


## Article

# Bayesian Bell Regression Model for Fitting of Overdispersed Count Data with Application

Ameer Musa Imran Alhseeni<sup>1</sup> and Hossein Bevrani<sup>1,2,\*</sup> <sup>1</sup> Department of Statistics, University of Tabriz, Tabriz 51666-15648, Iran; a.alhseeni@tabrizu.ac.ir<sup>2</sup> Department of Statistics, University of Kurdistan, Sanandaj 66177-15175, Iran

\* Correspondence: bevrani@tabrizu.ac.ir or hossein.bevrani@uok.ac.ir

## Abstract

The Bell regression model (BRM) is a statistical model that is often used in the analysis of count data that exhibits overdispersion. In this study, we propose a Bayesian analysis of the BRM and offer a new perspective on its application. Specifically, we introduce a G-prior distribution for Bayesian inference in BRM, in addition to a flat-normal prior distribution. To compare the performance of the proposed prior distributions, we conduct a simulation study and demonstrate that the G-prior distribution provides superior estimation results for the BRM. Furthermore, we apply the methodology to real data and compare the BRM to the Poisson and negative binomial regression model using various model selection criteria. Our results provide valuable insights into the use of Bayesian methods for estimation and inference of the BRM and highlight the importance of considering the choice of prior distribution in the analysis of count data.

**Keywords:** Bayesian estimation; Bell regression model; G-prior distribution; log-marginal pseudo-likelihood; deviance information criterion

**MSC:** 62F15; 62F25; 62J12



Academic Editor: Wei Zhu

Received: 5 September 2025

Revised: 4 October 2025

Accepted: 7 October 2025

Published: 10 October 2025

**Citation:** Imran Alhseeni, A.M.; Bevrani, H. Bayesian Bell Regression Model for Fitting of Overdispersed Count Data with Application. *Stats* **2025**, *8*, 95. <https://doi.org/10.3390/stats8040095>

**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

Count regression models are valuable for understanding the relationships between predictor variables and count outcomes in various domains, offering a flexible and powerful framework for analyzing discrete, non-negative data. Count data regression analysis has a wide range of important applications. In biological and genetic studies, it is used to analyze data such as the number of genes, genetic mutations, or disease occurrences over time [1]. In epidemiology and public health, this analysis helps assess disease incidence, mortality rates, and the frequency of health events in specific populations [2]. Social scientists employ count data regression to model phenomena such as criminal offenses, births, and deaths within particular areas. Economic research utilizes this method to examine events including business failures, patents, and the frequency of various economic activities [3]. In insurance and actuarial science, it is applied to model claim frequencies and policyholder behavior [4]. Finally, environmental studies benefit from count data regression to analyze ecological counts, such as species diversity in habitats and wildlife populations [5].

Undoubtedly, the Poisson regression model (PRM) is the primary choice for analyzing count data. However, this model has a significant limitation: it assumes that the variance in the count variable is equal to its mean. This assumption is often violated in real-world datasets due to overdispersion, where the variance exceeds the mean. Consequently,

the PRM's applicability is limited in such scenarios, prompting the need for alternative models. One such alternative is the Bell regression model (BRM), introduced by [6], which has been well-received. BRMs have been widely discussed in various contexts, for example, in the presence of multicollinearity [7–13], excess zeros [14,15], and shrinkage strategies [16,17].

In statistical modeling, Bayesian inference has emerged as a powerful approach for data analysis. By combining prior knowledge with observed data, Bayesian methods enable the quantification of uncertainty and the estimation of model parameters. Although the Bell regression model has been applied in many contexts, most existing work has focused on frequentist estimation or shrinkage techniques. A Bayesian treatment of the BRM that utilizes informative priors is largely absent from the literature. In particular, no study has explored the use of G-prior distributions in this setting, even though the choice of prior is known to strongly influence Bayesian inference. This paper aims to fill that gap with several contributions. First, we propose a Bayesian formulation of the Bell regression model and introduce the use of G-priors, with hyperparameters specified through KL-divergence. Second, we develop a tailored Metropolis–Hastings algorithm for efficient posterior inference. Third, we conduct a simulation study that directly compares G-priors with the commonly used flat-normal prior, showing that G-priors consistently yield more accurate and stable estimates. Finally, we demonstrate the practical value of our approach through an application to mine fracture data, where the BRM with G-priors provides a better fit than both the Poisson and negative binomial regression models, according to several Bayesian model selection criteria. Together, these results highlight how incorporating G-priors into Bayesian BRMs can strengthen inference and offer new insights for applied researchers working with overdispersed count data. This paper was previously published as a preprint [18].

