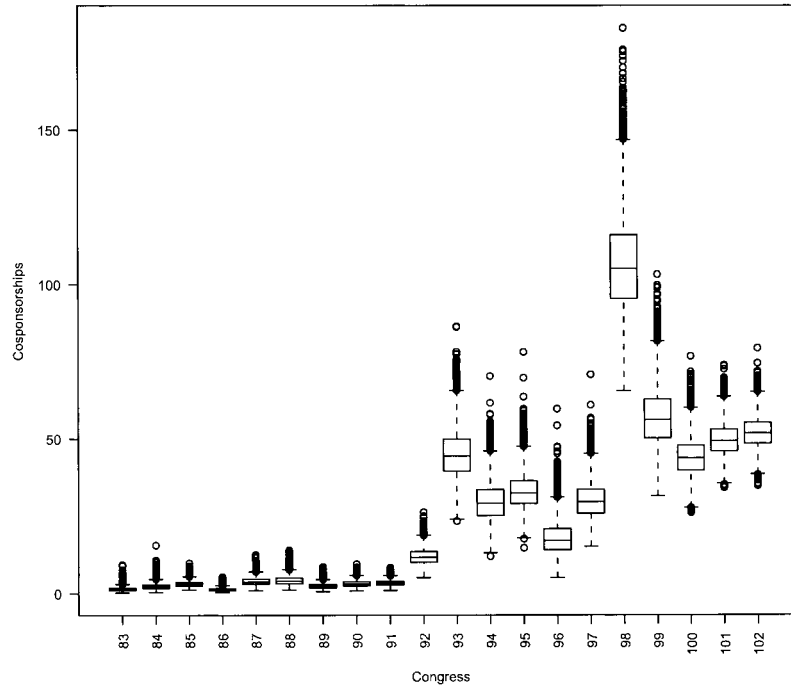
hypothetical member is predicted to cosponsor between 1.5 and 4 pieces of legislation per Congress. However, the agenda change in the 92nd Congress is quite apparent; this member is expected to cosponsor more than 11 bills in the 92nd Congress and more than 40 bills in the 93rd Congress and thereafter. The jump for the rules change in the 95th Congress is also apparent.

For the hypothetical moderate Democratic male in Figure 3, similar patterns are discernible. While the agenda change does not seem to have as dramatic an effect, the rules change in the 95th Congress seems to affect behavior dramatically. Indeed, this expectation for the hypothetical member went from substantially less than 1 cosponsorship per Congress until the 97th Congress, when the number jumps to 1.6 and then further jumps to 6.0 in the 98th Congress. Holding all else constant, the rules change dramatically affected cosponsorship behavior on women's rights bills, as did the change in the macro-political agenda. Clearly, contextual factors have a strong substantive impact on congressional cosponsorship behavior.

