

Article

Inference with Pólya-Gamma Augmentation for US Election Law

Adam C. Hall [†] and Joseph Kang ^{*,†}

US Census Bureau, 4600 Silver Hill Rd., Washington, DC 20233, USA; adam.c.hall@census.gov

* Correspondence: joseph.kang@census.gov

† These authors contributed equally to this work.

Abstract: Pólya-gamma (PG) augmentation has proven to be highly effective for Bayesian MCMC simulation, particularly for models with binomial likelihoods. This data augmentation strategy offers two key advantages. First, the method circumvents the need for analytic approximations or Metropolis–Hastings algorithms, which leads to simpler and more computationally efficient posterior inference. Second, the approach can be successfully applied to several types of models, including nonlinear mixed-effects models for count data. The effectiveness of PG augmentation has led to its widespread adoption and implementation in statistical software packages, such as version 2.1 of the R package BayesLogit. This success has inspired us to apply this method to the implementation of Section 203 of the Voting Rights Act (VRA), a US law that requires certain jurisdictions to provide non-English voting materials for specific language minority groups (LMGs). In this paper, we show how PG augmentation can be used to fit a Bayesian model that estimates the prevalence of each LMG in each US voting jurisdiction, and that uses a variable selection technique called stochastic search variable selection. We demonstrate that this new model outperforms the previous model used for 2021 VRA data with respect to model diagnostic measures.

Keywords: small area estimation; Pólya-gamma augmentation; stochastic search variable selection; Bayesian regression; Gibbs sampling; Voting Rights Act

MSC: 62F15



Academic Editor: Jose Luis Vicente Villardon

Received: 7 February 2025

Revised: 6 March 2025

Accepted: 11 March 2025

Published: 13 March 2025

Citation: Hall, A.C.; Kang, J. Inference with Pólya-Gamma Augmentation for US Election Law. *Mathematics* **2025**, *13*, 945. <https://doi.org/10.3390/math13060945>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. A Motivating Example

Section 203(b) of the Voting Rights Act (VRA) is an election law that requires some US states or their political subdivisions to make voting materials available in languages other than English. The US Census Bureau determines whether a jurisdiction, e.g., county, is subject to such requirements based on the rates of limited English language proficiency and illiteracy within designated population subgroups, called language minority groups (LMGs), in that jurisdiction. For the purposes of VRA coverage determinations, limited English proficiency (LEP) is defined as speaking a language other than English at home, and speaking English less than “Very Well”. Illiteracy (ILL) is defined as having less than a 5th-grade education. A more detailed discussion of the legal and statistical criteria used to make VRA coverage determinations is available in Slud et al. [1].

To make these determinations, the Census Bureau must estimate the number of voting-age (18 or over) people (VOT), voting-age citizens (CIT), voting-age citizens who have limited English proficiency (LEP), and voting-age citizens with limited English proficiency who are illiterate (ILL) in each of these geographies. The Census Bureau initially used “direct” estimates based on data from the long-form of the Decennial Census (later replaced

by the American Community Survey, or ACS) to approximate these quantities. The term “direct” indicates that an estimate does not use any statistical model or auxiliary information beyond the survey weights and observations within that area. However, since 2011, Section 203 determinations have been based on estimates from small-area models. These statistical models were updated in 2014 by Joyce et al. [2], who developed a Bayesian hierarchical model, and Slud et al. [3], who adopted a Dirichlet-multinomial model.

In the VRA 2021 determinations, model-based estimates were obtained using both frequentist and Bayesian methods [1]. The frequentist method used adaptive Gaussian quadrature [4] to fit random effects models for LMGs with a sample in less than 700 jurisdictions (counties and MCDs), or less than 3000 jurisdictions where most citizens did not have severe illiteracy. The Bayesian model was fitted using STAN [5], which only worked well for large LMGs. Models fit with STAN had relatively slow and unstable convergence in smaller LMGs even when using priors recommended for computational efficiency, such as the LKJ prior for the random-effects covariance distribution. In addition, both methods used covariates that were selected based on their performance in a few large LMGs to model all LMGs, rather than performing independent variable selection procedures in the models for smaller groups. More details on the VRA 2021 methods are described in Slud et al. [1].

Table 1 shows a dataset that contains hypothetical examples of jurisdiction-level estimates that VRA models might produce for an LMG. For this LMG, let VOT_j , CIT_j , LEP_j , and ILL_j represent VOT, CIT, LEP, and ILL, respectively, in jurisdiction j . Then, in Table 1, CITprop, LEPprop, and ILLprop represent the direct estimates of the ratios CIT_j/VOT_j , LEP_j/VOT_j , and ILL_j/VOT_j , respectively. The model-based predictions of these ratios are represented by π_{CIT} , π_{LEP} , and π_{ILL} , respectively.

Table 1. Hypothetical VRA proportions for an LMG.

Jurisdiction	# of Voting-Aged Persons	CITprop	LEPprop	ILLprop	π_{CIT}	π_{LEP}	π_{ILL}	X_j
Jurisdiction 1	1000	0.80	0.10	0.01	0.80	0.10	0.01	X_1
Jurisdiction 2	10	0.95	0.25	0.02	0.92	0.19	0.07	X_2
...
Jurisdiction N	1	1	0	0	0.93	0.10	0.01	X_N

Note that the model-based estimates are similar to the direct estimates for jurisdiction 1, which has a large number of voting-age residents. However, these estimates differ in jurisdictions with fewer voting-age persons. This pattern occurs because direct estimates perform well in large samples, but poorly when sample sizes are small. Model-based estimates use vectors of auxiliary covariates, denoted in Table 1 by X , to help improve their low sample performance.

The new model introduced in this article uses Pólya-gamma (PG) augmentation [6] to estimate Bayesian models for all LMGs and jurisdictions. This enables the new model to more easily implement stochastic search variable selection (SSVS) within each LMG to select covariates, which allows our new model to outperform the older models in terms of diagnostic metrics.

1.2. Utility of Pólya-Gamma Augmentation in Bayesian Methods

Pólya-gamma (PG) augmentation, as introduced by Polson et al. [6], is an efficient method for Bayesian inference in models with binomial likelihoods. The application of PG augmentation yields closed-form expressions for full conditionals of posterior distributions in sampling algorithms like Markov chain Monte Carlo (MCMC), which eliminates the need for analytic approximations, numerical integration, or Metropolis–Hastings algorithms. This allows the construction of simple and effective Gibbs samplers for models with bino-

mial likelihoods, particularly logistic regression. The method has been successfully applied to various models, including logistic regression and nonlinear mixed-effects models [6]. In practice, the PG Gibbs sampler for Bayesian logistic regression involves two main steps in each iteration: sampling PG random variables and drawing from a multivariate normal distribution. This approach has been implemented in various statistical software packages, such as Version 2.1 of the R package BayesLogit [7].

The remainder of this paper is organized into the following sections. Section 2 reviews the 2021 VRA model, introduces PG augmentation, and explains a basic application in the context of the VRA data. Section 3 builds on this foundation to introduce the new Bayesian multinomial logit model, and describes a Gibbs sampling algorithm that makes use of PG augmentation to fit it. Section 4 applies the new model to the data used for the 2021 VRA determinations and shows its superior diagnostic results compared with the 2021 model results. Finally, Section 5 discusses the results.

2. The Pólya-Gamma Distribution for Binomial VRA Datasets

2.1. The 2021 VRA Model

The 2021 VRA model was constructed by modeling the counts mentioned in Section 1 using a multinomial distribution. Specifically, for each jurisdiction j where $j \in \{1, \dots, 7859\}$, and n_j is the number of voting-age people in j , we model

$$M_{j,1}, M_{j,2}, M_{j,3}, M_{j,4} \sim \text{Multinom}(n_j, (p_{j,1}, p_{j,2}, p_{j,3}, p_{j,4})), \tag{1}$$

where $M_{j,1}, M_{j,2}, M_{j,3}$, and $M_{j,4}$ are $VOT_j - CIT_j, CIT_j - LEP_j, LEP_j - ILL_j$, and ILL_j , respectively.