The remainder of the paper is structured as follows: Section 2 provides a detailed explanation of the BRM, including prior specification, posterior inference, and model selection criteria. In Section 3, a simulation study is carried out to compare the proposed prior distributions for the BRM. The methodology is then illustrated in Section 3.2, where the BRM is compared to the Poisson regression model (PRM) and the negative binomial regression model (NBRM) using Bayesian inference. Finally, Section 4 provides a discussion and concluding remarks.

## 2. Material and Methods

### 2.1. Bell Regression Model

The discrete Bell distribution is introduced by [6] based on a series expansion due to [19,20]. Its probability mass function is defined as follows:

$$f(y) = \frac{\theta^y e^{1-e^\theta} B_y}{y!}, \quad y = 0, 1, 2, \dots; \quad \theta > 0, \quad (1)$$

where  $B_y = e^{-1} \sum_{k=0}^{\infty} \frac{k^y}{k!}$  indicates the Bell numbers. The important characteristics of the Bell distribution are given by the following:

$$\begin{aligned} \mathbb{E}(Y) &= \theta e^\theta, \\ \mathbb{V}(Y) &= \theta e^\theta (1 + \theta) = \mathbb{E}(Y)(1 + \theta). \end{aligned}$$

Since  $\theta > 0$ , the variance value is greater than its mean, which is known as overdispersion. Consequently, the regression model is defined based on Bell distribution, which will be suitable for count data with overdispersion. In regression contents, it is common

to assume that the mean of the distribution depends on the vector of covariates. Let  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  be a random sample from  $Bell(\theta_i)$ . We relate the  $\mu_i := \mathbb{E}(Y_i)$  to  $p$  covariates by the log link function, i.e.,

$$\log(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta}, \quad i = 1, 2, \dots, n \quad (2)$$

where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  is the model parameter vector and  $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$  is the  $i$ th observation for  $p$  model covariates. Due to (2), the parameter of the Bell distribution will be  $\theta_i = W_0(\mu_i)$  ( $i = 1, 2, \dots, n$ ), where  $W_0(\cdot)$  is the Lambert function [21]. Hence, the model in (1) can be reparameterized as follows:

$$f(y_i | \mathbf{X}_i) = \frac{B_{y_i}}{y_i!} e^{1 - e^{W_0(\mu_i)}} [W_0(\mu_i)]^{y_i}. \quad (3)$$

## 2.2. Bayesian Inference

Consider  $n$  observations of the response–covariates pair as  $D = \{(y_i, \mathbf{X}_i), i = 1, 2, \dots, n\}$ . The likelihood function of the BRM is given as follows:

$$L(D | \boldsymbol{\beta}) = \left( \prod_{i=1}^n \frac{B_{y_i}}{y_i!} \right) \left( \prod_{i=1}^n [W_0(\mu_i)]^{y_i} \right) e^{n - \sum_{i=1}^n e^{W_0(\mu_i)}}, \quad (4)$$

where  $\mu_i = \exp \{ \mathbf{X}_i^T \boldsymbol{\beta} \}$ . Bayesian regression allows for the incorporation of prior knowledge about the parameters, which is particularly useful when such information is available or when making use of expert knowledge. The inference is based on the posterior distribution of BRM, i.e.,

$$\begin{aligned} \pi(\boldsymbol{\beta} | D) &= L(D | \boldsymbol{\beta}) \pi(\boldsymbol{\beta}) \\ &= \left( \prod_{i=1}^n \frac{B_{y_i}}{y_i!} \right) \left( \prod_{i=1}^n [W_0(\mu_i)]^{y_i} \right) e^{n - \sum_{i=1}^n e^{W_0(\mu_i)}} \times \pi(\boldsymbol{\beta}). \end{aligned} \quad (5)$$

Thus, we first need to specify the prior distribution.

### 2.2.1. Specification of Priors

First, we consider a common prior on  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$  for our model, the flat-normal prior  $N_p(\mathbf{0}, \tau^2 \mathbf{I}_p)$ , where  $\tau > 0$ . When we set a large value for  $\tau^2$ , the resulting flat prior becomes a diffuse prior. This diffuse prior assigns equal probability to all possible values of the regression coefficients. Unfortunately, this can lead to overestimating the magnitude of the regression coefficient and being overconfident about its sign. This is problematic because when it comes to regression coefficients (other than the intercept), we are typically more interested in knowing the magnitude and sign of the effect. Thus, we consider the second prior on model parameters based on the idea presented in [22].