#### COMPARISONS WITH OTHER APPROACHES

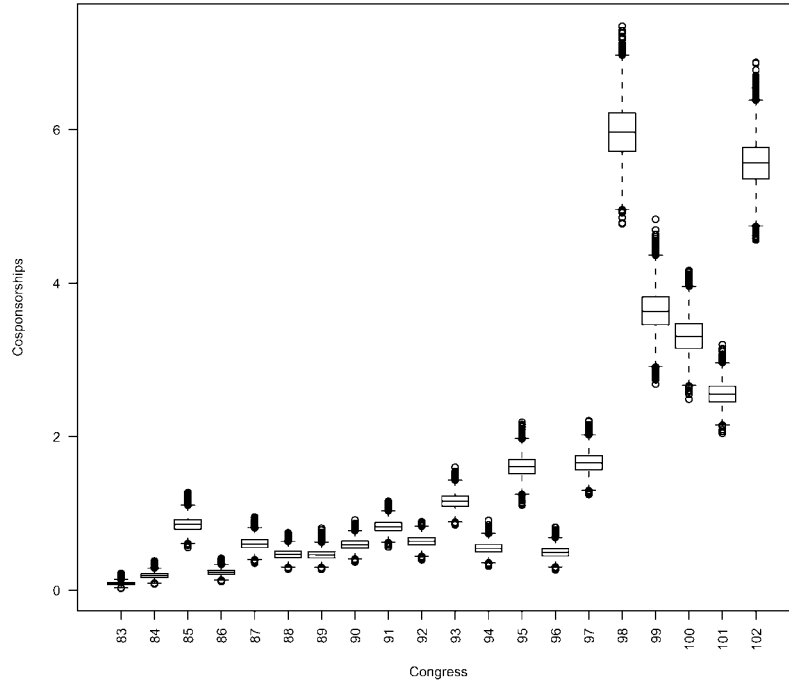
What advantages does this Bayesian approach afford over traditional approaches? One approach, oftentimes used in political science, is to account for contextual differences by including a fixed effect (i.e., a dummy variable) for each contextual unit—in this case, for each Congress. There are two weaknesses to this approach. First, it assumes that differences between contextual units exist in the mean level only; that is, none of the structural parameters differs across contexts. Second, while the model accounts for this specific type of heterogeneity, it does not *explain* it. It is thus not a particularly satisfying strategy if explanation of contextual effects is of interest.

To explicitly model heterogeneity in the structural parameters across contexts, there are essentially three strategies that can be used (Jones and Steenbergen 2002; Raudenbush and Bryk 2002). The first is simply to substitute the second-level equations (4) and (5) into the second line of equation (3). This results in an interactive model, which can easily be estimated using existing software for cross-sectional models. It is important to note that unlike the fixed-effects model mentioned above, this model allows structural parameters to differ



**Figure 2: Boxplots of Posterior Predictive Densities for a Liberal Republican Woman in the 83rd to 102nd Houses**

across contexts. The problem with this approach is not unlike the problem of measurement error. This model assumes that the first-level parameters are *perfectly* explained by the second-level covariates (technically, this assumes that  $\xi^{-1} = 0$ ). However, this is a very strong assumption that surely fails in realistic settings. If the assumption does not hold, parameter estimates can be inefficient and biased (Jones and Steenbergen 2002). For the sake of comparison, I present results from an interactive model in Table 3. The estimates for the interactive model are maximum likelihood estimates from a Poisson regression model. Since the MCMC algorithm employed for the hierarchical model is sampled from the exact posterior distribution, and since the priors used for the structural parameters and the variance terms are noninformative, the MCMC results in Table 2 can serve as



**Figure 3: Boxplots of Posterior Predictive Densities for a Moderate Democratic Man in the 83rd to 102nd Houses**

a baseline for comparison. The findings from the interactive model are inferior. First, the standard errors on the second-level parameters are entirely too small. Second, the magnitudes of the parameters in equations (4) and (5) are off the mark. Both of these are solely a result of the assumption that the second-level covariates perfectly explain the differences in the first-level parameters.

A second approach is to perform a two-stage procedure, first fitting separate Poisson regressions (using maximum likelihood) for the data from each context and then using the point estimates from these models as dependent variables in a second-stage regression analysis (estimated using ordinary least squares).<sup>14</sup> I report results from this meta model in Table 3. The problem with this approach is that it treats the first-level estimates as if they were known with certainty. This, in general, will produce overconfidence in results and may cause biases

**TABLE 3: Summaries of Alternative Models of Cosponsorships on Women’s Rights Legislation for the 83rd to the 102nd House of Representatives**