An equivalent formulation of this model treats the M variables in Equation (1) as the composition of three conditionally binomial (*Binom*) models:

$$\begin{aligned} M_{j,1} &\sim \text{Binom}(n_j, p_{j,1}) \\ M_{j,2} &\sim \text{Binom}\left(n_j - M_{j,1}, \frac{p_{j,2}}{1 - p_{j,1}}\right) \text{ given } M_{j,1} \\ M_{j,3} &\sim \text{Binom}\left(n_j - M_{j,1} - M_{j,2}, \frac{p_{j,3}}{1 - p_{j,1} - p_{j,2}}\right) \text{ given } M_{j,1}, M_{j,2}, \end{aligned} \tag{2}$$

The intuition behind the success probabilities in these distributions is as follows. In the multinomial model, $p_{j,1}$ represents the marginal probability that a voting-age person is not a citizen. Thus, if we ignore the other categories, $M_{j,1}$ can be modeled as a binomial distribution counting the number of voting-age non-citizens against the number of voting-age citizens. Similarly, the marginal probability that a person is a voting-age citizen and is not LEP is $p_{j,2}$ in the multinomial model. It can be inferred that $1 - p_{j,1}$ is the marginal probability that a voting-age person is a citizen. Thus, the probability of a voting-age person not having limited English proficiency conditional on being a citizen is $p_{j,2}/(1 - p_{j,1})$. By the same logic, the marginal probability that a voting-age person is a citizen and LEP but not ILL is $p_{j,3}$ in the multinomial model, and the marginal probability that a voting-age person is a citizen and is LEP is $1 - p_{j,1} - p_{j,2}$. Thus, the conditional probability of a voting-age person not being ILL given that they are a citizen and have LEP is $p_{j,3}/(1 - p_{j,1} - p_{j,2})$. $M_{j,4}$ is known given $M_{j,1}, M_{j,2}, M_{j,3}$, and n_j , and thus does not need to be modeled.

The complements of the “success” probabilities of the binomial distributions for $M_{j,1}, M_{j,2}$, and $M_{j,3}$ can be interpreted as estimates of the ratios $v_{j,1} = CIT_j/VOT_j$,

$v_{j,2} = LEP_j/CIT_j$, and $v_{j,3} = ILL_j/LEP_j$ in jurisdiction j . With this notation, the probabilities in Equation (1) can be written for each jurisdiction j as

$$(p_{j,1}, p_{j,2}, p_{j,3}, p_{j,4}) \equiv (1 - v_{j,1}, v_{j,1}(1 - v_{j,2}), v_{j,1}v_{j,2}(1 - v_{j,3}), v_{j,1}v_{j,2}v_{j,3})$$

This means that we can model the multinomial success probabilities in (1) by modeling the binomial success probabilities in (2), and we can model the binomial success probabilities in (2) by modeling the $v_{j,k}$ ratios. To model the $v_{j,k}$ ratios, we assume that

$$v_{j,k} = \frac{\exp(\eta_{j,k})}{1 + \exp(\eta_{j,k})}, \tag{3}$$

where

$$\eta_{j,k} \equiv X_j^T \beta_k + u_{j,k}$$

and the u are multivariate normal random variables with a zero mean and variance parameter

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}. \tag{4}$$

This specification is called the “multinomial logit normal” (MLN) model due to the multinomial outcome described in Equation (1), the logistic link function in Equation (3), and the multivariate normal random effects. When these parameters are estimated, we can use them to derive model-based estimates of π_{CIT} , π_{LEP} , and π_{ILL} similar to the hypothetical examples in Table 1. Specifically, for jurisdiction j , we estimate π_{CIT} as $v_{j,1}$, π_{LEP} as $v_{j,1} \times v_{j,2}$, and π_{ILL} as $v_{j,1} \times v_{j,2} \times v_{j,3}$. The difficulty of estimating these parameters stems from the nonlinearity of the logistic link function in Equation (3). The Pólya-gamma augmentation makes the estimation process linear, resulting in closed-form solutions for the posterior distributions of our parameters.

2.2. The Pólya-Gamma (PG) Distribution

A PG random variable is defined as an infinite weighted sum of gamma random variables. Specifically, if $\omega \sim PG(b, c)$, where $b > 0$ and $c \in \mathbb{R}$, then:

$$\omega \stackrel{d}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 0.5)^2 + \frac{c^2}{4\pi^2}}$$

where $\stackrel{d}{=}$ denotes equality in distribution, and $g_k \sim Gamma(b, 1)$ are independent gamma random variables. Polson et al. [6] proved the following useful property of the PG ω variable. Let ψ denote a constant, e.g., a linear combination of covariates X and their coefficients β . Then, the following equation holds for an exponential functional form:

$$\frac{(e^\psi)^a}{(1 + e^\psi)^b} = 2^{-b} e^{\kappa\psi} \int_0^\infty e^{-\omega\psi^2/2} P(\omega) d\omega \tag{5}$$

where $\kappa = a - b/2$ and $p(\omega) = PG(\omega|b, 0)$. The utility of the above equation is that for $j = 1, \dots, N$, we can express the contribution of jurisdiction j to the likelihood as

$$\begin{aligned} \mathcal{L}_j(\beta) &= \frac{(e^{\psi_j})^{y_j}}{(1 + e^{\psi_j})^{n_j}} \\ &= 2^{-n_j} e^{\kappa_j \psi_j} \int_0^\infty e^{-\omega_j \psi_j^2 / 2} p(\omega_j) d\omega_j, \end{aligned} \tag{6}$$

where $\kappa_j = y_j - n_j/2$. Note that if we condition on $\Omega = \text{diag}(\omega_1, \dots, \omega_N)$, the kernel of the posterior distribution for β looks similar to a Gaussian density function:

$$P(\beta|\Omega, y, n) = P(\beta) \prod_{j=1}^N \mathcal{L}_j(\beta|\omega_j) \tag{7}$$

$$\propto P(\beta) \prod_{j=1}^N e^{\kappa_j \psi_j} e^{-\omega_j \psi_j^2 / 2} \tag{8}$$

$$\propto P(\beta) \prod_{j=1}^N \exp\left\{-\frac{\omega_j}{2} \left(\psi_j - \frac{\kappa_j}{\omega_j}\right)^2\right\} \tag{9}$$

$$= P(\beta) \exp\left\{-\frac{1}{2}(Z - X\beta)^T \Omega (Z - X\beta)\right\} \tag{10}$$

where $Z = (\kappa_1/\omega_1, \dots, \kappa_N/\omega_N)$. Step (8) is justified because conditional on ω_j , the expectation in (6) is constant. Step (9) is derived by completing the square. From this, Step (10) can be derived using algebra.

In summary, if our prior on β is Gaussian (quadratic in β), then (7) has a closed-form solution. This enables us to construct a Gibbs sampler for β , by repeatedly sampling $\Omega|\beta$ and then sampling $\beta|\Omega$.

2.3. A Simplified Application of Pólya-Gamma Augmentation

To illustrate the application of PG augmentation to the 2021 VRA model, consider jurisdiction j , for $j \in \{1, \dots, N\}$, where N denotes the total number of jurisdictions. Let n_j denote the number of voting-age citizens, v_j denote the probability that a voting-age citizen is LEP, and y_j denote the number of voting-age citizens with LEP. For simplicity, we do not include random effects in the example for this section. Then, PG augmentation as introduced by Polson et al. [6] can be applied to write the probability of observing y_j out of n_j as

$$(1 - v_j)^{n_j - y_j} v_j^{y_j} = \frac{\exp(\psi_j)^{y_j}}{(1 + \exp(\psi_j))^{n_j}} \propto N(\kappa_j/\omega_j | X_j \beta, \omega_j^{-1}), \tag{11}$$

where $\kappa_j = (y_j - n_j/2)$, $\omega_j \sim PG(n_j, \psi_j)$, and $\psi_j = X_j \beta$. Following Polson et al. [6], we let $Z = (\kappa_1/\omega_1, \dots, \kappa_N/\omega_N)$, which allows us to write the β posterior distribution as

$$\begin{aligned} P(\beta|y, \omega) &\propto P(\beta) \prod_{j=1}^N L_j(\beta|\omega_j) \propto P(\beta) \exp\left\{-\frac{1}{2}(Z - X\beta)^T \Omega (Z - X\beta)\right\} \\ &= N(\beta|b, B) N(Z|X\beta, \Omega^{-1}) \\ &= N(\beta|m_\omega, V_\omega), \end{aligned}$$

where $X = [X_1^T, \dots, X_N^T]^T$, $\kappa = [\kappa_1, \dots, \kappa_N]^T$, $\Omega = \text{diag}(\omega_1, \dots, \omega_N)$, $V_\omega = (X^T \Omega X + B^{-1})^{-1}$, and $m_\omega = V_\omega (X^T \kappa + B^{-1}b)$ and a standard normal distribution $N(b, B)$ for the β prior. Thus, a simple Gibbs sampling algorithm to fit this model iterates:

$$(\omega_j|\beta) \sim PG(n_j, X_j^T \beta)$$

$$(\beta|y_j, n_j, \omega_j) \sim N(m_\omega, V_\omega).$$

This Gibbs sampling algorithm was implemented using the R package BayesLogit [7] and provides closed-form equations for the β posterior. This allows users to adopt methods developed for linear regressions, including the Bayesian model selection procedure [8]. Appendix A explains the derivation of the closed-form β posterior distribution.