In a proposed regression model, if the mean  $\mu$  of  $y_i$  is assumed to be in the range of  $(0, \infty)$ , and a subject matter expert has information on the marginal distribution of  $\mu$  characterized by inverse-gamma distribution with parameters  $a_\mu$  and  $b_\mu$ , i.e.,  $IG(a_\mu, b_\mu)$ , where  $a_\mu > 0$  and  $b_\mu > 0$  are known, the objective is to formulate a prior on  $\boldsymbol{\beta}$  that incorporates this prior information while adjusting for covariates. Following [23], we consider a G-prior distribution for the regression parameters. This choice is particularly suitable for BRMs because it allows for prior information on the marginal distribution of the mean response to be incorporated while adjusting for covariates, making it a natural and

flexible prior structure for regression models with overdispersed count data. The G-prior can be considered as follows:

$$\boldsymbol{\beta} \sim N_p \left( M\mathbf{u}, gn(\mathbf{X}^\top \mathbf{X})^{-1} \right), \quad (6)$$

where  $\mathbf{u} = (1, 0, \dots, 0)^\top$ ,  $M$  is a prior mean for the intercept and  $g > 0$  is a scaling constant. Suppose that  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \stackrel{iid}{\sim} H(x)$  with the mean  $A$  and covariance matrix  $\boldsymbol{\Sigma}$ . With  $\mathbf{X}_i$  including the intercept in the first element, the first element of  $A$  is one and entries in the first row and those in the first column of  $\boldsymbol{\Sigma}$  all are zero. For the new subject with covariates  $\mathbf{X} \sim H$  and response  $y$ , we can see that the mean of  $y$  equals  $\mu(\mathbf{X}_i) = h^{-1}(\mathbf{X}_i^\top \boldsymbol{\beta})$  and also assume that  $\mathbf{X}_i$  and  $\boldsymbol{\beta}$  are independent [22]. Therefore, we have

$$\mathbb{E}(\mathbf{X}_i^\top \boldsymbol{\beta}) = \mathbb{E}(\mathbb{E}(\mathbf{X}_i^\top \boldsymbol{\beta} | \mathbf{X}_i)) = \mathbb{E}(\mathbf{X}_i^\top M\mathbf{u}) = M,$$

and

$$\begin{aligned} \mathbb{V}(\mathbf{X}_i^\top \boldsymbol{\beta}) &= \mathbb{E}(\mathbb{V}(\mathbf{X}_i^\top \boldsymbol{\beta} | \mathbf{X}_i)) + \mathbb{V}(\mathbb{E}(\mathbf{X}_i^\top \boldsymbol{\beta} | \mathbf{X}_i)) \\ &= \mathbb{E}(gn\mathbf{X}_i^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}_i) + \mathbb{V}(M) \\ &= g \cdot \text{tr} \left( n(\mathbf{X}^\top \mathbf{X})^{-1} (\boldsymbol{\Sigma} + A A^\top) \right). \end{aligned}$$

Since  $n(\mathbf{X}^\top \mathbf{X})^{-1}$  converges in probability to  $(\boldsymbol{\Sigma} + A A^\top)^{-1}$ ,

$$\mathbb{V}(\mathbf{X}_i^\top \boldsymbol{\beta}) \xrightarrow{p} gp.$$

Ref. [24] found that for a various  $H(\cdot)$  considered in their simulations,  $\mathbf{X}_i^\top \boldsymbol{\beta}$  is approximately distributed as normal for any given value of  $\mathbf{X}_i$ . Therefore, it is reasonable to consider that  $\mathbf{X}_i^\top \boldsymbol{\beta} \sim N(M, gp)$ . Using these results, we can select the value of  $M$  and  $g$  in G-prior distribution so that the induced distribution of  $\mu(\mathbf{X}_i) = h^{-1}(\mathbf{X}_i^\top \boldsymbol{\beta})$  matches the marginal prior distribution  $\mu \sim IG(a_\mu, b_\mu)$ . We minimize the Kullback–Leibler divergence, specifically  $D_{KL}(P \parallel Q)$ , since this measures how well the distribution induced by our model ( $P$ ) approximates the prior information distribution ( $Q$ ). Thus, we have  $M = E(h(\mu))$  and  $g = \mathbb{V}(h(\mu))/p$ . In our regression model, we considered  $h(\cdot)$  as the log link function; hence,

$$M = \psi(a_\mu) + \log(b_\mu) \quad \text{and} \quad g = \frac{1}{p} \psi^{(1)}(a_\mu) \quad (7)$$

where  $\psi(\cdot)$  and  $\psi^{(1)}(\cdot)$  are digamma and trigamma function, respectively. When the values for  $a_\mu$  and  $b_\mu$  are not available, we use  $a_\mu = b_\mu = 1$  as the defaults, yielding relatively weak prior information on the location of  $\mu$ . It is worth noting that, under standard conditions, G-priors are known to yield consistent Bayesian estimators [23,24].