<i>Parameter</i>	Interactive ( <i>n</i> = 8, 808)		Meta ( <i>n</i> = 20)		W-Meta ( <i>n</i> = 20)		HGLM ( <i>n</i> = 8, 808)	
	<i>MLE</i>	<i>SE</i>	<i>OLS</i>	<i>SE</i>	<i>WLS</i>	<i>SE</i>	<i>PQL</i>	<i>SE</i>
$\alpha_{1,1}$ -Constant	-0.874	0.034	-1.010	0.281	-0.713	0.277	-1.765	0.323
$\alpha_{1,2}$ -Post-91st	1.100	0.037	0.765	0.506	0.573	0.435	3.334	0.239
$\alpha_{1,3}$ -Post-95th	0.737	0.028	1.117	0.528	1.231	0.395	-0.736	0.279
$\alpha_{2,1}$ -Constant	10.40	1.416	27.13	7.42	26.95	8.08	5.78	4.82
$\alpha_{2,2}$ -Heterogeneity	-50.33	7.523	-164.1	44.32	-167.1	47.11	-18.62	21.45
$\alpha_{2,3}$ -Party Difference	2.165	1.757	31.83	13.16	34.13	12.37	-6.95	5.01
$\alpha_{3,1}$ -Constant	-0.695	0.033	-0.815	0.083	-0.810	0.076	-0.804	0.079
$\alpha_{4,1}$ -Constant	1.066	0.025	1.006	0.107	1.008	0.074	0.992	0.076
$\xi_1^{-1}$ (Variance)	NA		0.710		0.667		1.573	
$\xi_2^{-1}$ (Variance)	NA		0.907		0.861		1.196	
$\xi_3^{-1}$ (Variance)	NA		0.137		0.342		0.081	
$\xi_4^{-1}$ (Variance)	NA		0.231		0.331		0.078	
Log-likelihood	-12785.04		NA		NA		-14777.07	

NOTE: The Interactive model contains maximum likelihood estimates (MLE) from a Poisson regression model with interactive terms between the Level 1 and Level 2 covariates. The Meta model contains ordinary least squares (OLS) estimates of the Level 2 parameters from a two-stage analysis in which the point estimates of Level 1 Poisson regressions are treated as known quantities. The W-Meta model contains weighted least squares (WLS) estimates of the Level 2 parameters from a two-stage analysis in which the point estimates of Level 1 Poisson regressions are treated as known quantities and the reciprocal of their estimated standard errors are used as weights. The hierarchical generalized linear model (HGLM) results are from a unit-specific model fit using penalized quasi-likelihood (PQL) in HLM (Raudenbush, Bryk, and Congdon 2000). SE denotes the (asymptotic) standard error.

if the errors among the first-level estimates are correlated (which they surely are). In this application, due to the small number of clusters and the large number of data points within each cluster, the meta-model does a reasonable job. All of the coefficient estimates are comparable with the posterior means from the Bayesian treatment. So too are the estimates of the conditional error variances in the second-level regressions. One manner in which one can account for the uncertainty in the first-level estimates using this two-stage strategy is to weight each observation in the second-level model by the reciprocal standard error from the first level. The third model in Table 3 contains these weighted least squares estimates.<sup>15</sup> In general, however, these meta-models will not provide satisfactory answers (Raudenbush and Bryk 2002). First, with a larger number of clusters or with fewer observations within each cluster, the amount of overconfidence in the findings will become apparent. Moreover, biases can enter the results if correlations exist between Level 1 estimates or between the Level 2 errors. The more principled approach is to treat the first-level estimates *as estimates* and account for that uncertainty in the subsequent analyses.

The third alternative approach is to fit the hierarchical model in equation (3) with frequentist methods. The most widely used software package to fit hierarchical models from a classical perspective is called HLM (Raudenbush, Bryk, and Congdon 2000). Our hierarchical Poisson regression is a special case of a hierarchical generalized linear model (HGLM). See Raudenbush and Bryk (2002, chaps. 10, 14) for an introduction to HGLMs. Estimating HGLMs from a frequentist perspective poses a number of difficulties due to the necessity of numerically integrating out random effects to compute the likelihood. While various approaches exist, HLM uses the penalized quasi-likelihood (PQL) approach of Breslow and Clayton (1993) to estimate HGLMs. The downside to using this approach is its difficulty dealing with a small number of clusters. Indeed, “the desirable properties [of the maximum likelihood approach] are based ... on large-sample theory. In the case of hierarchical models, the number of higher-level units ... is usually key in determining whether these large-sample properties will apply” (Raudenbush and Bryk 2002:14). In other words, the number of clusters  $K$  must be sufficiently large to trust the point estimates and the standard errors for any hierarchical model estimated using maximum likelihood. In this context, we are



faced with only 20 clusters. The results in the final columns of Table 3 are poor; only the parameters with no second-level covariates are estimated with the correct magnitude and standard errors.