3. New MLN Model with PG Augmentation

3.1. Basic Setup

In this section, we introduce the model specification for our new Bayesian VRA model. This new model is distinguished by our adoption of stochastic search variable selection [8], which is the most important difference between it and previous VRA models. However, the new model retains the multinomial logit normal structure from the 2021 VRA model [1] described in Section 2.1. In particular, we can write this model as a composition of three conditional binomial distributions with success probabilities $v_1, v_2,$ and v_3 . Table 2 summarizes the relationships between the multinomial category proportions $p_1, p_2, p_3,$ and p_4 and the modeled ratios $v_1, v_2,$ and v_3 in this model. It also shows the contribution that each jurisdiction, American Indian Area (AIA), and/or Alaskan Native Regional Corporation (ANRC) makes to the log-likelihood when fitting this statistical model.

Table 2. Likelihood contributions.

Binomial Prob.	Ratio	Equivalent Prob.	Log-Likelihood Contribution for a Jurisdiction
v_1	CIT/VOT	$1 - p_1$	$M_1 \log(p_1) + (n - M_1) \log(1 - p_1)$
v_2	LEP/CIT	$1 - \frac{p_2}{1-p_1}$	$M_2 \log\left(\frac{p_2}{1-p_1}\right) + (n - M_1 - M_2) \log\left(1 - \frac{p_2}{1-p_1}\right)$
v_3	ILL/LEP	$1 - \frac{p_3}{1-p_1-p_2}$	$M_3 \log\left(\frac{p_3}{1-p_1-p_2}\right) + (n - M_1 - M_2 - M_3) \log\left(1 - \frac{p_3}{1-p_1-p_2}\right)$

Note that the log-likelihood contribution for v_k compares M_k with $\sum_{c=k+1}^4 M_c$ for $k = 1, 2, 3$.

3.2. Model Likelihood and Priors

Regression coefficients β , the random effects u , and the covariance matrix Σ are specified with the following likelihood contribution for the j th jurisdiction:

$$LH_j = (1 - v_{j,1})^{VOT_j - CIT_j} v_{j,1}^{CIT_j} (1 - v_{j,2})^{CIT_j - LEP_j} v_{j,2}^{LEP_j} (1 - v_{j,3})^{LEP_j - ILL_j} v_{j,3}^{ILL_j}$$

$$= (1 - v_{j,1})^{M_{j,1}} v_{j,1}^{\sum_{c=2}^4 M_{j,c}} (1 - v_{j,2})^{M_{j,2}} v_{j,2}^{\sum_{c=3}^4 M_{j,c}} (1 - v_{j,3})^{M_{j,3}} v_{j,3}^{M_{j,4}}$$

$$= \prod_{c=1}^3 \left[\frac{(\exp(\eta_{j,c}))^{\sum_{k=c+1}^4 M_{j,k}}}{(1 + \exp(\eta_{j,c}))^{\sum_{k=c}^4 M_{j,k}}} \right],$$

Using this notation, the likelihood for all N jurisdictions multiplied by the prior distributions for the parameters can be expressed as

$$\prod_{j=1}^N LH_j \times \prod_{k=1}^3 P(\beta_k) P(\Sigma) P(U_{j,k} = u_{j,k}). \tag{12}$$

The Bayesian model in Equation (12) treats the random effects variables u as parameters to be estimated. In contrast, frequentist inference averages out random effects as “missing data” in the optimization process. However, the widely used frequentist adaptive Gaussian quadrature method [4] for fitting random effects models treats the random effects as parameters in the process of averaging them out. Thus, both Bayesian and frequentist inferences estimate random effects parameters from this technical perspective.

3.3. Conditional Posterior Distributions for Regression Coefficients and PG Variables

For each outcome $k \in \{1, 2, 3\}$ and jurisdiction $j \in \{1, \dots, 7859\}$, the regression coefficients β_k and PG random variables $\omega_{j,k}$ are sampled similarly to what was done in Polson et al. [6]. Recall the key result from Section 2 that, conditionally given ω , $\exp(\psi)^y / (1 + \exp(\psi))^n \propto N(\omega^{-1}(y - n/2) | \psi, \omega^{-1})$, which indicates a Gaussian normal distribution for $\omega^{-1}(y - n/2)$ with its mean ψ and variance ω^{-1} . Using this result it can be shown that, conditional on $\omega_{j,k}$, the posterior distributions for the β_k coefficients have the following closed-form representations:

- $VOT_j - CIT_j$ vs. CIT_j ($\sum_{k=1}^4 M_{j,k} - \sum_{k=2}^4 M_{j,k}$ vs. $\sum_{k=2}^4 M_{j,k}$)

$$\beta_1 \sim N(V_{\beta_1}^{-1} m_{\beta_1}, V_{\beta_1}^{-1}) \tag{13}$$

conditional on $\omega_{j,1}$ being drawn from $\omega_{j,1} \sim PG(\sum_{k=1}^4 M_{j,k}, \eta_{j,1}) \stackrel{d}{=} PG(VOT_j, \eta_{j,1})$ for each j .

- $CIT_j - LEP_j$ vs. LEP_j ($\sum_{k=2}^4 M_{j,k} - \sum_{k=3}^4 M_{j,k}$ vs. $\sum_{k=3}^4 M_{j,k}$)

$$\beta_2 \sim N(V_{\beta_2}^{-1} m_{\beta_2}, V_{\beta_2}^{-1}) \tag{14}$$

conditional on $\omega_{j,2}$ being drawn from $\omega_{j,2} \sim PG(\sum_{k=2}^4 M_{j,k}, \eta_{j,2}) \stackrel{d}{=} PG(CIT_j, \eta_{j,2})$ for each j .

- $LEP_j - ILL_j$ vs. ILL_j ($\sum_{k=3}^4 M_{j,k} - \sum_{k=4}^4 M_{j,k}$ vs. $\sum_{k=4}^4 M_{j,k}$)

$$\beta_3 \sim N(V_{\beta_3}^{-1} m_{\beta_3}, V_{\beta_3}^{-1}) \tag{15}$$

conditional on $\omega_{j,3}$ being drawn from $\omega_{j,3} \sim PG(\sum_{k=3}^4 M_{j,k}, \eta_{j,3}) \stackrel{d}{=} PG(LEP_j, \eta_{j,3})$ for each j .

Here, for each outcome category $k \in \{1, 2, 3\}$, we define: $\eta_{j,k} = X_j^T \beta_k + u_{j,k}$, $X = [X_1^T, \dots, X_N^T]^T$, $\kappa_{j,k} = \sum_{c=k+1}^4 M_{j,c} - \frac{1}{2} \sum_{c=k}^4 M_{j,c}$, $\kappa_k = [\kappa_{1,k}, \dots, \kappa_{N,k}]^T$, $\Omega_k = \text{diag}(\omega_{1,k}, \dots, \omega_{N,k})$, $U_k = [u_{1,k}^T, \dots, u_{N,k}^T]^T$, $V_{\beta_k} = X^T \Omega_k X + B^{-1}$, $m_{\beta_k} = X^T \kappa_k - X^T \Omega_k U_k + B^{-1} b$, and b and B are from the normal β_k prior $N(b, B)$. In general, the conditional β_k posterior for outcome category k is $P(\beta_k | X, \kappa_k, U_k, \Omega_k) \sim N(V_{\beta_k}^{-1} m_{\beta_k}, V_{\beta_k}^{-1})$. This posterior distribution is derived in Appendix B.

3.4. Conditional Posterior Distribution for the Variable Selection Parameters

George and McCulloch [8] developed a Bayesian variable selection procedure known as stochastic search variable selection (SSVS) in the context of linear regression. Recall that the regression in Equation (11) assumes the β prior distribution $N(b, B)$. To perform SSVS, we specify the prior distribution for β as a mixture of two normal distributions: one with a smaller variance to signify a coefficient close to zero, and the other with a larger variance to signify a coefficient farther away from zero. Specifically, George and McCulloch [8] introduced the

latent inclusion variable γ_i (1 if included, 0 otherwise) for the normal mixture prior distribution for the i th coefficient β_i ($i \in \{1, \dots, q\}$):

$$(1 - \gamma_i)N(0, \tau_i^2) + \gamma_i N(0, c_i^2 \tau_i^2). \tag{16}$$

The interpretation of this equation is as follows. When γ_i is 0, β_i is assumed to have $N(0, \tau_i^2)$, and when γ_i is 1, β_i is assumed to have $N(0, c_i^2 \tau_i^2)$. τ_i is set to be small so that if γ_i is 1, then β_i will be close to 0. c_i is set to be large ($c_i > 1$ always) so that if γ_i is 1, then a non-zero estimate of β_i will be included in the model. For specific choices of γ_i and c_i , George and McCulloch [8] recommended that users should consider varying the settings to extract more information rather than treat any particular settings as rules that guarantee good results.