## 2.2.2. MCMC Algorithm

Consider the posterior distribution by using flat-normal prior and G-prior distribution in (6), which is analytically intractable. Monte Carlo Chain Markov (MCMC) simulation methods, such as the Gibbs sampler and Metropolis–Hastings algorithm, are utilized to obtain a sample from the posterior distribution [25]. To implement the Metropolis–Hastings algorithm, the following steps are taken:

1. Start with any point  $\boldsymbol{\beta}^{(0)}$  and stage indicator  $k = 0$ .

2. Generate  $\tilde{\beta}$  according to the transitional kernel  $K(\tilde{\beta}, \beta^{(k)}) = N_p(\beta^{(k)}, \tilde{\Sigma})$ , where  $\tilde{\Sigma}$  is a known symmetric positive defined matrix.
3. Accept  $\tilde{\beta}$  as  $\beta^{(k+1)}$  with the following probability:

$$\min \left\{ 1, \frac{\pi(\tilde{\beta}|D)}{\pi(\beta^{(k)}|D)} \right\}. \quad (8)$$

4. By increasing the stage indicator, repeat steps (1) to (3) until the process reaches a stationary distribution.

The computational program is available upon request from the authors.

### 2.3. Model Selection Criteria

There are multiple techniques available to compare and select the best-fitting model among several competing models for a given dataset. One commonly used technique in applied research is the conditional predictive ordinate (CPO) statistic. To learn more about the CPO and its applications in model selection, see [26,27]. Suppose that  $D$  is the full data,  $D_{(-i)}$  for  $i = 1, 2, \dots, n$  denotes the data with the  $i$ th observation deleted, and the posterior distribution based on  $D_{(-i)}$  is denoted by  $\pi(\beta|D_{(-i)})$ . For the  $i$ th observation,  $CPO_i$  is defined as follows:

$$CPO_i = \left[ \int_{\beta} \frac{\pi(\beta|D)}{f(y_i|\beta)} d\beta \right]^{-1}.$$

The low CPO values indicate poor model fit, but a closed-form CPO is unavailable for the proposed model. However, a Monte Carlo estimate of the  $CPO_i$  can be obtained by using a single MCMC sample  $\{\beta^{(1)}, \beta^{(2)}, \dots, \beta^{(T)}\}$  from the posterior distribution  $\pi(\beta|D)$ . Thus, the  $CPO_i$  can be approximated by

$$\widehat{CPO}_i = \left[ \frac{1}{T} \sum_{k=1}^T \frac{1}{f(y_i|\beta^{(k)})} \right]^{-1}.$$

The statistic for model comparison is the log-marginal pseudo-likelihood (LMPL) defined as follows:

$$LMPL = \sum_{i=1}^n \log(\widehat{CPO}_i). \quad (9)$$

Therefore, the largest value of  $LMPL$  indicates that the data is well-fitted by the model under consideration.

The second criterion which is proposed by [28] is called the deviance information criterion (DIC). Based on MCMC samples, the DIC can be estimated as follows:

$$\widehat{DIC} = 2 \left[ -2 \frac{1}{T} \sum_{k=1}^T \log L(D|\beta^{(k)}) + \log L(D|\bar{\beta}) \right], \quad (10)$$

where  $\bar{\beta}$  is the mean of MCMC samples. The model with the lowest  $\widehat{DIC}$  value is considered as the best-fitted model.

The final two criteria considered here are the expected Akaike information criterion (EAIC) by [29] and the expected Bayesian information criterion (EBIC) by [30]. These criteria can also be estimated using the following:

$$\widehat{EAIC} = -2 \sum_{k=1}^T \log L(D|\beta^{(k)}) + 2p, \quad (11)$$

and

$$\widehat{EBIC} = -2 \sum_{k=1}^T \log L(D|\beta^{(k)}) + p \log(n), \quad (12)$$

where  $p$  is the number of model parameters. The model that exhibits the lowest value of these criteria, similar to the DIC, is considered to be a better fit for the data.

