

Article

Closed Form Bayesian Inferences for Binary Logistic Regression with Applications to American Voter Turnout

Kevin Dayaratna ^{1,*}, Jesse Crosson ² and Chandler Hubbard ³¹ Center for Data Analysis, The Heritage Foundation, 214 Massachusetts Ave. NE, Washington, DC 20002, USA² Department of Political Science, Purdue University, 100 North University Street, West Lafayette, IN 47907, USA³ Department of Economics, College of Business Department 3985, University of Wyoming, Laramie, WY 82071, USA

* Correspondence: kevin.dayaratna@heritage.org

Abstract: Understanding the factors that influence voter turnout is a fundamentally important question in public policy and political science research. Bayesian logistic regression models are useful for incorporating individual level heterogeneity to answer these and many other questions. When these questions involve incorporating individual level heterogeneity for large data sets that include many demographic and ethnic subgroups, however, standard Markov Chain Monte Carlo (MCMC) sampling methods to estimate such models can be quite slow and impractical to perform in a reasonable amount of time. We present an innovative closed form Empirical Bayesian approach that is significantly faster than MCMC methods, thus enabling the estimation of voter turnout models that had previously been considered computationally infeasible. Our results shed light on factors impacting voter turnout data in the 2000, 2004, and 2008 presidential elections. We conclude with a discussion of these factors and the associated policy implications. We emphasize, however, that although our application is to the social sciences, our approach is fully generalizable to the myriads of other fields involving statistical models with binary dependent variables and high-dimensional parameter spaces as well.

Keywords: Bayesian inference; Empirical Bayes; U.S. elections; voter turnout



Citation: Dayaratna, K.; Crosson, J.; Hubbard, C. Closed Form Bayesian Inferences for Binary Logistic Regression with Applications to American Voter Turnout. *Stats* **2022**, *5*, 1174–1194. <https://doi.org/10.3390/stats5040070>

Academic Editor: Wenhao Gui

Received: 20 September 2022

Accepted: 9 November 2022

Published: 17 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Logistic Regression and Improvements to Bayesian Computation

From questions ranging from factors influencing voting decisions, to the determinants of child poverty, to the success/failure of terrorist attacks, to medical outcomes, to consumer choice, to baseball player hitting performance among others, modeling binary outcomes is a fundamental question in myriads of fields [1–21]. With the proliferation of statistical computing capabilities over the last several decades, incorporation of both unit-specific as well as individual-level heterogeneity has become increasingly commonplace in applied statistical research. Researchers have developed a variety of methodologies for incorporating heterogeneity including parametric and non-parametric Bayesian approaches, frequentist finite mixture approaches, as well as combinations of both [22–28].

Yet despite improvements in statistical computing power over the course of the past two decades, a practical challenge for such models—particularly their parametric Bayesian implementations—lies in their computational costliness. Both in terms of memory and computation time, estimating high-parameter models while incorporating unit-specific or individual-level heterogeneity constitutes a non-trivial challenge. In the social sciences in particular, these challenges are especially acute in certain applications with binary outcomes, as many of these applications such as modeling vote or roll call choice [3–6], voter turnout [7–9], and similar binary decisions [10,11] lend themselves to applications of logistic regression.

In this paper, we introduce an alternative method of estimation for logistic regression models incorporating individual-level heterogeneity that renders associated estimation considerably more computationally tractable. We then apply this method, based on polynomial expansions, to a political behavior—voter turnout—that is both well-suited for individual-level heterogeneity but nevertheless faces computational challenges due to the large size of the data and associated parameter space. More specifically, following one seminal paper on multilevel regression and poststratification (MRP) (Ghitza and Gelman 2013 [7]), we examine a dataset regarding voter turnout in three presidential elections (2000, 2004, and 2008) using Bayesian logistic regression to understand the factors influencing turnout. Our analysis of this dataset, currently incapable of being estimated via existing Markov Chain Monte Carlo (MCMC) methods, sheds light on the factors influencing voter turnout over the course of three U.S. Presidential elections. Although our application is to the social sciences, our method is fully generalizable to the myriads of other settings where logistic regression is applicable as well.

2. Individual-Level Heterogeneity and Computational Challenges

A wide variety of applications in the social sciences benefit from the incorporation of individual heterogeneity, at least in theory. Models of public opinion in particular benefit from the incorporation of unit-specific or individual-level heterogeneity: there is no *a priori* reason to believe a voter in New Jersey, for example, will respond to stimuli identically to a voter in Texas. Similarly, certain individuals in the Deep South tend to differ in important ways from their northern counterparts [29]. Nevertheless, given the scale of common public opinion datasets such as the American National Election Studies (ANES), Current Population Survey (CPS), and Cooperative Congressional Election Study (CCES), incorporating individual-level heterogeneity likely implies an intractably extensive parameter space. Ghitza and Gelman (2013), for example, explore voter turnout using the CPS with Bayesian logistic regression analysis, in an important early political science adaptation of MRP. However, even in their extensive application, the authors were restricted in limiting their heterogeneity to particular subgroups within their dataset [7]. In sum, although often theoretically justified, incorporating individual-level heterogeneity is often concomitant with the drawbacks of the computational complexity associated with numerical computation.

Although methodologists have developed an impressive number of estimation techniques to deal with large numbers of estimands, most are nevertheless computationally intensive as they pertain to individual-level heterogeneity. Frequently, researchers have attempted to incorporate heterogeneity with parametric Bayesian models. However, with limited data per individual in a data set, assuming a different parameterization for each individual can potentially render a model statistically unidentifiable, making estimation virtually impossible. In response, researchers will typically assume that these individual response coefficients are drawn from their own lower-dimensional probability distribution. They can then estimate these models from an empirical Bayesian perspective, or from a fully Bayesian perspective by imposing priors on the parameters of the heterogeneity distributions themselves [22,30]. Nevertheless, in addition to their computational intensiveness, commonly-used MCMC methods suffer from the drawback of sensitivity to starting values and can consequently result in a significant amount of simulation error. Likewise, numerical methods such as quadrature and simulated maximum likelihood methods can be difficult and time consuming to implement, especially for large data sets involving high-dimensional parameter spaces [31,32]. Given that the social sciences have seen a growing interest in ever-larger data sets, such computational drawbacks are of importance.

To address this problem, we utilize polynomial approximations to develop an alternative, “MCMC-free” approach initially developed in Dayaratna (2014) [1]. As we will show, this approach preserves the theoretical benefits of incorporating individual-heterogeneity while providing a framework for tractable, practical estimation. We then apply this approach to the same data presented by Ghitza and Gelman (2013) and demonstrate not only

the computational gains made possible by our method, but also the empirical insights enabled by heightened attention to individual-level heterogeneity [7].

Our work builds upon previous scholarship that has employed some similar approaches for other models. For example, Everson and Bradlow (2002) used polynomial expansions to approximate the posterior distributions of the beta-binomial random variables using a class of prior distributions previously considered non-conjugate [33]. Similarly, Bradlow et al. (2002) used polynomial expansions to improve on researchers' ability to make posterior inferences about the negative binomial distribution [34]. In subsequent research, McShane et al. (2008) used similar techniques to improve on Weibull count model estimation [35]. Finally, Miller et al. (2006) used polynomial expansions to examine the same problem examined here, namely for binary logistic regression [36]. Miller et al.'s approach, however, suffered from a serious limitation by requiring that the prior distribution be single-sided. Consequently, their result, although nice in principle, is limited in scope, as it is often unrealistic to be able to assume *a priori* that all coefficients have the same sign.

We address this limitation in this research by allowing the researcher to draw from one of the richest and most commonly used two-sided prior distributions—the normal distribution. Specifically, we offer a novel adaptation of Bayesian logistic regression with an application to voter behavior. In particular, we examine a dataset of voter behavior from three presidential elections (2000, 2004, and 2008) using our new approach to understand the various factors influencing voter turnout. This dataset consists of over 140,000 observations encompassing over 500 explanatory variables. As a result, incorporation of individual level heterogeneity in analyzing a dataset of this size results in a parameter space consisting of over 70 million parameters, thus far too large to estimate using standard MCMC methods. We instead develop an alternative empirical Bayesian approach utilizing series expansions that represents a viable alternative to existing MCMC methods. Although our approach is not fully Bayesian, it nevertheless serves as an approximation to the fully hierarchical Bayesian approach that is currently incapable of being estimated via existing MCMC methods in this setting [30]. On a tangible level, our approach enables us to draw inferences about the factors influencing voter turnout during these three elections. We conclude with a discussion of these results, policy implications, and potential avenues of future research.

In the following section, we develop our alternative estimation technique for binary logistic regression via polynomial expansions. Subsequently, we demonstrate the efficacy of our approach by estimating our model on a dataset on voter turnout, which, as we demonstrate, cannot be estimated by existing MCMC methods [7]. We emphasize, however that our approach is fully generalizable to many other settings involving binary dependent variables with high-dimensional parameter spaces as well.