These results suggest that a traditional frequentist analysis of these data would be less than satisfying. Beyond epistemological debates about the strengths and weaknesses of Bayesian inference, the results here suggest that Bayesian methods are a promising modeling tool when dealing with multilevel data with a small number of contextual units (Western and Jackman 1994). In sociology, political science, and other social sciences, this is a commonly faced problem. Of course, these findings beg a number of broader questions, including the following: "What is a large enough number of clusters to use frequentist methods?" While general answers to this question probably do not exist, some careful Monte Carlo studies could shed some light on the issue. However, as long as one is willing to posit prior beliefs about the parameters of interest and can write down an algorithm to simulate from the posterior density, the Bayesian approach provides a satisfactory solution.

### CONCLUSION

The analysis presented above demonstrates the usefulness of Bayesian estimation using MCMC for the analysis of clustered, heterogeneous data. The model above, however, is just the tip of the iceberg. Indeed, there are many more sophisticated models one can use to model all sorts of contextual effects. Not only can these models be generalized with more levels of analysis (e.g., looking at roll call voting by members of every state legislature for a 50-year time period), but contextual differences could also be modeled across states *and* across time using three levels of analysis. One could also develop more elaborate cross-cutting clustering schemes or more elaborate hyperstructures that include autoregressive components and allow for various types of curve fitting for the hyperparameters using splines and nonparametric forms (see, e.g., Berry et al. 1999). One could also simultaneously model within-cluster heterogeneity and across-cluster heterogeneity using both strategies outlined above in tandem.

The advantage of the hierarchical modeling strategy is that it allows social scientists to systematically answer questions that were previously unanswerable using standard statistical methods. Indeed, context affects all manner of political behavior. Hierarchical models allow the researcher to systematically evaluate these explanations. It is also important to note that the methodology presented herein does not constrain the practitioner. The only constraints are the data and the imagination of the researcher. When modeling event counts or dealing with a small number of clusters, Bayesian approaches are preferred over other alternatives.

To conclude, it is worth mentioning other models one can use when dealing with other types of heterogeneous event counts. In some situations, one is faced with multivariate count data—that is, observing a vector of counts  $y_i$  for each observational unit that are likely to be related. An example would be modeling the number of cosponsorships and the number of discharge petitions signed by a single member of Congress. King (1989a) offers a seemingly unrelated Poisson regression model that is useful for modeling multivariate counts. Chib and Winkelmann (1999) discuss a flexible Bayesian estimation strategy for such counts. In some applications, observed counts do not follow a Poisson distribution but rather include a disproportionate number of zeros. Substantively, this is indicative of a threshold effect. To model this sort of data, one must turn to the class of zero-inflated models, which is essentially a mixture of a binomial choice model and an event count regression. These models, reviewed by Zorn (1998), are useful when modeling event counts with an inordinate number of zeros. Finally, when dealing with time-series event counts, one must be mindful of dealing with the problem of persistence. Brandt, Williams, Fordham, and Pollins (2000) offer a time-series regression model called the PEWMA model, which is particularly useful when dealing with time series.

## *APPENDIX*

### *BUGS Code*

This appendix contains BUGS code for all models presented in the text. These programs can be easily modified to fit models with

other data. All programs are available in electronic form at the replication Web site (as both plain text files and as WinBUGS ODC files), along with the data set employed. Since BUGS code is somewhat dependent on the platform used, here I present the key model definition statements for the models presented in this article. The data input statements and the specification of starting values differ slightly across platforms. All of the models presented here were estimated using WinBUGS 1.3. BUGS minimizes the start-up cost of performing Bayesian inference by alleviating the need to program in a high-level language and choosing an appropriate MCMC algorithm given the problem at hand. BUGS also comes with a suite of routines one can use with R and S-Plus to test for convergence called CODA (Best, Cowles, and Vines 1997). For a comprehensive treatment of how to use BUGS, compile and estimate models, and test for convergence, consult (2000) Spiegelhalter et al. (1997) and Best et al. (1997).