### 3. Results

#### 3.1. Simulation Study

In this section, we conduct a simulation study to illustrate the implementation of the proposed regression methodology.

The model we consider here is given by

$$\mu_i = \exp \{ \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \}, \quad i = 1, 2, \dots, n. \quad (13)$$

Here, we consider the model with intercept; thus, we set  $x_{i1} = 1$ . The observations for the covariates, i.e.,  $x_{i2}, \dots, x_{ip}$ , are generated from standard normal distribution. The real value of the parameters in model (13) is considered as follows:

$$\beta = (0, -0.5, \underbrace{1, 1, \dots, 1}_{p-2})^\top. \quad (14)$$

In summary, we obtained the response observations from  $Bell(W_0(\mu_i))$ . In order to evaluate the effectiveness of our proposed method, we tested it with different sample sizes  $n = 50, 100, 200$  to represent low, medium, and high sample sizes, and with different numbers of covariates  $p = 3, 6$ . For the prior distributions, we consider flat-normal distribution with  $\tau = 10^2$  and G-prior defined in (6) with hyperparameters obtained from (7) by setting  $a_\mu = b_\mu = 1$ .

We generate two parallel independent MCMC runs of size  $T = 50,000$  for each posterior distribution and discard the 10,000 first generated sample as the burn-in to eliminate the impact of initial values. For computing the posterior estimates, we used every 20th sample to reduce the autocorrelations of the generated chains and yield better convergence results. In the Metropolis–Hastings algorithm, the proposal covariance was set equal to the covariance matrix of the maximum likelihood estimator. The convergence of the MCMC chains was monitored using the trace and autocorrelation function (ACF), Heidelberger–Welch, and Gelman–Rubin convergence diagnostics [31]. Additionally, we performed a small sensitivity study to evaluate the robustness of the model with respect to the choice of hyperparameters in the prior distributions by testing different values of  $\tau$  and  $M$  for prior distributions.

The summaries of the parameters in the posterior distribution exhibit minimal differences and do not impact the results presented in Tables 1 and 2. These tables display the results of posterior inference for both prior distributions. These tables show the estimated values using the squared loss function, posterior standard deviations (PSD), and 95% highest posterior density (HPD) intervals. The results show that as the sample size grows, both the posterior standard deviations (PSDs) and the widths of the 95% HPD intervals become smaller. Across all cases, the G-prior gives tighter intervals and lower PSDs than the flat-normal prior, which means that it provides more precise estimates. In addition, the acceptance rate of the algorithm stayed high (90% to 95%) for both priors, indicating that the sampling procedure worked efficiently.

Table 3 presents the mean squared errors (MSEs) and the mean absolute errors (MAEs) for the estimates found in Tables 1 and 2. Table 3 shows that the MSE and MAE of the

G-prior are consistently lower than those of the flat-normal prior, except in the case  $p = 6$  and  $n = 200$ . This suggests that the G-prior generally produces more accurate and stable parameter estimates, with particularly clear benefits when the sample size is small or moderate. For both prior choices, we also observe that increasing the number of covariates leads to higher MSE and MAE values.

**Table 1.** Bayesian inference of model (13) based on prior distributions when  $p = 3$ .

		G-Prior				Flat Normal			
		Estimate	PSD	95% HPD		Estimate	PSD	95% HPD	
True Value				Lower	Upper			Lower	Upper
<i>n</i> = 50									
$\beta_1$	0	−0.0984	0.2022	−0.4795	0.3073	−0.1088	0.2135	−0.5245	0.3017
$\beta_2$	−0.5	−0.4743	0.1830	−0.8319	−0.0946	−0.4990	0.1874	−0.8604	−0.1053
$\beta_3$	1	1.0369	0.1440	0.7547	1.3125	1.0629	0.1482	0.7726	1.3460
<i>n</i> = 100									
$\beta_1$	0	−0.1609	0.1430	−0.4254	0.1273	−0.1636	0.1465	−0.4449	0.1175
$\beta_2$	−0.5	−0.5825	0.1168	−0.8042	−0.3377	−0.5903	0.1172	−0.8164	−0.3505
$\beta_3$	1	1.0240	0.0876	0.8479	1.1900	1.0329	0.0891	0.8569	1.2067
<i>n</i> = 200									
$\beta_1$	0	−0.1171	0.1051	−0.3133	0.0911	−0.1202	0.1081	−0.3203	0.0926
$\beta_2$	−0.5	−0.4977	0.0817	−0.6432	−0.3273	−0.5007	0.0822	−0.6459	−0.3300
$\beta_3$	1	1.0404	0.1067	0.8329	1.2342	1.0527	0.1091	0.8369	1.2478