3. Problem Formulation

As prefaced above, we utilize polynomial approximations as the basis of our alternative “MCMC-free” approach to estimation. We formulate our model as follows. Consider a data set obtained from $i \in \{1, \dots, I\}$ individuals (units) having $j \in \{1, \dots, J\}$ categories measured on $t \in \{1, \dots, N_i\}$ occasions (repeated measures). As is standard, we define

$$y_{ijt} = \begin{cases} 1 & \text{if outcome occurs for individual } i \text{ pertaining to category } j \text{ at time } t \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where $p_{ijt} = \text{Prob}(y_{ijt} = 1)$ is the probability of a particular outcome occurring (e.g., choosing to vote, choosing to form an alliance, living in poverty) for the i th individual pertaining to the j th category on the t th occasion. Additionally, let $p = 1, \dots, P$ represent a set of attributes pertaining to the covariates, with corresponding values $x_{ijt,p} \geq 0$ such that $X_{ijt}^T = (x_{ijt,1}, \dots, x_{ijt,P})$. To account for residual effects not manifested in the coefficient estimates for the explanatory variables, we can allow $x_{ijt,1} = 1$ defining category-level intercepts.

Multiplying over all individuals, categories, and occasions, we obtain the standard logit likelihood of the data, $Y = (y_{ijt})$:

$$P(Y|\beta) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}}, \quad (2)$$

where $\beta_i = (\beta_{i,1}, \dots, \beta_{i,p})$ is the coefficient vector for the i th individual with p th variable-specific coefficient, $\beta_{i,p}$ and $\beta = (\beta_1, \dots, \beta_I)$.

As argued above, our goal is to model individual- or unit-level heterogeneity, as there is no reason to believe that all individuals will behave in an identical manner. We can model this heterogeneity across individuals by allowing each $\beta_{i,p}$ to be drawn from probability distributions. In doing so, we build upon the Miller et al. (2006) [36] approach, and take a variety of steps, detailed below, to ameliorate the limitations of their approach.

3.1. Polynomial Expansions of the Binary Logit Model

We are interested in the following marginalized likelihood which we intend to maximize over our parameter space:

$$P(Y|\Omega) = \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta. \quad (3)$$

In the above equation, $P(Y|\beta)$ is our standard logit likelihood with a prior distribution $N(\beta|\Omega)$ and Ω represents the hyperparameters of our prior distribution. As mentioned above, we have non-negative explanatory variables $X_{ijt}^T = (x_{ijt,1}, \dots, x_{ijt,p})$ and binary dependent variables $Y = (y_{ijt})$. We intend to maximize the above marginalized likelihood over our prior distribution's parameter space. Specifically, our logit likelihood is:

$$P(Y|\beta) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}}, \quad (4)$$

It is the heterogeneity across $i = 1, \dots, I$ individuals in their $\beta_{i,p}$ coefficients that we are modeling by allowing these parameters to follow $N(\beta|\Omega)$. Due to the fact that the β_i appears in both the numerator and denominator of (4), performing the integration in (3) analytically for most choices of heterogeneity distributions without any numerical approximations is essentially impossible.

We can, however, take a series expansion approach to this problem and rewrite $P(Y|\beta)$ as follows:

$$\begin{aligned} P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} \\ &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{y_{ijt} X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{1}{1 + e^{X_{ijt}^T \beta_i}} \\ P(Y|\beta) &= P_1(Y|\beta) P_2(Y|\beta). \end{aligned} \quad (5)$$

We refer to the second factor above as $P_2(Y|\beta)$ although it does not depend on Y . If we assume $X_{ijt}^T \beta_i < 0$, we can expand $P_2(Y|\beta)$ via a geometric series expansion as follows [36]:

$$\begin{aligned} P_2(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{1}{1 + e^{X_{ijt}^T \beta_i}} \\ &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{k_{ijt} X_{ijt}^T \beta_i}. \end{aligned} \quad (6)$$

Putting together the pieces, we therefore have when $X_{ijt}^T \beta_i < 0$:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{y_{ijt} X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{1}{1 + e^{X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{y_{ijt} X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{k_{ijt} X_{ijt}^T \beta_i} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{y_{ijt} X_{ijt}^T \beta_i + k_{ijt} X_{ijt}^T \beta_i} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt} + k_{ijt}) X_{ijt}^T \beta_i} \\
P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt} + k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}}. \tag{7}
\end{aligned}$$

If, on the other hand, we assume $X_{ijt}^T \beta_i > 0$, we can also use a geometric series expansion:

$$\begin{aligned}
P_2(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{1}{1 + e^{X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{-X_{ijt}^T \beta_i}}{1 + e^{-X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{-X_{ijt}^T \beta_i} \frac{1}{1 + e^{-X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{-X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{-k_{ijt} X_{ijt}^T \beta_i} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{-X_{ijt}^T \beta_i - k_{ijt} X_{ijt}^T \beta_i} \\
P_2(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{-(1+k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}}. \tag{8}
\end{aligned}$$

Furthermore, again putting together the pieces:

$$\begin{aligned}
P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{y_{ijt} X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{1}{1 + e^{X_{ijt}^T \beta_i}} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} e^{y_{ijt} X_{ijt}^T \beta_i} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{-(1+k_{ijt}) X_{ijt}^T \beta_i} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{y_{ijt} X_{ijt}^T \beta_i - (1+k_{ijt}) X_{ijt}^T \beta_i} \\
&= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt} - 1 - k_{ijt}) X_{ijt}^T \beta_i} \\
P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt} - 1 - k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}}. \tag{9}
\end{aligned}$$

In the next section we utilize these series expansions to derive closed-form expressions from which we can make Bayesian inferences.

3.2. Closed Form Bayesian Inference via Polynomial Expansions as Described in Miller et al. (2006)

Ideally, one would like to allow each $\beta_{i,p}$ to follow two sided prior distribution, under such circumstances we would have a combination of both of the above situations, as well as when $X_{ijt}^T \beta_i = 0$:

$$\begin{aligned}
 P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} \\
 &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} (1) \right] \\
 &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} \left[I(X_{ijt}^T \beta_i > 0) + I(X_{ijt}^T \beta_i < 0) + I(X_{ijt}^T \beta_i = 0) \right] \right] \\
 P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} I(X_{ijt}^T \beta_i > 0) + \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} I(X_{ijt}^T \beta_i < 0) \right. \\
 &\quad \left. + \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} I(X_{ijt}^T \beta_i = 0) \right].
 \end{aligned}$$

This can be rewritten as:

$$\begin{aligned}
 P(Y|\beta) &= \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt}-1-k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} > 0\right) \right. \\
 &\quad + \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt}+k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} < 0\right) \\
 &\quad \left. + \frac{e^{y_{ijt} \sum_{p=1}^P x_{ijt,p} \beta_{i,p}}}{1 + e^{\sum_{p=1}^P x_{ijt,p} \beta_{i,p}}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} = 0\right) \right]. \quad (10)
 \end{aligned}$$

As a result, the marginalized likelihood is:

$$\begin{aligned}
 P(Y|\Omega) &= \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta \\
 &= \int_{\beta} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} N(\beta|\Omega) d\beta \\
 &= \int_{\beta} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \frac{e^{y_{ijt} X_{ijt}^T \beta_i}}{1 + e^{X_{ijt}^T \beta_i}} dP_{\beta_{i,p}} \\
 P(Y|\Omega) &= \int_{\beta} P(Y|\beta) dP_{\beta_{i,p}}
 \end{aligned}$$

where $P_{\beta_{i,p}}$ is the measure induced by $\beta_{i,p}$ on measurable space $(S_{i,p}, F_{i,p})$.