To implement the normal mixture prior in Equation (16), the prior covariance matrix B from the prior distribution $N(b, B)$ is specified as $D_\gamma R D_\gamma$, where $\gamma = (\gamma_1, \dots, \gamma_q)$, R is the prior correlation matrix (an identity matrix), and

$$D_\gamma \equiv \text{diag}(a_1 \tau_1, \dots, a_q \tau_q),$$

with $a_i = 1$ if $\gamma_i = 0$ and $a_i = c_i$ if $\gamma_i = 1$. That is, $a_i \tau_i = \{(1 - \gamma_i) + \gamma_i c_i\} \times \tau_i$ and $a_i^2 \tau_i^2 = \{(1 - \gamma_i) + \gamma_i c_i^2\} \times \tau_i^2$. Thus, D_γ determines the scaling of the prior covariance matrix in such a way that Equation (16) is satisfied.

The conditional posterior distribution of $\gamma_i^{(m)}$ for the i th covariate at the m th MCMC sampling step is

$$P(\gamma_i^{(m)} = 1 | \beta^{(m-1)}, \gamma_{(i)}^{(m)}),$$

where $\gamma_{(i)}^{(m)} = (\gamma_1^{(m)}, \dots, \gamma_{i-1}^{(m)}, \gamma_{i+1}^{(m-1)}, \dots, \gamma_p^{(m-1)})$, and $\beta^{(m-1)}$ is the coefficients vector for the $(m - 1)$ th MCMC sample with $m \in \{1, \dots, 10^5\}$. Notice that the posterior for $\gamma_i^{(m)}$ does not depend on the rest of the parameters, including the outcomes. This yields a substantial simplification that reduces computational requirements and allows for faster convergence of the MCMC subsequence for all γ_i . Finally,

$$P(\gamma_i^{(m)} = 1 | \beta^{(m-1)}, \gamma_{(i)}^{(m)}) = \frac{a}{a + b},$$

where

$$a = P(\beta^{(m)} | \gamma_{(i)}^{(m)}, \gamma_i^{(m)} = 1) \times p_i$$

and

$$b = P(\beta^{(m)} | \gamma_{(i)}^{(m)}, \gamma_i^{(m)} = 0) \times (1 - p_i),$$

where p_i was chosen to be 0.5, implying equal weights for a and b .

3.5. Conditional Posterior Distribution for Random Effects Covariance Σ

Recall $U = (u_1, u_2, \dots, u_N)$ and $u_j = (u_{j,1}, u_{j,2}, u_{j,3})$. Σ was defined in Equation (4). Assuming U has a multivariate normal distribution $MVN(0, \Sigma)$, the conjugate Σ posterior is

$$IW\left(\sum_{j=1}^N u_j u_j^T + V_0, N + s_0\right),$$

where V_0 can be set to be $I_{3 \times 3}$, and s_0 can be set to 3, which is the number of rows of Σ . These are typical choices for the inverse Wishart priors. s_0 can be chosen to be bigger than 3,

but larger values of s_0 tend to attenuate the random effect covariance elements because the elements of the posterior mean $(\sum_{j=1}^N u_j u_j^T + V_0) / (N + s_0 - 3 - 1)$ will get closer to zero as s_0 increases. We keep s_0 at 3, and set the off-diagonal elements to zero. Slud et al. [1] called the multinomial logit model with this random effects covariance structure “MLN-D”. When the off-diagonal elements were not set to zero, Slud et al. [1] called the model “MLN-F”. Fitting MLN-D is generally more stable than fitting MLN-F, but this does not necessarily mean that one is better than the other in terms of diagnostic results.

3.6. Conditional Posterior Distribution for Random Effects Variables

Consider a Gaussian prior distribution $N(0, \Sigma)$ on $u_j = (u_{j,1}, u_{j,2}, u_{j,3})$. Then, the conditional posterior $P(u_j | \Sigma)$ can be shown to be $N(V_{j,u}^{-1} m_{j,u}, V_{j,u}^{-1})$, where $V_{j,u} = \Omega_j^* + \Sigma^{-1}$, $m_{j,u} = \Omega_j^* M_j^{*u}$, $\Omega_j^* = \text{diag}(\omega_{j,1}, \omega_{j,2}, \omega_{j,3})$, $M_{j,k}^{*u} = M_{j,k}^* - X_j \beta_k$, $M_{j,k}^* = \omega_{j,k}^{-1} (\sum_{c=k+1}^4 M_{j,c} - \frac{1}{2} \sum_{k=c}^4 M_{j,c})$, and $\eta_{j,k} \equiv X_j^T \beta_k + u_{j,k}$.

Note that $m_{j,u} = \Omega_j^* M_j^{*u}$ can be expressed as:

$$\begin{pmatrix} \omega_{j,1} & 0 & 0 \\ 0 & \omega_{j,2} & 0 \\ 0 & 0 & \omega_{j,3} \end{pmatrix} \times \begin{pmatrix} M_{j,1}^* - X_j \beta_1 \\ M_{j,2}^* - X_j \beta_2 \\ M_{j,3}^* - X_j \beta_3 \end{pmatrix} = \begin{pmatrix} \sum_{k=2}^4 M_{j,k} - \frac{1}{2} \sum_{k=1}^4 M_{j,k} - \omega_{j,1} \times X_j \beta_1 \\ \sum_{k=3}^4 M_{j,k} - \frac{1}{2} \sum_{k=2}^4 M_{j,k} - \omega_{j,2} \times X_j \beta_2 \\ \sum_{k=4}^4 M_{j,k} - \frac{1}{2} \sum_{k=3}^4 M_{j,k} - \omega_{j,3} \times X_j \beta_3 \end{pmatrix}.$$

This is computationally helpful because $M_{j,k}^*$ is equal to $\omega_k^{-1} (\sum_{c=k+1}^4 M_{j,c} - \frac{1}{2} \sum_{k=c}^4 M_{j,c})$. Calculating this expression requires taking the inverse of $\omega_{j,k}$, which may be close to zero. For numerical stability, we offset $\omega_{j,k}^{-1}$ by multiplying it by $\omega_{j,k}$ and set the minimum value of $\omega_{j,k}$ to 10^{-16} .

3.7. Gibbs Sampling Algorithm

The following Gibbs sampling steps are for respective posterior distributions for the aforementioned parameters— β , ω , γ , Σ , and u —with each step being conditioned on all other sampled parameters from previous steps. The initial values were drawn from either a standard normal distribution or a uniform distribution.

1. $\omega_{j,k}$ (see Section 3.3): Draw PG random values $\omega_{j,k}$ for each jurisdiction $j \in \{1, \dots, N\}$ and outcome category $k \in \{1, 2, 3\}$ as follows:

$$\omega_{j,k} \sim PG\left(\sum_{c=k}^4 M_{j,c}, \eta_{j,k}\right) \tag{17}$$

where $\eta_{j,k}$ is as defined in Section 3.3.

2. β_k (see Section 3.3): Draw coefficients β_k for $k \in \{1, 2, 3\}$ as follows:

$$\beta_k \sim N(V_{\beta_k}^{-1} m_{\beta_k}, V_{\beta_k}^{-1}) \tag{18}$$

where V_{β_k} and m_{β_k} are defined as in Section 3.3, and the prior covariance matrix B for β_k is defined as in Section 3.4.

3. γ_i (see Section 3.4): For each covariate $i \in \{1, \dots, p\}$, draw

$$\gamma_i \sim P(\gamma_i = 1 | Y, \beta, \gamma_{(i)}) = \frac{N(\beta | 0, D_\gamma R D_\gamma; \gamma_{(i)}, \gamma_i = 1) \times p_i}{\sum_{k \in \{0,1\}} N(\beta | 0, D_\gamma R D_\gamma; \gamma_{(i)}, \gamma_i = k) p_i^k (1 - p_i)^{1-k}}$$

where the matrices D_γ and R , as well as the proportions p_i for each i , are as defined in Section 3.4.