#### *POISSON REGRESSION WITH POSTERIOR PREDICTION*

The first model is the Poisson regression model with the estimation of posterior predictive distributions. The results from this model appear in Table 1 and Figure 1. Note that I use the same model definition notation used in the article. Since the posterior predictive distributions are just Monte Carlo averages over the posterior density of  $\beta$ , they too can be simulated within the sampler.

```
model cosponsor;
{
for (i in 1:N)
{
log(mu[i]) <-
(beta[1] * cons[i] +
beta[2] * dnom[i] +
beta[3] * party[i] +
beta[4] * gender[i]) ;
cospon[i] ~ dpois(mu[i]);
}

# Priors
for(j in 1:K)
```

```
{  
  beta[j] ~ dnorm(0.0, 0.0001);  
}  
  
# Posterior Predictive Distributions  
pred[1] <- exp(beta[1] + -0.627 * beta[2]  
  + 0 * beta[3] + 1 * beta[4]); # libRW  
pred[2] <- exp(beta[1] + -0.027 * beta[2]  
  + 0 * beta[3] + 1 * beta[4]); # modRW  
pred[3] <- exp(beta[1] + 0.573 * beta[2]  
  + 0 * beta[3] + 1 * beta[4]); # conRW  
pred[4] <- exp(beta[1] + -0.627 * beta[2]  
  + 1 * beta[3] + 1 * beta[4]); # libDW  
pred[5] <- exp(beta[1] + -0.027 * beta[2]  
  + 1 * beta[3] + 1 * beta[4]); # modDW  
pred[6] <- exp(beta[1] + 0.573 * beta[2]  
  + 1 * beta[3] + 1 * beta[4]); # conDW  
pred[7] <- exp(beta[1] + -0.627 * beta[2]  
  + 0 * beta[3] + 0 * beta[4]); # libRM  
pred[8] <- exp(beta[1] + -0.027 * beta[2]  
  + 0 * beta[3] + 0 * beta[4]); # modRM  
pred[9] <- exp(beta[1] + 0.573 * beta[2]  
  + 0 * beta[3] + 0 * beta[4]); # conRM  
pred[10] <- exp(beta[1] + -0.627 * beta[2]  
  + 1 * beta[3] + 0 * beta[4]); # libDM  
pred[11] <- exp(beta[1] + -0.027 * beta[2]  
  + 1 * beta[3] + 0 * beta[4]); # modDM  
pred[12] <- exp(beta[1] + 0.573 * beta[2]  
  + 1 * beta[3] + 0 * beta[4]); # conDM  
}
```

To allow for overdispersion, one only needs to add the random effect to the  $\log(\mu[i])$  term:

```
lambda[i];
```

To define its distribution, one needs to add the following within the model definition loop:

```
lambda[i] ~ dnorm(0.0, tau);
```

Finally, to place a prior on the precision, one should add the following in the section that defines the priors:

```
tau ~ dgamma(0.001,0.001);
```