When Miller et al. (2006) looked at this problem, the authors attempted to integrate each $\beta_{i,p}$ individually for every potential value of i and p [36]. For a two-sided heterogeneity distribution, such as a normal heterogeneity distribution, this marginalization would involve:

$$\begin{aligned}
P(Y|\Omega) &= \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta \\
&= \int_{\beta} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt}-1-k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} > 0\right) \right. \\
&\quad + \sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt}+k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} < 0\right) \\
&\quad \left. + \frac{e^{y_{ijt} \sum_{p=1}^P x_{ijt,p} \beta_{i,p}}}{1 + e^{\sum_{p=1}^P x_{ijt,p} \beta_{i,p}}} I\left(\sum_{p=1}^P x_{ijt,p} \beta_{i,p} = 0\right) \right] \cdot \prod_{p=1}^P \frac{1}{\sqrt{2\pi}\sigma_p} \cdot e^{-\frac{(\beta_{i,p}-\mu_p)^2}{2\sigma_p^2}} d\mathbf{f}_{i,p}. \quad (11)
\end{aligned}$$

As the range of $\beta_{i,p}$ is the entire real line, the limits of the integration space differ for the first and second integrals depending on whether $\sum_{p=1}^P x_{ijt,p} \beta_{i,p} < 0$ or $\sum_{p=1}^P x_{ijt,p} \beta_{i,p} > 0$. Miller et al. (2006) noted that integrating over both spaces would result in “numerous, complicated subdivisions of the integration space.” These subdivisions, they argued, rendered the integration “untenable” and precluded the derivation of “tractable closed-form expansions.” As a result, the authors restricted their model to adhere to only one of the above cases by requiring that X_{ijt}^T be non-negative and stipulated that the density $N(\beta|\Omega)$ being integrated over a probability distribution with positive support. Making the assumption that $N(\beta|\Omega)$ was composed of independent gamma distributions $g(\beta_{i,p}|b_p, n_p)$ with parameters b_p and n_p (i.e., $N(\beta|\Omega) = \prod_{p=1}^P g(\beta_{i,p}|b_p, n_p)$), they derived the marginalized likelihood as follows:

$$\begin{aligned}
P(Y|\Omega) &= \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta \\
&= \int_{\beta} \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^{N_i} \left[\sum_{k_{ijt}=0}^{\infty} (-1)^{k_{ijt}} e^{(y_{ijt}+k_{ijt}) \sum_{p=1}^P x_{ijt,p} \beta_{i,p}} \prod_{p=1}^P g(\beta_{i,p}|b_p, n_p) d\mathbf{f}_{i,p} \right] \\
&= \prod_{i=1}^I \sum_{k_{i11}=0}^{\infty} \dots \sum_{k_{iJN_i}=0}^{\infty} (-1)^{\sum_{j=1}^J \sum_{t=1}^{N_i} k_{ijt}} \prod_{p=1}^P \int_{\beta_{i,p}=0}^{\infty} e^{-\sum_{j=1}^J \sum_{t=1}^{N_i} (y_{ijt}+k_{ijt}) x_{ijt,p} \beta_{i,p}} \\
&\quad \cdot \frac{1}{b_p \Gamma(n_p)} \left(\frac{\beta_{i,p}}{b_p} \right)^{n_p-1} e^{-\beta_{i,p}/b_p} d\mathbf{f}_{i,p} \\
&= \prod_{i=1}^I \sum_{k_{i11}=0}^{\infty} \dots \sum_{k_{iJN_i}=0}^{\infty} (-1)^{\sum_{j=1}^J \sum_{t=1}^{N_i} k_{ijt}} \prod_{p=1}^P \left(\frac{1}{1 + b_p \sum_{j=1}^J \sum_{t=1}^{N_i} (y_{ijt} + k_{ijt}) x_{ijt,p}} \right)^{n_p}
\end{aligned}$$

Having made assumptions that the explanatory variables $X_{ijt,p}$ were restricted to the set of non-negative integers, Miller et al. (2006) borrowed tools from analytic number theory to rewrite the above equation in terms of solutions to a system of Diophantine equations, which made estimating the model significantly more feasible from a computational perspective [37]. The interested reader is referred to Miller et al. (2006) for a complete discussion of this methodology [36].

3.3. Bayesian Inference via Polynomial Expansions Using a Two-Sided Heterogeneity Distribution

Although the Miller et al. (2006) result is elegant mathematically, it is not particularly useful to implement in practice as in most applications it is generally unrealistic to a priori assume that the regression coefficients all have the same sign. However, for the case when $J = 1$ and $N_i = 1$, a simple transformation of variables leads to very clean and tractable integration, allowing us to integrate within distinct regions along the real line. Restricting J and N_i in this manner is quite reasonable for many applied statistical problems including

cross-sectional data analysis with a single category (such as the voter turnout application looked at later in this study), longitudinal analysis of a single individual (such as the baseball player hitting streak analysis conducted in Albright (1993) [15], or analysis where the heterogeneity can be assumed across all observations of the data set (such as the terrorist attack data analysis conducted in Kyung et al. (2011) or the data set used in the analysis of medical outcomes in Wisner (1990) [2,14,15,36].

Specifically, if we make the assumption that $p_i = \text{Prob}(y_i = 1)$ is the probability of a particular outcome occurring (e.g., choosing to vote, living in poverty, occurrence of a war, whether a citizen votes, decision to cosponsor legislation, etc.) for the i th individual and again let $p = 1, \dots, P$ represent a set of attributes describing the covariates, with corresponding values $x_{i,p}$ such that $X_i^T = (x_{i,1}, \dots, x_{i,P})$ and take the product across all individuals i , the likelihood function is:

$$P(Y|\beta) = \prod_{i=1}^I \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}}, \quad (12)$$

where $\beta_i = (\beta_{i,1}, \dots, \beta_{i,P})$ and β are defined as before. Upon making these assumptions, we can recall the marginalization presented in (11):

$$\begin{aligned} P(Y|\Omega) &= \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta \\ &= \int_{\beta} \prod_{i=1}^I \left[\sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i-1-k_i) \sum_{p=1}^P x_{i,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{i,p} \beta_{i,p} > 0\right) \right. \\ &\quad \left. + \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i+k_i) \sum_{p=1}^P x_{i,p} \beta_{i,p}} I\left(\sum_{p=1}^P x_{i,p} \beta_{i,p} < 0\right) + \frac{e^{y_i \sum_{p=1}^P x_{i,p} \beta_{i,p}}}{1 + e^{\sum_{p=1}^P x_{i,p} \beta_{i,p}}} I\left(\sum_{p=1}^P x_{i,p} \beta_{i,p} = 0\right) \right] \\ &\quad \cdot \prod_{p=1}^P \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{(\beta_{i,p}-\mu_p)^2}{2\sigma_p^2}} d\mathbf{f}_{i,p}. \end{aligned} \quad (13)$$

In particular, since we are assuming that the $\beta_{i,p}$ follow independent normal distributions for $p = 1, \dots, P$ (i.e., $\beta_{i,p} \sim N(\mu_p, \sigma_p^2)$) it follows that $z_i = \sum_{p=1}^P x_{i,p} \beta_{i,p} \sim N(\sum_{p=1}^P x_{i,p} \mu_p, \sum_{p=1}^P x_{i,p}^2 \sigma_p^2)$. Therefore, if we define P_{z_i} as the measure induced by z_i on measurable space (T_i, G_i) having density with respect to Lebesgue measure:

$$f(z_i) = \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} e^{-\frac{(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}},$$

then:

$$\begin{aligned} P(Y|\Omega) &= \int_{\beta} P(Y|\beta) N(\beta|\Omega) d\beta \\ &= \int_{\beta} \prod_{i=1}^I \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}} N(\beta|\Omega) d\beta \\ &= \int_{\beta} \prod_{i=1}^I \frac{e^{y_i X_i^T \beta_i}}{1 + e^{X_i^T \beta_i}} dP_{\beta_{i,p}} \\ P(Y|\Omega) &= \int_{\beta} P(Y|\beta) dP_{\beta_{i,p}}. \end{aligned}$$

Applying our transformation we can see that [38]:

$$\begin{aligned}
 P(Y|\Omega) &= \int_{\beta} P(Y|\beta) dP_{\beta_{i,p}} \\
 &= \int_{T_1} \dots \int_{T_1} \prod_{i=1}^I P(y_i|z_i) dP_{z_i} \\
 &= \prod_{i=1}^I \int_{T_i} P(y_i|z_i) dP_{z_i} \\
 &= \prod_{i=1}^I \int_{z_i} P(y_i|z_i) \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i \\
 P(Y|\Omega) &= \prod_{i=1}^I H_i,
 \end{aligned} \tag{14}$$

where:

$$\begin{aligned}
 H_i &= \int_{-\infty}^{\infty} \left[\sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i-1-k_i)z_i} I(z_i > 0) \right. \\
 &\quad \left. + \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{(y_i+k_i)z_i} I(z_i < 0) + \frac{e^{y_i z_i}}{1 + e^{z_i}} I(z_i = 0) \right] \\
 &\quad \cdot \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i.
 \end{aligned}$$

We can decompose H_i into a sum of three integrals, $H_{i,1}$, $H_{i,2}$, and $H_{i,3}$ where $H_i = H_{i,1} + H_{i,2} + H_{i,3}$.

Before we proceed, however, we present a simple integration lemma for integrating an exponential against a normal distribution with mean μ and variance σ^2 , which involves completing the square of the function.

Lemma 1 (Integrating an Exponential Against a Normal Distribution).

$$\int_c^{\infty} e^{kx} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(\frac{k\sigma^2 - c + \mu}{\sigma}\right) \tag{15}$$

$$\int_{-\infty}^c e^{kx} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = e^{k\mu + \frac{k^2\sigma^2}{2}} \Phi\left(-\frac{k\sigma^2 - c + \mu}{\sigma}\right) \tag{16}$$

where $\Phi(x)$ is the normal cumulative distribution function.