4. Σ (see Section 3.5): Draw

$$\Sigma \sim IW\left(\sum_{j=1}^N u_j u_j^T + V_0, N + s_0\right)$$

where u_j is defined as in Section 3.5, V_0 is set to be the 3×3 identity matrix $I_{3 \times 3}$, and s_0 is set to 3, due to the three disjoint outcome categories.

5. u_j (see Section 3.6): For each $j \in \{1, \dots, N\}$, draw

$$u_j = (u_{j,1}, u_{j,2}, u_{j,3}) \sim N\left(V_{j,u}^{-1} m_{j,u}, V_{j,u}^{-1}\right),$$

where $V_{j,u}$ and $m_{j,u}$ are as defined in Section 3.6.

4. 2021 VRA Data Analysis Results

4.1. Data Description

This paper uses the ACS 2015–2019 5-year data, which was used for the most recent 2021 VRA determination. The ACS releases 1-year and 5-year data products. The 5-year products aggregate and re-weight data collected over a 5-year period, allowing increased precision of population estimates. The 5-year data are particularly useful for estimating features of small geographic areas or small domains in which 1-year estimates are too imprecise for release under Census Bureau statistical quality guidelines.

Data for the four disjoint categorical outcomes described in Equation (1) were obtained from the ACS 2015–2019 data, as were the following covariates:

1. Logit-transformed fraction of voting-age persons who are citizens;
2. Logit-transformed fraction of citizens that are limited English-proficient;
3. Proportion of voting-age persons who are non-Hispanic White in each geography;
4. Proportion of voting-age persons with no college education in each geography;
5. Average number of voting-age people per housing unit in each geography;
6. Average age among voting-age persons in any AIAN LMG in each geography;
7. Proportion of voting-age persons in poverty in each geography;
8. Proportion of voting-age persons speaking a language other than English at home in each jurisdiction;
9. Proportion of foreign-born voting-age persons in each jurisdiction;
10. Average years in US (as of 2019) of voting-age foreign-born persons in each jurisdiction.

4.2. Model Comparison

This section highlights the improvement in model diagnostic results made by the Bayesian model described in this paper over the diagnostic results for the model used to produce the 2021 VRA determinations. We estimate the new model for 73 Asian, Hispanic, and Native American LMGs in the 7859 jurisdictions with respondents in the 2015–2019 ACS.

Note that direct population estimates from the ACS are reliable when they are made within domains with large sample sizes. Slud et al. [1] evaluate the performance of models by examining the discrepancy between their estimates and direct ACS estimates when aggregated over larger domains. The model is judged to be performing well when the direct and model-based estimates agree in these large sample domains.

We follow Slud et al. [1] in using three metrics for model evaluation: Δ , $\text{PctRel}\Delta$, and $\text{Stdiz}\Delta$. Δ is the difference between model predictions \tilde{N} and the ACS direct estimates \hat{N} . $\text{PctRel}\Delta$ is the percent relative difference $100 \cdot (\tilde{N} - \hat{N}) / \hat{N}$. $\text{Stdiz}\Delta$ is the standardized relative difference $(\tilde{N} - \hat{N}) / SE(\hat{N})$, where $SE(\hat{N})$ is the square root of the “successive difference replication” (SDR) estimate of the variance of the direct estimate \hat{N}

described in the Appendix of Slud et al. [1]. This is the standard error estimate generally used by the ACS, and for large domains, the standardized difference $(\hat{N} - N_{true})/SE(\hat{N})$ is distributed approximately as a standard normal. Because of this, $Stdiz\Delta$ roughly measures the departure from the null hypothesis that the \tilde{N} prediction and \hat{N} estimate are the same. However, since model-based predictions are conditioned on the direct estimate, \tilde{N} and \hat{N} will be positively correlated, and $Stdiz\Delta$ will be smaller in absolute value than a $\mathcal{N}(0, 1)$ deviate. We also follow Slud et al. [1] by evaluating these estimates in domains created by aggregating jurisdictions with voting-age sample sizes in the following pre-specified ranges: 1–4, 5–12, 13–25, 26–50, 51–200, and 201–10⁶ (the release of the statistics in this paper has been approved by the Census Bureau’s Disclosure Review Board (DRB) with the approval number CBDRB-FY25-0130). Priors were chosen such that model-diagnostic results were optimized.

4.2.1. Bangladeshi LMG Results

Table 3 compares the 2021 VRA model results for the Bangladeshi LMG published by Slud et al. [1] to the model results for the same LMG from our new Bayesian MLN model. In Table 3, the first three lines are results from the 2021 VRA model and the next three lines are results from the new model.

Table 3. Comparing new Bayes model and older results for Bangladeshi LMG.

	[1, 4]	[5, 12]	[13, 25]	[26, 50]	[51, 200]	[201, 1 × 10 ⁶]
Delta (Old)	611.9	9.7	−143.0	−47.5	−319.9	−64.3
PctRelΔ (Old)	47.2	0.4	−6.2	−1.9	−3.1	−0.4
StdizΔ (Old)	3.2	0.0	−0.5	−0.2	−0.6	−0.1
Delta (New)	223.6	−117.3	−126.1	−62.4	42.4	8.5
PctRelΔ (New)	17.3	−5.3	−5.5	−2.5	0.4	0.0
StdizΔ (New)	1.2	−0.4	−0.5	−0.2	0.1	0.0

For the jurisdictions with the number of voting-age persons larger than 4, i.e., [5, 12], [13, 25], . . . , [201, 1 × 10⁶], both the new Bayesian MLN model and the 2021 VRA model provided comparable $Stdiz\Delta$ values that were less than 1.96, the 97.5th percentile point of the standard normal distribution. The new Bayesian MLN model outperformed the 2021 VRA model for the jurisdictions with smaller number of voting-age persons ranging from 1 to 4; the old model resulted in a $StdzDelt$ of 3.2, while the new model produced 1.2. The enhanced prediction accuracy in small areas can likely be attributed to the impact of the SSVS variable selection procedure in the new model.

4.2.2. Sri Lankan LMG Results

Table 4 shows similar results for the Sri Lankan LMG, which was also used as an example in Slud et al. [1]. Like the Bangladeshi LMG, the new Bayes MLN model resulted in more accurate prediction for jurisdictions with a voting-age sample size ranging from 1 to 4.

Table 4. Comparing new Bayes model and older results for Sri Lankan LMG.

	[1, 4]	[5, 12]	[13, 25]	[26, 50]	[51, 200]	[201, 1 × 10 ⁶]
Delta (Old)	228.5	−143.3	−55.9	−8.3	−49.6	−11.4
PctRelΔ (Old)	54.3	−18.4	−6.7	−1.9	−3.6	−1.4
StdizΔ (Old)	2.5	−1.0	−0.3	−0.1	−0.2	−0.1
Delta (New)	87.2	−174.3	11.5	62.9	−2.5	−9.7
PctRelΔ (New)	20.7	−22.4	1.4	14.4	−0.2	−1.2
StdizΔ (New)	0.9	−1.2	0.1	0.6	0.0	−0.1

4.2.3. Overall Results

Table 5 compares the results for the new Bayesian MLN and old 2021 VRA models across all LMGs. The sum of absolute $\text{Stdiz}\Delta$ values across all 73 LMGs was produced for each sample size bin. The 73 LMGs include 21 Asian LMGs, 51 American Indian and Alaska Native (AIAN) LMGs, and a single Hispanic LMG that includes all racial groups. Overall, the sum of the absolute StdzDelt values across all of these LMGs is smaller for the new model than it was for the 2021 VRA model.

Table 5. Comparing new Bayes model and older results across all LMGs.

	[1, 4]	[5, 12]	[13, 25]	[26, 50]	[51, 200]	[201, 1×10^6]
Old	104.9	50.1	43	28.7	21.7	12.9
New	79.7	44.0	39	22.0	17.0	8.5

4.2.4. Improvement Rates

For each sample size bin, the first row of Table 6 shows the number of LMGs with absolute $\text{Stdiz}\Delta$ statistics larger than 1.96 in the 2021 VRA models, and the second shows the number of LMGs with absolute $\text{Stdiz}\Delta$ statistics larger than 1.96 in the new Bayes model. The third row shows the number of LMGs with absolute $\text{Stdiz}\Delta$ values greater than 1.96 in at least one of the two models, and the fourth row shows the number of LMGs where both models produce absolute $\text{Stdiz}\Delta$ values greater than 1.96. The fifth row of Table 6 shows the number of LMGs where $|\text{Stdiz}\Delta| > 1.96$ for both the old and new models, but the new model produces better absolute $\text{Stdiz}\Delta$ values than the old model. Finally, the sixth row shows the same data as the fifth row as a percentage of the number of LMGs where at least one model has $|\text{Stdiz}\Delta| > 1.96$.