#### *HIERARCHICAL POISSON REGRESSION*

The final code I present is from the hierarchical Poisson regression model. Here the data are loaded from three files: one that contains the Level 1 covariates sorted by Congress, one that contains the number of observations  $N_k$  within each cluster, and one that contains the contextual (second-level) covariates. This code begins with the first level, assuming a Poisson regression within each cluster. Then, across each cluster  $k$ , I assume independent regressions are noted in the text. The code concludes by placing priors on all of the hyperparameters.

```
model hierarchical;
{
# Model Level One
for (k in 1:K)
  {
  for (i in lower[k]:upper[k])
    {
    log(mu[i]) <-
      (beta[1,k] * cons[i] +
       beta[2,k] * dnom[i] +
       beta[3,k] * party[i] +
       beta[4,k] * gender[i]);
    cospon[i] ~ dpois(mu[i]);
    }
  }
}
# Model Level Two
for (k in 1:K)
  {
  nu[1,k] <- alpha1[1] + alpha1[2] * era1[k]
    + alpha1[3] * era2[k];
  nu[2,k] <- alpha2[1] + alpha2[2] * hsdnom[k]
    + alpha2[3] * ptydist[k];
```

```
nu[3,k] <- alpha3[1];
nu[4,k] <- alpha4[1];
for (j in 1:J)
  {
    beta[j,k] ~ dnorm(nu[j,k],gamma[j]);
  }
}

# Priors
for(m1 in 1: M1)
  {
    alpha1[m1] ~ dnorm(0.0, 0.0001);
  }
for(m2 in 1: M2)
  {
    alpha2[m2] ~ dnorm(0.0, 0.0001);
  }
for(m3 in 1: M3)
  {
    alpha3[m3] ~ dnorm(0.0, 0.0001);
  }
for(m4 in 1: M4)
  {
    alpha4[m4] ~ dnorm(0.0, 0.0001);
  }
for (j in 1:J)
  {
    gamma[j] ~ dgamma(0.001,0.001);
    igamma[j] <- 1/gamma[j];
  }
}
```

## NOTES

1. Observations are said to be exchangeable if the model results are not changed by relabeling the data values. This implies that the data are generated conditional on the model parameters in precisely the same way for each observation (Gill 2002).

2. Other priors could just as easily be applied. One could assume a multivariate Normal prior  $\beta \sim \mathcal{N}_J(\beta_0, B_0^{-1})$  or an improper uniform prior on  $\mathbb{R}$ . Unless using informative priors, this choice usually makes very little difference in practice. If it does make a difference in a particular application, the prior, not the likelihood, is driving the inference.

3. In addition, for all of the models presented below, I tested for sensitivity of the results to various prior specifications. For all models, the substantive conclusions remain the same given other prior specifications, including reasonable informative priors. I have also tested for convergence using the Geweke (1992) and Heidelberger and Welch (1981) diagnostics implemented in CODA (Best, Cowles, and Vines 1997).

4. This software is available on the Web for Windows, Unix, and Linux from the MRC Biostatistics Unit and the Imperial College School of Medicine at St. Mary's, London, at <http://www.mrc-bsu.cam.ac.uk/bugs/>. There are many helpful examples at this Web site. Gilks, Richardson, and Spiegelhalter (1996) also include some examples of models estimated with BUGS (Bayesian inference Using Gibbs Sampling).

5. As noted below in the application, the inferences drawn about the amount of heterogeneity in the data are essentially the same when mixing the Poisson with a Gamma or a Normal distribution. The choice to use the Normal distribution is made for computational reasons; the current WinBUGS release does not include support for the negative binomial distribution.

6. Clearly, there are measures that better capture women's right preferences (see Martin and Wolbrecht 2000). However, I employ these scores here because they are available for a long period of time.

7. In the classical paradigm, all parameters are fixed but unobservable. Thus, the probability that a parameter is positive equals 0 or 1, yet it cannot be observed. This requires mental gymnastics to interpret confidence intervals. The 95 percent credible intervals, on the other hand, can simply be interpreted as intervals where the (random) parameter falls 95 percent of the time.

8. Due to the large sample size and the noninformative priors, the posterior means are almost the same as the maximum likelihood estimates. So too is the estimated amount of overdispersion; for a negative binomial regression, the estimated  $\alpha$  parameter is 1.63.

9. A more general model would be to model the distribution of the  $\beta_k$  vectors with a multivariate regression model:  $\beta_k \sim \mathcal{N}_j(A_k\alpha, \Omega^{-1})$ . However, when there are many first-level parameters and few clusters, it becomes difficult, if not impossible, to estimate  $\Omega$  with precision. In the application below, with 4 first-level covariates and 20 contexts,  $\Omega$  has 10 free elements, which would be impossible to reliably estimate.