Proof.

$$\begin{aligned}
 \int_c^\infty e^{kx} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= \int_c^\infty \frac{1}{\sqrt{2\pi\sigma}} e^{kx - \frac{(x-\mu)^2}{2\sigma^2}} dx \\
 &= \int_c^\infty \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2 + kx} dx \\
 &= \int_c^\infty \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2\sigma^2} [x - (\sigma^2 k + \mu)]^2 - (\sigma^2 k + \mu)^2 + \mu^2} dx \\
 &= \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2\sigma^2} [-(\sigma^2 k + \mu)^2 + \mu^2]} \int_c^\infty e^{\frac{-1}{2\sigma^2} [x - (\sigma^2 k + \mu)]^2} dx \\
 &= e^{\frac{\sigma^2 k^2}{2} + k\mu} \left[1 - \Phi \left(\frac{c - (\sigma^2 k + \mu)}{\sigma} \right) \right] \\
 \int_c^\infty e^{kx} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx &= e^{k\mu + \frac{k^2 \sigma^2}{2}} \Phi \left(\frac{k\sigma^2 - c + \mu}{\sigma} \right). \quad (17)
 \end{aligned}$$

The computation of the second integral is quite similar. As a result of (Lemma 1),

$$\begin{aligned}
 H_{i,1} &= \int_0^\infty \sum_{k_i=0}^\infty (-1)^{k_i} e^{(y_i-1-k_i)z_i} \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i \\
 &= \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \sum_{k_i=0}^\infty (-1)^{k_i} \int_0^\infty e^{(y_i-1-k_i)z_i} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i \\
 H_{i,1} &= \sum_{k_i=0}^\infty (-1)^{k_i} e^{\frac{(y_i-1-k_i)^2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}{2} + (y_i-1-k_i) \sum_{p=1}^P x_{i,p} \mu_p} \\
 &\quad \cdot \Phi \left(\frac{(y_i-1-k_i) \sum_{p=1}^P x_{i,p}^2 \sigma_p^2 + \sum_{p=1}^P x_{i,p} \mu_p}{\sqrt{\sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \right). \quad (18)
 \end{aligned}$$

$$\begin{aligned}
 H_{i,2} &= \int_{-\infty}^0 \sum_{k_i=0}^\infty (-1)^{k_i} e^{(y_i+k_i)z_i} \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i \\
 &= \frac{1}{\sqrt{2\pi \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \sum_{k_i=0}^\infty (-1)^{k_i} \int_{-\infty}^0 e^{(y_i+k_i)z_i} e^{\frac{-(z_i - \sum_{p=1}^P x_{i,p} \mu_p)^2}{2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} dz_i \\
 H_{i,2} &= \sum_{k_i=0}^\infty (-1)^{k_i} e^{\frac{(y_i+k_i)^2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}{2} + (y_i+k_i) \sum_{p=1}^P x_{i,p} \mu_p} \\
 &\quad \cdot \Phi \left(-\frac{(y_i+k_i) \sum_{p=1}^P x_{i,p}^2 \sigma_p^2 + \sum_{p=1}^P x_{i,p} \mu_p}{\sqrt{\sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \right). \quad (19)
 \end{aligned}$$

$H_{i,3} = 0$ as it is an integral against a density on a set of Lebesgue measure zero. As a result, as $P(Y|\Omega) = \prod_{i=1}^I H_i$, we estimate our model via maximum likelihood estimation by maximizing $\log P(Y|\Omega)$ which is equivalent to:

$$\begin{aligned}
\log P(Y|\Omega) &= \log \prod_{i=1}^I H_i \\
&= \sum_{i=1}^I \log(H_i) \\
&= \sum_{i=1}^I \log(H_{i,1} + H_{i,2}),
\end{aligned} \tag{20}$$

where $H_{i,1}$ and $H_{i,2}$ are defined as above. We state this result as a theorem as it is the main result of this section. \square

Theorem 1 (Marginalized Logit Likelihood Assuming Independent Normal Prior Distributions). *The log marginalized likelihood of (12) assuming independent normal heterogeneity distributions, based on a convergent series approximation to (3), is provided by:*

$$\log P(Y|\Omega) = \sum_{i=1}^I \log(H_{i,1} + H_{i,2}), \tag{21}$$

where:

$$\begin{aligned}
H_{i,1} &= \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{(y_i-1-k_i)^2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}{2} + (y_i-1-k_i) \sum_{p=1}^P x_{i,p} \mu_p} \\
&\quad \cdot \Phi \left(\frac{(y_i-1-k_i) \sum_{p=1}^P x_{i,p}^2 \sigma_p^2 + \sum_{p=1}^P x_{i,p} \mu_p}{\sqrt{\sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \right)
\end{aligned} \tag{22}$$

$$\begin{aligned}
H_{i,2} &= \sum_{k_i=0}^{\infty} (-1)^{k_i} e^{\frac{((y_i+k_i))^2 \sum_{p=1}^P x_{i,p}^2 \sigma_p^2}{2} + (y_i+k_i) \sum_{p=1}^P x_{i,p} \mu_p} \\
&\quad \cdot \Phi \left(-\frac{(y_i+k_i) \sum_{p=1}^P x_{i,p}^2 \sigma_p^2 + \sum_{p=1}^P x_{i,p} \mu_p}{\sqrt{\sum_{p=1}^P x_{i,p}^2 \sigma_p^2}} \right)
\end{aligned} \tag{23}$$

Theorem 1 provides the marginalized likelihood, and we can estimate our parameters μ_p and σ_p^2 for $p = 1, \dots, P$ by maximizing the above equation and compute associated p-values to determine the statistical significance of the resulting estimates. This marginalization reduces the parameter space from one of IP dimensions to $2P$ dimensions, making model estimation on large data sets considerably more practical. Of course, Equations (22) and (23) that form the two fundamental components of Equation (21) are convergent series and in practice need to be truncated. In Section 4.2, we discuss a heuristic for determining this truncation level.

As is the case with empirical Bayesian modeling, the approach presented here is inherently more frequentist than Bayesian in nature as the model is estimated via the method of maximum marginalized likelihood (MML) and p-values for associated coefficient estimates are reported. Nevertheless, this empirical Bayesian approach is an approximation for fully hierarchical Bayesian inferences and enables estimation for large data sets—including the one presented in this study - that would otherwise not be feasible to estimate in real time. Furthermore, analysis using both models would ultimately examine the same information setting - namely estimates for lower dimensional parameters for the purpose of making inferences regarding various sub-groups.

4. Application: Individual-Level Heterogeneity and Voter Turnout

Having established the computational advantages of our approach, we now highlight how these gains can aid important substantive applications. As we have underscored already, the dramatic growth in the size and scope of datasets in political science has placed increased importance on the computational efficiency of various methods. This importance is compounded when estimated models with large numbers of parameters—particularly those that seek to incorporate individual-level heterogeneity. Yet in spite of the theoretical reasons to account for individual-heterogeneity, the computational expense associated with doing so may deter researchers. In this section, we show that individual-level heterogeneity is not only substantively consequential in practice, but that our estimation methodology ameliorates the computational challenges associated with such heterogeneity.

4.1. Data

To demonstrate the substantive and methodological advantages of our approach, we re-examine voter turnout data from the 2000, 2004, and 2008 presidential elections, previously analyzed by [7]. Ghitza and Gelman (2013) use the data in a particularly high-parameter application, MRP, leading them to explore a variety of combinations of variables and groupings in an effort to generate models with strong capability for predicting future voter-turnout numbers in various states.

The data for these analyses originally come from the CPS Post-Election Voting and Registration Supplement (summarized in Table 1), and our dependent variable is intent to vote in the specified presidential election. Explanatory variables include the election year, state, ethnicity, income, age, sex, education, marital status, and the presence of children in the survey participants' households. Analysis consisted of 147,689 respondents having complete data amongst the variables examined. Variables were benchmarked with respect to the first variable in the Categories column. As Table 2 shows, incorporation of interactions among variables in the dataset, similar to Ghitza and Gelman (2013), yielded a total 546 explanatory variables in the analysis [7]. Since each normally distributed prior distribution contains two parameters to estimate, our resulting model involved estimating 1092 variables.

Table 1. CPS Post-Election Voting and Registration Supplement.