Table 6. Comparing new Bayes (denoted by “New”) and older results across all LMGs.

	(0, 4]	(4, 12]	(12, 25]	(25, 50]	(50, 200]	(200, 1×10^6]	All
# LMGs with $ \text{Stdiz}\Delta > 1.96$ (Old)	15	5	2	1	1	0	24
# LMGs with $ \text{Stdiz}\Delta > 1.96$ (New)	7	6	2	1	0	0	16
# LMGs with $ \text{Stdiz}\Delta > 1.96$ (Old and New)	7	4	2	1	0	0	14
# LMGs with $ \text{Stdiz}\Delta > 1.96$ (Old or New)	15	7	2	1	1	0	26
# LMGs with $ \text{Stdiz}\Delta > 1.96$ (Improved)	14	4	2	1	1	0	22
% LMGs with $ \text{Stdiz}\Delta > 1.96$ (Improved)	93%	57%	100%	100%	100%	NA	85%

It should be noted that in most LMGs, the (0, 4] sample size bin contains more jurisdictions than the other bins, since many LMGs are sparsely sampled. In the first two rows of this table, we see that the number of LMGs with model estimates that are more than 1.96 SDR SDs away from the direct estimates in the (0, 4] sample size bin decreases by almost half in the new model compared to the old model. These numbers remain the same as the old model in the other sample size bins, except in the (4, 12] bin, for which there is one additional LMG that is “far” from the direct estimates. We also see that the new model performs better than the old model even when the model estimates are far from the direct estimates. In particular, the final row of Table 6 shows that in 85% of LMGs where the model estimates were more than 1.96 SDs away from the direct estimates, the new model performed better than the old model.

4.2.5. Residual Plots

Figure 1 compares the performance of the old and new models. Each point represents a jurisdiction across all 73 LMGs. The residual value here is the difference between “Direct” and “Model” estimates. LEPprop denotes the proportion of voting-age citizens who have limited English proficiency (or LEP / CIT), while ILLrat represents the illiteracy rate among citizens with limited English proficiency (or ILL / LEP). If all the points in this residual plot were on the 45-degree line, that would indicate that there was no difference between the old and new models. The closer the points are to a horizontal line in the center of the plot, the better the new model is performing for those jurisdictions. The fact that both plots show several points below the 45-degree lines and above the horizontal line at the center indicates that the old model has larger residuals than the new model for those jurisdictions. This pattern is consistent through all sample size bins.

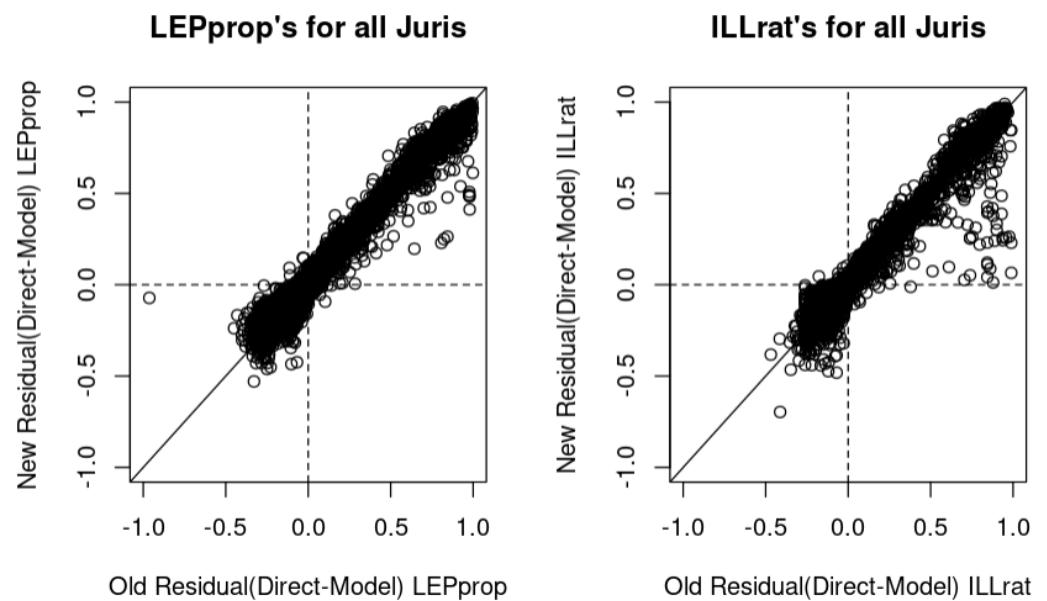


Figure 1. Residual plots.

5. Discussion

In this paper, a new statistical methodology was proposed for the Bayesian multinomial logit model for US Voting Rights Act Section 203 determinations. Our methodological contribution involved adopting PG augmentation to linearize the nonlinear optimization problem, resulting in closed-form solutions for posterior distributions. This enhanced computational stability in the Gibbs sampling steps, which enabled us to fit one model for all jurisdictions. The former 2021 VRA determinations used different models for different jurisdictions—that is, if one model did not fit, then other models were tried for the same jurisdiction data.

Our method was developed based on the works of George and McCulloch [8] and Polson et al. [6]. George and McCulloch [8] did not suggest formal convergence criteria or diagnostics. Instead, they relied on empirical observations of convergence. Polson et al. [6] implemented the PG sampling algorithm in the R package BayesLogit. Note that while convergence diagnostics such as trace plots can indicate non-convergence, they cannot definitively prove convergence, especially given the substantial number of parameters in our analysis. The ultimate evidence of convergence was the improvement in the new model’s diagnostic measures over the 2021 VRA model.

Using SSVS requires the specification of the prior parameters c_i and τ_i . Setting τ_i to 0.5 and c_i to 10 produced desirable diagnostic results for all LMGs, as discussed in

Section 4. In exchange for the additional effort required to identify suitable SSVS priors, the new PG-based Bayesian model demonstrates improved efficiency compared to the previous STAN-based Bayesian model developed for the 2021 VRA project. For example, the new model produced parameter estimates for the Bangladeshi LMG's 517 jurisdictions in approximately 8 min. In contrast, the 2021 Bayesian VRA model for this LMG could not be estimated at all using STAN. The use of other Bayesian variable selection procedures, such as Bayesian LASSO [9], is a topic of interest for future research.

The estimates from the new model can be viewed as Bayesian model averages, because the posterior parameters were sampled and averaged over various models that had different sets of selected covariates. The overall results for the 2021 VRA data showed that the new model, which used stochastic search variable selection (SSVS) [8] separately in each LMG, outperformed the previous model, which used the same covariates in every LMG. The diagnostic data presented for the Bangladeshi and Sri Lankan LMGs are examples of how the new model yielded better results than the 2021 VRA models, especially in jurisdictions with small sample sizes of between one and four voting-age persons.

The nonlinearity of the logistic link function of the MLN model and the unseen latent random effects variables were computationally challenging issues. Both the frequentist adaptive Gaussian quadrature procedure and Bayesian inference can handle random effects, as was demonstrated in previous 2021 VRA determinations [1]. The frequentist model was fit by using adaptive Gaussian quadrature to optimize the likelihood function, which required the random effects variable to be averaged out. This approach makes it technically difficult to adopt variable selection procedures, since the entire score function would have to be modified for such an additional procedure. In contrast, Bayesian methods treat the random effects as just another parameter, which can be drawn in Gibbs sampling. STAN was used to sample the random effects and their covariance matrix for the 2021 VRA Bayesian model, but it did not converge for smaller LMGs. Our new Bayesian model used PG augmentation to obtain closed-form solutions for posterior distributions, which made implementing Bayesian stochastic search variable selection computationally feasible in more LMGs. This led to noticeable improvements in our overall diagnostic results.