10. This model explicitly ignores the the temporal ordering of the clusters. Since every member of the House stands for reelection every two years, it is not unreasonable to assume that, conditional on the second-level covariates, each Congress is exchangeable, and each member within each Congress is exchangeable. Given more data, it would be possible to explore the time-series error structure (Brandt, Williams, Fordham, and Pollins 2000).

11. These results are available on the replication Web site.

12. These results are also available on the replication Web site.

13. To test for convergence, I employed the Geweke (1992) and Heidelberger and Welch (1981) diagnostics implemented in CODA (Best et al., 1997).

14. I thank an anonymous reviewer for this suggestion.

15. It is important to note that the weighted least squares approach accounts for the *relative* amount of uncertainty between each observation, but it does not account for the overall amount of uncertainty.

## REFERENCES

- Albert, James. 1992. "A Bayesian Analysis of a Poisson Random Effects Model for Home Run Hitters." *The American Statistician* 46:246-53.
- Bartels, Larry M. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40:905-42.

- Berry, Scott M., C. Shane Reese, and Patrick D. Larkey. 1999. "Bridging Different Eras in Sports." *Journal of the American Statistical Association* 94:661-86.
- Best, Nicky, Mary Kathryn Cowles, and Karen Vines. 1997. *Convergence Diagnostics and Output Analysis Software for Gibbs Sampling Output, Version 0.4*. Cambridge, UK: MRC Biostatistics Unit.
- Brandt, Patrick T., John T. Williams, Benjamin O. Fordham, and Brian Pollins. 2000. "Dynamic Modeling for Persistent Event Count Time Series." *American Journal of Political Science* 44:823-43.
- Breslow, N. E. 1984. "Extra-Poisson Variation in Log-Linear Models." *Applied Statistics* 33:38-44.
- Breslow, N. E. and D. G. Clayton. 1993. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88:9-25.
- Cameron, A. Colin and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press.
- Campbell, James E. 1982. "Cosponsoring Legislation in the U.S. Congress." *Legislative Studies Quarterly* 7:415-22.
- Chib, Siddhartha, Edward Greenberg, and Rainer Winkelmann. 1998. "Posterior Simulation and Bayes Factors in Panel Count Data Models." *Journal of Econometrics* 86:33-54.
- Chib, Siddhartha and Rainer Winkelmann. 2001. "Markov Chain Monte Carlo Analysis of Correlated Count Data." *Journal of Business and Economic Statistics* 19:428-35.
- Christiansen, Cindy L. and Carl N. Morris. 1997. "Hierarchical Poisson Regression Modeling." *Journal of the American Statistical Association* 92:618-32.
- Cox, Gary W. and Mathew D. McCubbins. 1993. *Legislative Leviathan: Party Government in the House*. Berkeley: University of California Press.
- Efron, B. and C. Morris. 1973. "Combining Possibly Related Estimation Problems." *Journal of the Royal Statistical Society B* 35:379-421.
- Gelfand, Alan E. and Adrian F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85:398-409.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman & Hall.
- Geman, S. and D. Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions of Pattern Analysis and Machine Intelligence* 6:721-41.
- Geweke, John. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments." In *Bayesian Statistics 4*, edited by J. M. Bernardo, J. O. Berger, and A. F. M. Smith. Oxford, UK: Clarendon.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.
- Gill, Jeff. 2002. *Bayesian Methods for the Social and Behavioral Sciences*. London: Chapman & Hall.
- Greenwood, Major and G. Udny Yule. 1920. "An Inquiry Into the Nature of Frequency Distributions Representative of Multiple Happenings With Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents." *Journal of the Royal Statistical Society* 83:255-79.
- Heidelberger, P. and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulation." *Communications of the ACM* 24:233-45.
- Jones, Bradford S. and Marco R. Steenbergen. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46:218-37.