Intent to Vote	Percentage	Ethnicity	Percentage	Income	Percentage
Yes	71.60%	White	80.35%	\$0–20k	11.60%
No	28.40%	Black	8.10%	\$20–40k	23.10%
		Latinx	6.79%	\$40–75k	29.40%
		Other	4.76%	\$75–150k	22.10%
				\$150k+	13.80%
Age	Percentage	Sex	Percentage	Education	Percentage
18–29	10.80%	Male	46.20%	<HS	10.20%
30–44	33.30%	Female	53.80%	HS	32.00%
45–64	40.40%			Some Coll	28.50%
65+	15.50%			Coll	19.00%
				Post-Grad	10.3%
Marital	Percentage	Kids	Percentage		
Married	86.40%	Kids	45.80%		
Single	13.60%	No Kids	54.20%		

Table 2. Variables in Regression Model.

Variable	Categories	Number of Variables	After Benchmarking First Variable
Main Effects Terms			
Intercept Term		1	1
Election Year	2000, 2004, 2008	3	2
State	51 States (including District of Columbia)	51	50
Ethnicity	White, Black, Latinx, Other	4	3
Income	\$0–20k, \$20–40k, \$40–75k, \$75–150k, \$150k+	5	4
Age	18–29, 30–44, 45–64, 65+	4	3
Sex	Male, Female	2	1
Education	<HS, HS, Some Coll, Coll, Post-Grad	5	4
Marital Status	Married, Single	2	1
Kids	Kids, No Kids	2	1
Interaction Terms			
Variable	Number of Interactions		
Ethnicity \times Income	$4 \times 5 = 20$	20	19
State \times Income	$50 \times 5 = 250$	255	254
Ethnicity \times State	$4 \times 50 = 200$	204	203
Total number of terms		558	546

4.2. Truncation Levels

Equations (22) and (23) are convergent series that can be made arbitrarily close in practice upon truncating the series after an a priori specified number of terms. As higher-order approximations provide greater accuracy at the expense of computational cost/time, however, there is a significant balancing act between these two factors in which the researcher must engage. A key question therefore is the determination of a reasonable truncation level for the series approximations. We developed a heuristic to do so by computing Equation (21) under randomly chosen parameterizations for a variety of truncation levels as presented in Figure 1. The heuristic bears a significant amount of similarity to the “elbow-method” used to determine the optimal number of clusters in *k*-means cluster analysis [39].

As Figure 1 illustrates, however, the marginalized likelihood in Equation (21) for estimation of the dataset discussed in Section 4.1 steadily increases after 25 terms but stabilizes after 200 terms. As a result, 200 terms is thus a reasonable and defensible truncation level in terms of this application.

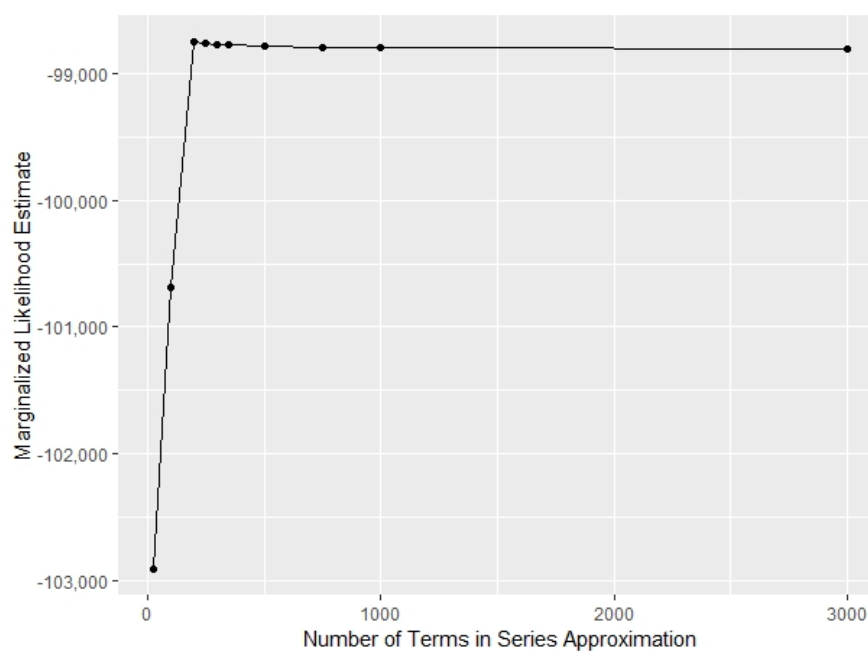


Figure 1. Marginalized Likelihood Estimate Based on Various Truncation Levels of Series Approximation for Marginalized Likelihood in Equation (21).

4.3. Model Estimation and Results

4.3.1. Existing MCMC Methods

We initially attempted to answer the question about the factors influencing voter turnout by estimating a hierarchical binomial logistic regression model in R via MCMC methods using the R package *bayesm* [40] to obtain fully Bayesian inferences. Unfortunately, due to the large number of explanatory variables and high dimensionality of the associated parameter space as a result of imposing prior distributions on every survey respondent (over 70 million parameters), we could not obtain even ten MCMC iterations over the course in 48 hours of run time on Williams College’s High Performance Computing Cluster (hereafter referred to as WHPCC), thus demonstrating the inability of existing MCMC methods to be able to answer the question of interest.

4.3.2. Closed Form Bayesian Inferences Using Convergent Polynomial Approximations

As existing MCMC methods are incapable of answering our question of interest, we drew upon our new MCMC-free approach to estimate the marginalized likelihood from Equations (21)–(23). In particular, we utilized 64 cores and 128 GB of RAM on WHPCC [41] to estimate our model via bootstrapping sub-samples of size 100,000 across 30 iterations using the Benham et al. (2015) cross-entropy parallelized optimization routine [42]. We truncated the series expansions depicted in Equations (21)–(23) and using the heuristic presented in Section 4.2.

Regression coefficients largest in magnitude are presented in Figures 2 and 3. Tables S10 and S11 from our supplemental material present statistically significant factors μ_p and σ_p from Equations (21)–(23) influencing voter turnout based on the 2.5th, 50th, and 97.5th percentiles from this bootstrapping process. Table S10 contains the 2.5th, 50th, and 97.5th percentiles for statistically significant parameter estimates μ_p from the parameter sample garnered via bootstrapping, estimated by standard smoothing techniques. Specifically, if a particular percentile is based on smoothing the k^{th} and $k + 1^{st}$ elements, then the closest element to the estimated percentile is used as the basis for determining the corresponding variance estimate. Table S11 contains the corresponding elements from this bootstrapped sample with respect to Table S10. Figures 2 and 3 condense this information, summarizing the main effects and interactions with the largest magnitudes from the analysis.