Author Contributions: Conceptualization, J.K. and A.C.H.; data curation, A.C.H.; formal analysis, J.K. and A.C.H.; methodology, J.K.; project administration, J.K.; software, J.K. and A.C.H.; supervision, J.K.; validation, J.K.; visualization, J.K.; writing—original draft preparation, J.K.; writing—review and editing, A.C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets presented in this article are not readily available because they are protected by Title 13, U.S.C., in the interests of respondent privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Derivation of the β Posterior [6]

Without loss of generality, we omit the outcome category indicator k from variable notations in this section. This means that β_k will be written as β , U_k will be written as U , Ω_k will be written as Ω , etc. Recall that $Z=(z_1, \dots, z_N)^T$, z_j denote $(y_j - n_j/2)/\omega_j$, and suppose ω_j is sampled from $PG(n_j, X_j\beta)$. Then, $P(Z|\beta, X)P(\beta|b, B)$ is proportional to

$$\begin{aligned}
 & \exp\left(-\frac{1}{2}\left[(Z - X\beta)^T \Omega (Z - X\beta) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z - X\beta)^T \left(\Omega^{\frac{1}{2}}\right)^T \Omega^{\frac{1}{2}}(Z - X\beta) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[\left(\Omega^{\frac{1}{2}}(Z - X\beta)\right)^T \left(\Omega^{\frac{1}{2}}(Z - X\beta)\right) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^* - X^*\beta)^T (Z^* - X^*\beta) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^* - X^*\hat{\beta} + X^*\hat{\beta} - X^*\beta)^T (Z^* - X^*\hat{\beta} + X^*\hat{\beta} - X^*\beta) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^* - X^*\hat{\beta})^T (Z^* - X^*\hat{\beta}) + (X^*\beta - X^*\hat{\beta})^T (X^*\beta - X^*\hat{\beta}) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^* - X^*\hat{\beta})^T (Z^* - X^*\hat{\beta}) + (\beta - \hat{\beta})^T X^{*T} X^* (\beta - \hat{\beta}) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z - X\hat{\beta})^T \Omega (Z - X\hat{\beta}) + (\beta - \hat{\beta})^T X^T \Omega X (\beta - \hat{\beta}) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[(\beta - \hat{\beta})^T X^T \Omega X (\beta - \hat{\beta}) + (\beta - b)^T B^{-1}(\beta - b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left(\beta - V_{\beta}^{-1} m_{\beta}\right)^T V_{\beta} \left(\beta - V_{\beta}^{-1} m_{\beta}\right)\right)
 \end{aligned}$$

where V_{β} is $X^T \Omega X + B^{-1}$ and m_{β} is $X_{p \times N}^T \Omega_{N \times N} X_{N \times p} \hat{\beta}_{p \times 1} + B^{-1} b$. Note that $X^T \Omega X \hat{\beta}$ is $X^T \Omega Z$ because $\hat{\beta}$ is such that $X^{*T} (Z^* - X^* \hat{\beta}) = 0$, where $A^* = \Omega^{\frac{1}{2}} A$ and because $X^{*T} (Z^* - X^* \hat{\beta})$ is the same as $X^T \Omega Z - X^T \Omega X \hat{\beta}$.

Thus, the posterior distribution of $\beta | Z, X$ is

$$N\left(V_{\beta}^{-1} m_{\beta}, V_{\beta}^{-1}\right),$$

where V_{β} is $X^T \Omega X + B^{-1}$ and m_{β} is $X^T \Omega z + B^{-1} b$, which is $X^T \kappa + B^{-1} b$, where $\kappa_i = (y_i - n_i / 2)$.

Appendix B. Derivation of the β Posterior for the VRA Model

Without loss of generality, we omit the outcome category indicator k from variable notations in this section. This means that β_k will be written as β , U_k will be written as U , Ω_k will be written as Ω , etc. Recall X is the matrix of covariates. Let U denote the $N \times 1$ matrix of random effects variables, $N(b, B)$ denote a standard normal prior for β , and the likelihood function $P(Z | \beta, U, X)$ denote $N(U + X\beta, \Omega^{-1})$. Then, $P(Z | \beta, U, X)P(\beta | b, B)$ is proportional to

$$\begin{aligned}
 & \exp\left(-\frac{1}{2}\left[(Z-U-X\beta)^T\Omega(Z-U-X\beta)+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(\tilde{Z}-X\beta)^T(\Omega^{\frac{1}{2}})^T\Omega^{\frac{1}{2}}(\tilde{Z}-X\beta)+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(\Omega^{\frac{1}{2}}(\tilde{Z}-X\beta))^T(\Omega^{\frac{1}{2}}(\tilde{Z}-X\beta))+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^*-X^*\beta)^T(Z^*-X^*\beta)+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^*-X^*\hat{\beta}+X^*\hat{\beta}-X^*\beta)^T(Z^*-X^*\hat{\beta}+X^*\hat{\beta}-X^*\beta)+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^*-X^*\hat{\beta})^T(Z^*-X^*\hat{\beta})+(X^*\beta-X^*\hat{\beta})^T(X^*\beta-X^*\hat{\beta})+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(Z^*-X^*\hat{\beta})^T(Z^*-X^*\hat{\beta})+(\beta-\hat{\beta})^TX^TX^*(\beta-\hat{\beta})+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[(\tilde{Z}-X\hat{\beta})^T\Omega(\tilde{Z}-X\hat{\beta})+(\beta-\hat{\beta})^TX^T\Omega X(\beta-\hat{\beta})+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[(\beta-\hat{\beta})^TX^T\Omega X(\beta-\hat{\beta})+(\beta-b)^TB^{-1}(\beta-b)\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left(\beta-V_{\beta}^{-1}m_{\beta}\right)^TV_{\beta}\left(\beta-V_{\beta}^{-1}m_{\beta}\right)\right)
 \end{aligned}$$

where V_{β} is $X^T\Omega X+B^{-1}$ and m_{β} is $X^T\Omega X\hat{\beta}+B^{-1}b$, A^* denotes $\Omega^{\frac{1}{2}}A$ for a variable A , and y and n denote y_1,\dots,y_N and n_1,\dots,n_N , respectively, as described in Section 2. Note that $X^T\Omega X\hat{\beta}$ is $X^T(y-n/2)-X^T\Omega U$ because $\hat{\beta}$ is such that $X^{*T}(Z^*-X^*\hat{\beta})=0$, and hence $X^T\Omega\tilde{Z}=X^T\Omega X\hat{\beta}$. Now, $X^T\Omega\tilde{Z}$ is $X^T\Omega(Z-U)$ and $X^T(y-n/2)-X^T\Omega U$. Thus, $X^T(y-n/2)-X^T\Omega U=X^T\Omega X\hat{\beta}$.

Appendix C. Derivation of $P(U|M, X, \beta, \omega)$

Let us assume a Gaussian prior distribution $N(0, \Sigma)$ for the random effects variables u_j ($=u_{j,1},u_{j,2},u_{j,3}$) for jurisdiction j . Recall that M_j denoted the counts of persons with language conditions. The conjugate Gaussian posterior for u_j is propotional to $P(M_j|u_j)\times P(u_j|\Sigma)$, as shown as follows.

$$\begin{aligned}
 & \frac{(\exp(\eta_{j,1}))^{\sum_{k=2}^4 M_{j,k}}}{(1+\exp(\eta_{j,1}))^{\sum_{k=1}^4 M_{j,k}}}\frac{(\exp(\eta_{j,2}))^{\sum_{k=3}^4 M_{j,k}}}{(1+\exp(\eta_{j,2}))^{\sum_{k=2}^4 M_{j,k}}}\frac{(\exp(\eta_{j,3}))^{\sum_{k=4}^4 M_{j,k}}}{(1+\exp(\eta_{j,3}))^{\sum_{k=3}^4 M_{j,k}}}\times \exp\left(-\frac{1}{2}u_j^T\Sigma^{-1}u_j\right) \\
 &= N\left(M_{j,1}^*|\eta_{j,1},\omega_{j,1}^{-1}\right)\times N\left(M_{j,2}^*|\eta_{j,2},\omega_{j,2}^{-1}\right)\times N\left(M_{j,3}^*|\eta_{j,3},\omega_{j,3}^{-1}\right)\times \exp\left(-\frac{1}{2}u_j^T\Sigma^{-1}u_j\right) \\
 &\propto \exp\left(-\frac{1}{2}\left[\omega_{j,1}\left(M_{j,1}^*-\eta_{j,1}\right)^2+\omega_{j,2}\left(M_{j,2}^*-\eta_{j,2}\right)^2+\omega_{j,3}\left(M_{j,3}^*-\eta_{j,3}\right)^2+u_j^T\Sigma^{-1}u_j\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[\omega_{j,1}\left(u_{j,1}-M_{j,1}^{*u}\right)^2+\omega_{j,2}\left(u_{j,2}-M_{j,2}^{*u}\right)^2+\omega_{j,3}\left(u_{j,3}-M_{j,3}^{*u}\right)^2+u_j^T\Sigma^{-1}u_j\right]\right) \\
 &= \exp\left(-\frac{1}{2}\left[\left(u_j-M_j^{*u}\right)_{[1\times 3]}^T\Omega_{j[3\times 3]}^*\left(u_j-M_j^{*u}\right)_{[3\times 1]}+\left(u_j-0\right)^T\Sigma^{-1}\left(u_j-0\right)\right]\right) \\
 &\propto N\left(V_{j,u}^{-1}m_{j,u},V_{j,u}^{-1}\right)
 \end{aligned}$$

where $V_{j,u}=\Omega_j^*+\Sigma^{-1}$, $m_{j,u}=\Omega_j^*M_j^{*u}$, $\Omega_j^*=diag(\omega_{j,1},\omega_{j,2},\omega_{j,3})$, $M_{j,k}^{*u}=M_{j,k}^*-X_j\beta_k$, $M_{j,1}^*=\omega_1^{-1}\left(\sum_{k=2}^4 M_{j,k}-\frac{1}{2}\sum_{k=1}^4 M_{j,k}\right)$, $M_{j,2}^*=\omega_2^{-1}\left(\sum_{k=3}^4 M_{j,k}-\frac{1}{2}\sum_{k=2}^4 M_{j,k}\right)$, and