- Kessler, Daniel and Keith Krehbiel. 1996. "Dynamics of Cosponsorship." *American Political Science Review* 90:555-66.
- King, Gary. 1988. "Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model." *American Journal of Political Science* 32:838-63.
- . 1989a. "A Seemingly Unrelated Poisson Regression-Model." *Sociological Methods and Research* 17:235-55.
- . 1989b. "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator." *American Journal of Political Science* 33:762-84.
- King, Gary, Ori Rosen, and Martin A. Tanner. 1999. "Binomial-beta Hierarchical Models for Ecological Inference." *Sociological Methods and Research* 28:61-90.
- Martin, Andrew D. 2001. "Congressional Decision Making and the Separation of Powers." *American Political Science Review* 95:361-78.
- Martin, Andrew D. and Christina Wolbrecht. 2000. "Partisanship and Pre-Floor Behavior: The Equal Rights and School Prayer Amendments." *Political Research Quarterly* 53:711-30.
- Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. Oxford, UK: Oxford University Press.
- Raudenbush, Stephen W. and Anyhony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage.
- Raudenbush, Stephen W., Anthony S. Bryk, and Richard Congdon. 2000. *HLM 5: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.
- Rubin, Donald B. 1984. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician." *Annals of Statistics* 12:1151-72.
- Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." *Science* 277:918-24.
- Schiller, Wendy J. 1995. "Senators as Political Entrepreneurs: Bill Sponsorship to Shape Legislative Agendas." *American Journal of Political Science* 39:186-203.
- Spiegelhalter, David J., Andrew Thomas, Nicky Best, and Wally R. Gilks. 2000. *WinBUGS 1.3: Bayesian Inference Using Gibbs Sampling*. Cambridge, UK: MRC Biostatistics Unit.
- Tanner, M. A. and W. Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82:528-50.
- Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *American Journal of Political Science* 42:1233-59.
- Western, Bruce and Simon Jackman. 1994. "Bayesian Inference for Comparative Research." *American Political Science Review* 88:412-23.
- Wolbrecht, Christina. 2000. *The Politics of Women's Rights: Parties, Positions, and Change*. Princeton, NJ: Princeton University Press.
- Young, Cheryl D. and Rick K. Wilson. 1997. "Copponsorship in the United States Congress." *Legislative Studies Quarterly* 22:25-43.
- Zorn, Christopher J. W. 1998. "An Analytic and Empirical Examination of Zero-Inflated and Hurdle Poisson Specifications." *Sociological Methods and Research* 26:368-400.

*Andrew D. Martin is an assistant professor of political science at Washington University in St. Louis. He holds a Ph.D. in political science from Washington University and an A.B. in mathematics from the College and William and Mary. He specializes in political*

*methodology and Bayesian statistics. His substantive area of research is American political institutions, with work appearing in the American Political Science Review, American Journal of Political Science, Political Analysis, Statistical Science, and other journals.*

## **Request Permission or Order Reprints Instantly**

Interested in copying, sharing, or the repurposing of this article? U.S. copyright law, in most cases, directs you to first get permission from the article's rightsholder before using their content.

To lawfully obtain permission to reuse, or to order reprints of this article quickly and efficiently, click on the "Request Permission/ Order Reprints" link below and follow the instructions. For information on Fair Use limitations of U.S. copyright law, please visit [Stamford University Libraries](#), or for guidelines on Fair Use in the Classroom, please refer to [The Association of American Publishers' \(AAP\)](#).

All information and materials related to SAGE Publications are protected by the copyright laws of the United States and other countries. SAGE Publications and the SAGE logo are registered trademarks of SAGE Publications. Copyright © 2003, Sage Publications, all rights reserved. Mention of other publishers, titles or services may be registered trademarks of their respective companies. Please refer to our user help pages for more details: <http://www.sagepub.com/cc/faq/SageFAQ.htm>

**[Request Permissions / Order Reprints](#)**