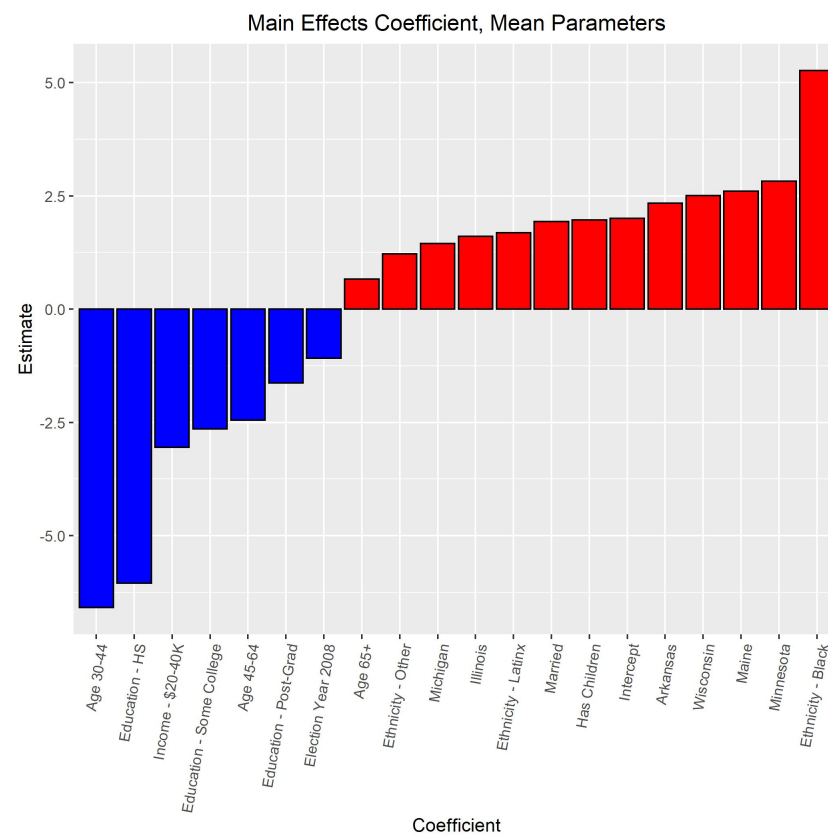


Figure 2. Selected Main Effects

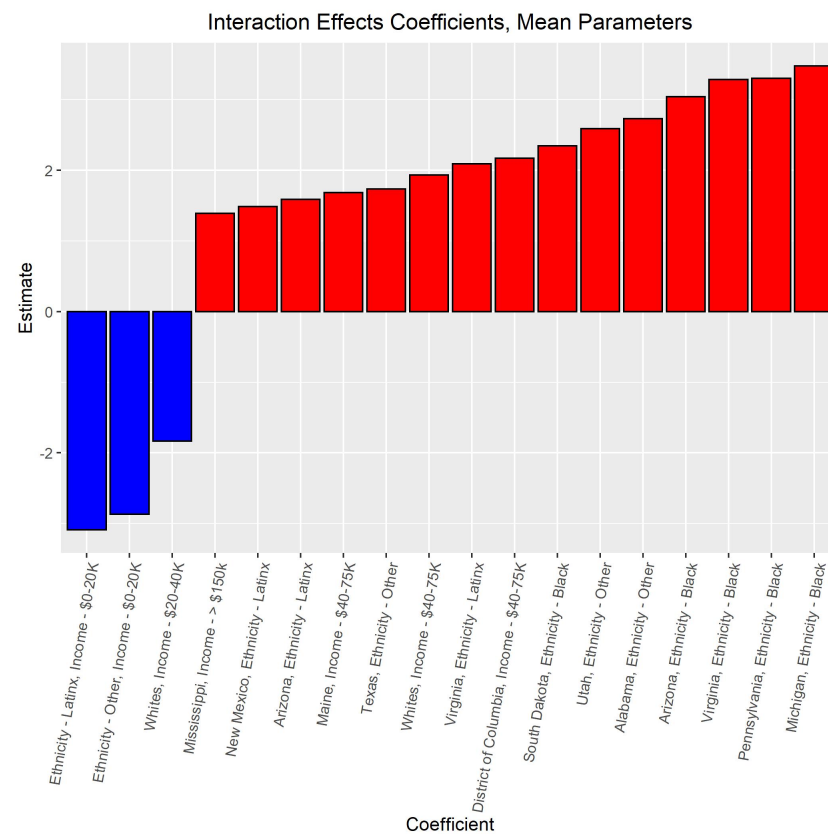


Figure 3. Selected Interactions

In terms of baseline associations, our results offer some interesting insights. For example, Election Year 2008 was statistically significant, having a negative coefficient estimate of -1.082 . This estimate suggests, according to the CPS data, that there was overall a lower propensity to turnout than the benchmark year 2000 as well as 2004—in spite of the fact that 2008 is regarded as a high-turnout election. In fact, 2008 ultimately had higher turnout than 2000 and 2004. However, as Ghitza and Gelman (2013) found, increased turnout in the 2008 election appears to have been driven primarily by particular racial minority groups. Black, Latinx, and other non-white voters saw a higher probability of voter turnout with coefficients of 5.263, 1.684, and 1.218, respectively, with respect to white voters. However, with respect to income, the only group that reached statistical significance was the \$20–40k income group, which had a coefficient of -3.053 . Thus, while some traditionally disadvantaged populations turned out in impressive numbers, others actually had a *lower* probability of voting than others than other income groups including the \$0–20k benchmark.

Like Ghitza and Gelman (2013), we also recover considerable heterogeneity by state—much of which not explained by traditional “battleground state” explanations. Arkansas, Illinois, Maine, Michigan, Minnesota, and Wisconsin had the highest propensity of voter turnout with coefficients ranging from 1.448 (Michigan) to 2.821 (Minnesota). Other states were not statistically significant (within the 95% level). Additionally, age groups exhibit considerable variation. Age groups from 30–44 and 45–64 have negative coefficients of -6.586 and -2.447 , respectively, while the group 65+ had a positive coefficient of 0.665 with respect to the under 30 age group.

Similar to Ghitza and Gelman (2013)—though in a fraction of the computation time relative to MCMC—our analysis underscores the importance of including a wide variety of interactions. Indeed, as Figure 3 indicates, variables such as race behave quite differently, conditional on other factors. For example, white voters in the \$20–40k income bracket had a coefficient of -1.834 , thus exhibiting a lower propensity to turnout than other groups, while those in the \$40–75 k bracket had a positive coefficient of 1.932 and thus had a greater propensity to vote. Similar to white voters, Latinx and other non-white voters earning \$20–40k also had negative coefficients of -3.090 and -2.867 , again suggesting that these groups had a lower propensity to turnout. With regard to voter location, D.C.’s and Maine’s \$40–75 k income bracket, as well as the bracket exceeding \$150k income bracket of Mississippi, exhibited a high probability of turnout—while no other state-income interactions reached significance. These estimates are consistent with what others have expected of income to voter turnout ratios [43]. Finally, while black voters exhibited higher turnout overall compared to other groups, some particular subsets turnout at especially high rates. Black voters in Arizona, Michigan, Pennsylvania, South Dakota, and Virginia, for example, exhibited considerable turnout rates. Likewise, Latinx voters in Arizona, New Mexico, and Virginia, alongside other non-white voters from Alabama, Texas, and Utah, saw a similarly turnout probability.

Beyond replicating findings from previous work, the computational advantages of our approach enabled us to estimate unit-level heterogeneity, some of which is captured in the variance parameters displayed in Figures 4 and 5. Indeed, whereas particular groups and group intersections exhibited notably higher and lower levers of turnout, our results also indicate that many of these groups and intersections displayed considerable *variance* in their turnout rates. In other words, while some groups have high *mean* turnout, for example, average turnout could be quite misleading for *individuals* within those groups and subgroups.

Geographically, some states such as Arkansas, Maine, Michigan, and Minnesota exhibited a high degree of variance, suggesting that there is a significant amount of heterogeneity among voters in these states. This finding is important because small area estimation techniques, such as MRP, rely upon national-level demographic trends to interpolate outcomes of interest within geographic subunits. These results suggest that differential levels of heterogeneity could complicate this process. Moreover, even within demographic groups

themselves, particular groups exhibit considerable heterogeneity. In particular, those with a high school education and those with some college education display a high degree of variance. Similarly, middle-aged (Age 30–44 and 45–64 categories) appear quite heterogeneous in their behavior, again complicating their inclusion in predictive models used for small area estimation.

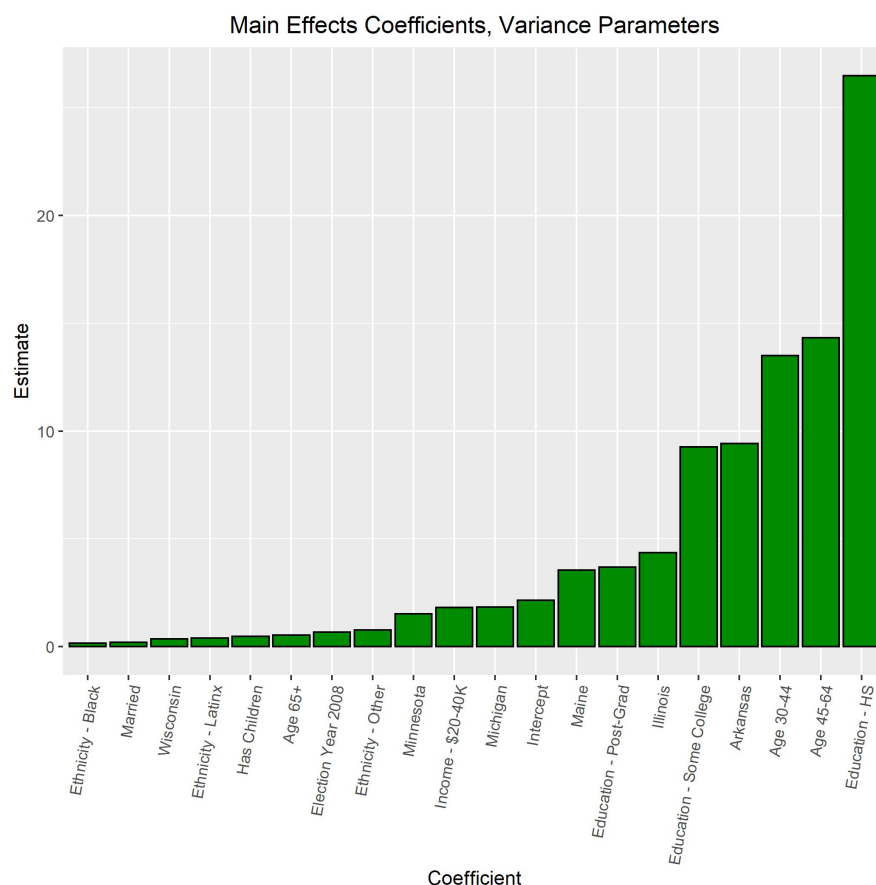


Figure 4. Selected Variance Parameters, Main Effects.