$M_{j,3}^* = \omega_3^{-1} \left(\sum_{k=4}^4 M_{j,k} - \frac{1}{2} \sum_{k=3}^4 M_{j,k} \right)$, and $\eta_{j,k} \equiv X_j^T \beta_k + u_{j,k}$. The last two lines of the equations result by the algebra in Appendix E.

Appendix D. Derivation of $P(\Sigma|U)$

Assume a multivariate normal distribution $MVN(0, \Sigma)$ on U , which is

$$\begin{aligned} P(U|\Sigma) &= \prod_{j=1}^N (2\pi)^{-\frac{1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} u_j^T \Sigma^{-1} u_j\right) \\ &\propto |\Sigma|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^N u_j^T \Sigma^{-1} u_j\right) \\ &= |\Sigma|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_{j=1}^N u_j u_j^T \Sigma^{-1}\right)\right). \end{aligned}$$

An inverse Wishart $IW(V_0, s_0)$ for the prior distribution with known scale matrix V_0 and the degrees of freedom s_0 for a $p \times p$ covariance matrix Σ is

$$P(\Sigma) = \frac{|V_0|^{\frac{1}{2}s_0}}{2^{s_0 p/2} \Gamma_p(s_0/2)} |\Sigma|^{-\frac{1}{2}(s_0+p+1)} \exp\left(-\frac{1}{2} \text{tr}(V_0 \Sigma^{-1})\right)$$

The conditional posterior $P(\Sigma|U)$ is proportional to

$$\begin{aligned} &P(U|\Sigma) \times P(\Sigma) \\ &\propto |\Sigma|^{-\frac{1}{2}N} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_{j=1}^N u_j u_j^T \Sigma^{-1}\right)\right) \times |\Sigma|^{-\frac{1}{2}(s_0+p+1)} \exp\left(-\frac{1}{2} \text{tr}(V_0 \Sigma^{-1})\right) \\ &= |\Sigma|^{-\frac{1}{2}(N+s_0+p+1)} \exp\left(-\frac{1}{2} \text{tr}\left(\sum_{j=1}^N u_j u_j^T \Sigma^{-1} + V_0 \Sigma^{-1}\right)\right) \\ &= |\Sigma|^{-\frac{1}{2}(N+s_0+p+1)} \exp\left(-\frac{1}{2} \text{tr}\left(\left\{\sum_{j=1}^N u_j u_j^T + V_0\right\} \Sigma^{-1}\right)\right). \end{aligned}$$

Appendix E. Algebraic Basics: Sum of Quadratic Matrices

Appendix E.1. Sum of Two Quadratic Matrices

$$\begin{aligned} &(x - \mu_A)^T \Sigma_A^{-1} (x - \mu_A) + (x - \mu_B)^T \Sigma_B^{-1} (x - \mu_B) \\ &= x^T \Sigma_A^{-1} x - 2\mu_A^T \Sigma_A^{-1} x + \mu_A^T \Sigma_A^{-1} \mu_A + x^T \Sigma_B^{-1} x - 2\mu_B^T \Sigma_B^{-1} x + \mu_B^T \Sigma_B^{-1} \mu_B \\ &= x^T (\Sigma_A^{-1} + \Sigma_B^{-1}) x - 2(\Sigma_A^{-1} \mu_A + \Sigma_B^{-1} \mu_B)^T x + \mu_A^T \Sigma_A^{-1} \mu_A + \mu_B^T \Sigma_B^{-1} \mu_B \\ &= x^T V x - 2m^T x + R \\ &= (x - V^{-1} m)^T V (x - V^{-1} m) - m^T V^{-1} m + R \\ &\rightarrow N(V^{-1} m, V^{-1}) \end{aligned}$$

where $V = \Sigma_A^{-1} + \Sigma_B^{-1}$ and $m = \Sigma_A^{-1} \mu_A + \Sigma_B^{-1} \mu_B$. Thus, $\mu_A|x \sim N(\mu_A, \Sigma_A)$ and $x \sim N(\mu_B, \Sigma_B)$, so $x|\mu_A \sim N(V^{-1} m, V^{-1})$. $V^{-1} = (\Sigma_A^{-1} + \Sigma_B^{-1})^{-1}$ and $m = \Sigma_A^{-1} \mu_A + \Sigma_B^{-1} \mu_B$.

Appendix E.2. Sum of N Quadratic Matrices

Consider x to be a vector of the VRA’s three disjoint outcomes. Then,

$$\begin{aligned} & \sum_{j=1}^N (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) \\ &= (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) + (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) + \dots + (x - \mu_N)^T \Sigma_N^{-1} (x - \mu_N) \\ &= x^T \Sigma_1^{-1} x - 2\mu_1^T \Sigma_1^{-1} x + \mu_1^T \Sigma_1^{-1} \mu_1 + x^T \Sigma_2^{-1} x - 2\mu_2^T \Sigma_2^{-1} x + \mu_2^T \Sigma_2^{-1} \mu_2 + \dots + x^T \Sigma_N^{-1} x - 2\mu_N^T \Sigma_N^{-1} x + \mu_N^T \Sigma_N^{-1} \mu_N + R \\ &= x^T (\Sigma_1^{-1} + \Sigma_2^{-1} + \dots + \Sigma_N^{-1}) x - 2(\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2 + \dots + \Sigma_N^{-1} \mu_N)^T x + R \\ &= x^T V_N x - 2m_N^T x + R \\ &= (x - V_N^{-1} m_N)^T V_N (x - V_N^{-1} m_N) + R \\ &\rightarrow N(V^{-1} m, V^{-1}) \end{aligned}$$

$V_N = \sum_{j=1}^N \Sigma_j^{-1}$ and $m_N = \sum_{j=1}^N \Sigma_j^{-1} \mu_j$. Thus, $x | \mu_1, \dots, \mu_N \sim N(V_N^{-1} m_N, V_N^{-1})$.

References

1. Slud, E.V.; Franco, C.; Hall, A.; Kang, J. Statistical Methodology (2021) for Voting Rights Act, Section 203 Determinations. 2022. Available online: <https://www.census.gov/library/working-papers/2022/adrm/RRS2022-06.html> (accessed on 9 April 2024).
2. Joyce, P.M.; Malec, D.; Little, R.J.; Gilary, A.; Navarro, A.; Asiala, M.E. Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations. *J. Am. Stat. Assoc.* **2014**, *109*, 36–47. [CrossRef]
3. Slud, E.V.; Ashmead, R.; Joyce, P.; Wright, T. Statistical Methodology (2016) for Voting Rights Act, Section 203 Determinations. 2018. Available online: <https://www.census.gov/library/working-papers/2018/adrm/RRS2018-12.html> (accessed on 9 April 2024).
4. Pinheiro, J.C.; Bates, D.M. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *J. Comput. Graph. Stat.* **1995**, *4*, 12–35. [CrossRef]
5. Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*, 1–32. [CrossRef] [PubMed]
6. Polson, N.G.; Scott, J.G.; Windle, J. Bayesian Inference for Logistic Models Using Polya Gamma Latent Variables. *J. Am. Stat. Assoc.* **2013**, *108*, 1339–1349. [CrossRef]
7. Polson, N.G.; Scott, J.G.; Windle, J. Package: BayesLogit 2.1. 2019. Available online: <https://cran.r-project.org/web/packages/BayesLogit/index.html> (accessed on 9 April 2024).
8. George, E.I.; McCulloch, R.E. Variable Selection via Gibbs Sampling. *J. Am. Stat. Assoc.* **1993**, *88*, 881–889. [CrossRef]
9. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* **2008**, *103*, 681–686. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.