Subunits and intersectionalities also present similar patterns of heterogeneity. For example, African-Americans in particular states such as Michigan, Arizona, South Dakota, and Pennsylvania exhibited rather high levels of heterogeneity in turnout, as did Latinx Americans in New Mexico, Arizona, and Virginia. Other non-whites also vary considerably by geographic units: such voters in Texas, Alabama, and Utah rank especially high in their estimated variance parameters. Finally, our estimation procedure uncovered notable heterogeneity among a variety of income-related subunits. For instance, low-income (\$0–20k) Latinx Americans and other non-black/Latinx non-whites displayed notable variance in their turnout patterns. Likewise, lower-middle (\$20–40k) and middle-income whites displayed notable variances. Interactions between income and state also uncovered notable heterogeneity on occasion, with high-income Mississippians exhibit large variance, along with middle-income citizens in Maine and DC.

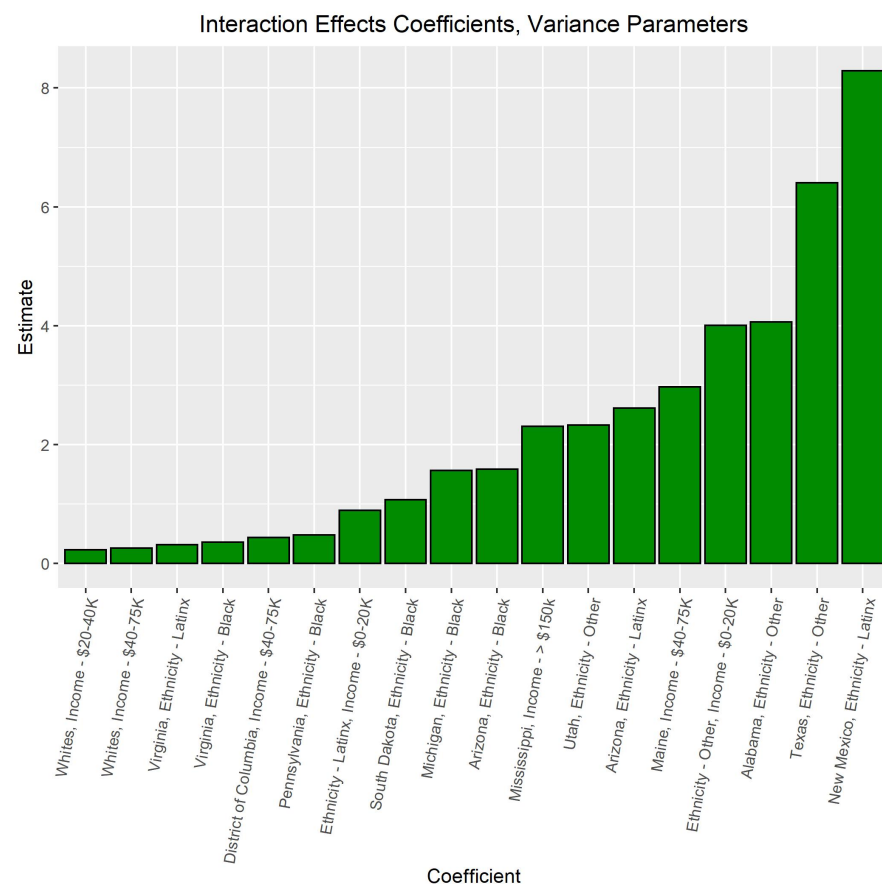


Figure 5. Selected Variance Parameters, Interactions.

4.4. Discussion and Implications

Typically, incorporation of individual-level heterogeneity into an analysis of large datasets such as the one used here in this study can be computationally infeasible under standard MCMC methods. As we have demonstrated, however, our polynomial expansion approach renders such models to indeed be estimated in real time.

Substantively, our results continue to affirm that black and young Latinx voters have been key drivers of turnout in 21st Century presidential contests, while low income voters continue to turn out much less frequently—even within high-turnout groups. From a policy perspective, some have suggested that mandatory voter registration (MVR) could help aid voter turnout [44]. However, in addition to concerns about election integrity that have arisen in response to federal MVR proposals (e.g., von Spakovsky 2013 [45]), our results also suggest that federal level policy may be poorly suited to the clear heterogeneity found among groups within different states. That is, given how differently the same groups and subgroups perform across state lines, single federal-level policy would almost surely be poorly suited to address these differences. We believe that future applications may be able to use our approach to further investigate these possibilities.

5. Conclusions and Future Research

High parameter models, particularly with large datasets, are becoming increasingly common in statistical modeling. For example, the use of small-area estimation techniques, such as MRP and Bayesian item response theory have generated a heightened interest in the estimation of models with large numbers of parameters [46–52]. Although the inclusion of such heterogeneity may not be feasible using existing MCMC methods, as we have demonstrated in this paper, we offer an alternative approach via polynomial expansions that offers insight regarding such heterogeneity amongst various groups groups and subgroups as evident in Figures 2–5.

From a statistical perspective, there are many potential avenues of future research that this study should encourage. For example, in order to render the integration associated with our polynomial expansions more feasible, we make the assumption in this study that $JN_i = 1$ in Equation (4). Future research can potentially look into weakening this restriction, which would enable this model to be applicable to settings with repeated measures. Additionally, a potential avenue of future research could be to explore other methods of polynomial expansions and compare them to the approach using geometric series expansions here. Furthermore, although we primarily concentrated on one particular model (binary logistic regression) and one class of priors (the normal distribution), we hope this study spurs research on closed-form Bayesian inferences for other models as well. In particular, a positive feature of members of the exponential family is that each member has a particular conjugate prior. It could be useful from a computational perspective to use polynomial expansions to approximate posterior distributions within this family for a choice of priors previously considered non-conjugate. Additionally, the binary logistic regression model discussed here belongs to a larger class of generalized linear models (GLMs) commonly called upon in applied research. A potential avenue of future research could be to utilize polynomial expansions to allow researchers to make closed-form Bayesian inferences based on other GLMs. Additionally, deriving a polynomial expansion approach for the multinomial logistic regression model, a workhorse model in applied econometrics, would also be a worthy endeavor of future research [13,53].

In sum, our methodology provides a series of both methodological and potential substantive advantage that can improve statistical modeling in both political science, public policy, as well as many other fields including where logistic regression is called upon including medicine, marketing, and sports modeling [2,13,15,20]. We thus hope that this approach provides yet another useful additional to the applied statistician's toolbox.

Supplementary Materials: Supplementary Materials referenced in this manuscript with additional information is available at: <https://www.mdpi.com/article/10.3390/stats5040070/s1>.

Author Contributions: Conceptualization, K.D. and J.C.; Data curation, C.H.; Formal analysis, K.D. and C.H.; Investigation, K.D.; Methodology, K.D.; Project administration, K.D.; Writing-original draft, K.D. and J.C.; Writing-review & editing, K.D., J.C. and 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 data used in this study has been made publicly available as part of the Ghitza and Gelman (2013) study [7].

Acknowledgments: We would like to thank the editors and the reviewers for useful comments on this manuscript. Additionally, we would like to thank Steven J. Miller, Norbert Michel, Benjamin Kedem, Tobias Von Petersdorff, P.K. Kannan, Paxton Wagner, Hans von Spakovsky, Parker Sheppard, Lavania Dayaratna, and Natashka Dayaratna for comments on prior versions of this manuscript as well as Bryan Wooley for help with compiling the data. We would also like to thank Steven J. Miller for providing us use of WHPCC for the model estimation presented in this paper [41].

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

References

1. Dayaratna, K. Contributions to Bayesian Statistical Modeling in Public Policy Research. Ph.D. Thesis, University of Maryland, College Park, MD, USA, 2014.
2. Kyung, M.; Gill, J.; Casella, G. New findings from terrorism data: Dirichlet process random-effects models for latent groups. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2011**, *60*, 701–721. [CrossRef]
3. Segal, J.A.; Cameron, C.M.; Cover, A.D. A spatial model of roll call voting: Senators, constituents, presidents, and interest groups in Supreme Court confirmations. *Am. J. Political Sci.* **1992**, *36*, 96–121. [CrossRef]
4. Swers, M.L. Are women more likely to vote for women's issue bills than their male colleagues? *Legis. Stud. Q.* **1998**, *23*, 435–448. [CrossRef]
5. Lauderdale, B.E.; Clark, T.S. Estimating Vote-Specific Preferences from Roll-Call Data Using Conditional Autoregressive Priors. *J. Politics* **2016**, *78*, 1153–1169. [CrossRef]

6. Hellwig, T.T. Interdependence, government constraints, and economic voting. *J. Politics* **2001**, *63*, 1141–1162. [\[CrossRef\]](#)
7. Ghitza, Y.; Gelman, A. Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *Am. J. Political Sci.* **2013**, *57*, 762–776. [\[CrossRef\]](#)
8. Powell, G.B., Jr. American voter turnout in comparative perspective. *Am. Political Sci. Rev.* **1986**, *80*, 17–43. [\[CrossRef\]](#)
9. Krupnikov, Y. When does negativity demobilize? Tracing the conditional effect of negative campaigning on voter turnout. *Am. J. Political Sci.* **2011**, *55*, 797–813. [\[CrossRef\]](#)
10. Krehbiel, K. Cosponsors and Wafflers from A to Z. *Am. J. Political Sci.* **1995**, *39*, 906–923. [\[CrossRef\]](#)
11. Howard, N.O.; Roberts, J.M. The Politics of Obstruction: Republican Holds in the US Senate. *Legis. Stud. Q.* **2015**, *40*, 273–294. [\[CrossRef\]](#)
12. Rector, R.; Johnson, K. Roles of Couples' Relationship Skills and Fathers' Employment in Encouraging Marriage. In *Heritage Foundation Center for Data Analysis Report*; The Heritage Foundation: Washington, DC, USA, 2004; Volume 4.
13. Guadagni, P.M.; Little, J.D. A logit model of brand choice calibrated on scanner data. *Mark. Sci.* **1983**, *2*, 203–238. [\[CrossRef\]](#)
14. Wisner, D.H. A stepwise logistic regression analysis of factors affecting morbidity and mortality after thoracic trauma: Effect of epidural analgesia. *J. -Trauma-Inj. Infect. Crit. Care* **1990**, *30*, 799–805. [\[CrossRef\]](#)
15. Albright, S.C. A statistical analysis of hitting streaks in baseball. *J. Am. Stat. Assoc.* **1993**, *88*, 1175–1183. [\[CrossRef\]](#)
16. Bucklin, R.E.; Gupta, S. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *J. Mark. Res.* **1992**, *29*, 201–215. [\[CrossRef\]](#)
17. Guadagni, P.M.; Little, J.D. When and what to buy: A nested logit model of coffee purchase. *J. Forecast.* **1998**, *17*, 303–326. [\[CrossRef\]](#)
18. Gupta, S.; Chintagunta, P.K. On using demographic variables to determine segment membership in logit mixture models. *J. Mark. Res.* **1994**, *31*, 128–136. [\[CrossRef\]](#)
19. Zhang, J.; Kai, F.Y. What is the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. *J. Am. Med. Assoc.* **1998**, *280*, 1690–1691. [\[CrossRef\]](#)
20. Bender, R.; Grouven, U. Using binary logistic regression models for ordinal data with non-proportional odds. *J. Clin. Epidemiol.* **1998**, *51*, 809–816. [\[CrossRef\]](#)
21. Kedem, B.; Fokianos, K. *Regression Models for Time Series Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2005; Volume 488.
22. Gelfand, A.E.; Hills, S.E.; Racine-Poon, A.; Smith, A.F. Illustration of Bayesian inference in normal data models using Gibbs sampling. *J. Am. Stat. Assoc.* **1990**, *85*, 972–985. [\[CrossRef\]](#)
23. Jochmann, M.; Leon-Gonzalez, R. Estimating the demand for health care with panel data: A semiparametric Bayesian approach. *Health Econ.* **2004**, *13*, 1003–1014. [\[CrossRef\]](#)
24. Kamakura, W.A.; Russell, G.J. A probabilistic choice model for market segmentation and elasticity structure. *J. Mark. Res.* **1989**, *26*, 379–390. [\[CrossRef\]](#)
25. McFadden, D.; Train, K. Mixed MNL models for discrete response. *J. Appl. Econom.* **2000**, *15*, 447–470. [\[CrossRef\]](#)
26. Allenby, G.M.; Arora, N.; Ginter, J.L. On the heterogeneity of demand. *J. Mark. Res.* **1998**, *35*, 384–389. [\[CrossRef\]](#)
27. Oh, M.S.; Choi, J.W.; Kim, D.G. Bayesian inference and model selection in latent class logit models with parameter constraints: An application to market segmentation. *J. Appl. Stat.* **2003**, *30*, 191–204. [\[CrossRef\]](#)
28. Dayaratna, K.D.; Kannan, P.K. A mathematical reformulation of the reference price. *Mark. Lett.* **2012**, *23*, 839–849. [\[CrossRef\]](#)
29. Williams, P. The Politics of Place: How Southern Identity Shapes Americans' Political Attitudes Furthermore, Policy Preferences. Ph.D. Thesis, University of Michigan, Ann Arbor, MI, USA, 2021.
30. Morris, C.N. Parametric empirical Bayes inference: Theory and applications. *J. Am. Stat. Assoc.* **1983**, *78*, 47–55. [\[CrossRef\]](#)
31. Revelt, D.; Train, K. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Rev. Econ. Stat.* **1998**, *80*, 647–657. [\[CrossRef\]](#)
32. Gelfand, A.E.; Smith, A.F. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **1990**, *85*, 398–409. [\[CrossRef\]](#)
33. Everson, P.J.; Bradlow, E.T. Bayesian inference for the beta-binomial distribution via polynomial expansions. *J. Comput. Graph. Stat.* **2002**, *11*, 202–207. [\[CrossRef\]](#)
34. Bradlow, E.T.; Hardie, B.G.; Fader, P.S. Bayesian inference for the negative binomial distribution via polynomial expansions. *J. Comput. Graph. Stat.* **2002**, *11*, 189–201. [\[CrossRef\]](#)
35. McShane, B.; Adrian, M.; Bradlow, E.T.; Fader, P.S. Count models based on Weibull interarrival times. *J. Bus. Econ. Stat.* **2008**, *26*, 369–378. [\[CrossRef\]](#)
36. Miller, S.J.; Bradlow, E.T.; Dayaratna, K. Closed-form Bayesian inferences for the logit model via polynomial expansions. *Quant. Mark. Econ.* **2006**, *4*, 173–206. [\[CrossRef\]](#)
37. Miller, S.J.; Takloo-Bighash, R. *An Invitation to Modern Number Theory*; Princeton University Press: Princeton, NJ, USA, 2006; Volume 13.
38. Shiryaev, A.N. *Probability*; Springer: New York, NY, USA, 1996.
39. Humaira, H.; Rasyidah, R. Determining the appropriate cluster number using Elbow method for K-Means algorithm. In *Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA)*, 2020.
40. Rossi, P.; McCulloch, R. Bayesm: Bayesian Inference for Marketing/Microeconometrics. R package. Available online: <https://cran.r-project.org/web/packages/bayesm/index.html> (accessed on 15 November 2022).

41. Williams College, Office of Information Technology. High Performance Computing Cluster Overview. Available online: <https://oit.williams.edu/itech/hpcc/> (accessed on 15 November 2022).
42. Benham, T.; Duan, Q.; Kroese, D.P.; Liqueur, B. CEoptim: Cross-entropy R package for optimization. *arXiv* **2015**, arXiv:1503.01842.
43. Verba, S. *Voice and Equality: Civic Voluntarism in American Politics*; Harvard University Press: Cambridge, MA, USA, 1995.
44. Chen, M.; Rosenberg, J. *Expanding Democracy: Voter Registration Around the World*; Brennan Center for Justice: New York, NY, USA, 2009.
45. von Spakovsky, H. Mandatory Voter Registration: How Universal Registration Threatens Electoral Integrity. *Heritage Foundation Backgrounder*; Oxford University Press: New York, NY, USA, 2013.
46. Molina, I.; Rao, J. *Small Area Estimation on Poverty Indicators*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
47. Hobza, T.; Morales, D.; Santamaría, L. Small area estimation of poverty proportions under unit-level temporal binomial-logit mixed models. *Test* **2018**, *27*, 270–294. [[CrossRef](#)]
48. Molina, I.; Rao, J. Small area estimation of poverty indicators. *Can. J. Stat.* **2010**, *38*, 369–385. [[CrossRef](#)]
49. Haslett, S.; Isidro, M.; Jones, G. Comparison of survey regression techniques in the context of small area estimation of poverty. *Surv. Methodol.* **2010**, *36*, 157–170.
50. Buil-Gil, D.; Moretti, A.; Shlomo, N.; Medina, J. Worry about crime in Europe: A model-based small area estimation from the European Social Survey. *Eur. J. Criminol.* **2021**, *18*, 274–298. [[CrossRef](#)]
51. Ghosh, M. Small area estimation: Its evolution in five decades. *Stat. Transit. New Ser.* **2020**, *21*, 1–22. [[CrossRef](#)]
52. Kubacki, J.; Jędrzejczak, A. Small area estimation of income under spatial SAR model. *Stat. Transit. New Ser.* **2016**, *17*, 365–390. [[CrossRef](#)]
53. McFadden, D. Conditional logit analysis of qualitative choice behavior. *Front. Econom.* **1973**, 105–142.