**Table 2.** Bayesian inference of model (13) based on prior distributions when  $p = 6$ .

		G-Prior				Flat Normal			
				95% HPD				95% HPD	
	True Value	Estimate	PSD	Lower	Upper	Estimate	PSD	Lower	Upper
$n = 50$									
$\beta_1$	0	0.0045	0.2046	−0.4025	0.3899	−0.0801	0.2286	−0.5406	0.3511
$\beta_2$	−0.5	−0.5517	0.1238	−0.8012	−0.3143	−0.5742	0.1282	−0.8326	−0.3259
$\beta_3$	1	0.9397	0.1027	0.7257	1.1282	0.9798	0.1089	0.7535	1.1779
$\beta_4$	1	0.8249	0.1167	0.5838	1.0548	0.8750	0.1202	0.6326	1.1170
$\beta_5$	1	0.9214	0.1046	0.7168	1.1318	0.9283	0.1087	0.7099	1.1381
$\beta_6$	1	1.0821	0.1615	0.7607	1.3963	1.1515	0.1770	0.8025	1.4978
$n = 100$									
$\beta_1$	0	−0.1425	0.1471	−0.4346	0.1352	−0.2118	0.1588	−0.5130	0.0938
$\beta_2$	−0.5	−0.4448	0.0803	−0.6019	−0.2862	−0.4491	0.0810	−0.6077	−0.2882
$\beta_3$	1	1.0378	0.0571	0.9199	1.1507	1.0550	0.0585	0.9329	1.1725
$\beta_4$	1	0.8337	0.1112	0.6046	1.0405	0.8784	0.1147	0.6420	1.0970
$\beta_5$	1	0.8680	0.0866	0.7029	1.0404	0.9007	0.0889	0.7356	1.0770
$\beta_6$	1	1.0771	0.1061	0.8626	1.2812	1.1223	0.1132	0.8899	1.3387
$n = 200$									
$\beta_1$	0	−0.0152	0.0981	−0.2078	0.1788	−0.0406	0.1016	−0.2321	0.1737
$\beta_2$	−0.5	−0.4573	0.0542	−0.5611	−0.3551	−0.4670	0.0546	−0.5710	−0.3645
$\beta_3$	1	0.8729	0.0674	0.7455	1.0197	0.8881	0.0684	0.7556	1.0359
$\beta_4$	1	0.9738	0.0577	0.8647	1.0938	0.9837	0.0596	0.8677	1.1038
$\beta_5$	1	1.0507	0.0552	0.9496	1.1600	1.0565	0.0562	0.9520	1.1654
$\beta_6$	1	0.9395	0.0658	0.8172	1.0727	0.9539	0.0669	0.8329	1.0912



**Table 3.** MSEs and MAEs for the estimated values reported in Tables 1 and 2.

<i>n</i>	MSE				MAE			
	G-Prior		Flat Normal		G-Prior		Flat Normal	
	<i>p</i> = 3	<i>p</i> = 6	<i>p</i> = 3	<i>p</i> = 6	<i>p</i> = 3	<i>p</i> = 6	<i>p</i> = 3	<i>p</i> = 6
50	0.0036	0.0087	0.0052	0.0093	0.0537	0.0754	0.0576	0.0871
100	0.0111	0.0126	0.0120	0.0150	0.0891	0.1018	0.0956	0.1102
200	0.0051	0.0042	0.0057	0.0035	0.0533	0.0537	0.0579	0.0507

### 3.2. Application

In this section, we apply our methodology to a mine fracture dataset that was primarily analyzed by [32]. The dataset includes four variables: the thickness of the inner burden in feet ( $x_1$ ), the percentage of extraction from the previously mined lower seam ( $x_2$ ), the height of the lower seam ( $x_3$ ), and the time since the mine has been opened ( $x_4$ ). The number of fractures in the mine is denoted by  $y$ , and there are 44 observations available for each variable.

We examine three count regression models: the Bell regression model (BRM), the Poisson regression model (PRM), and the negative binomial regression model (NBRM). To begin, we assess whether the response variable follows each distribution using a chi-square goodness-of-fit test at the 95% confidence level. The results, presented in Table 4, indicate that the Poisson distribution does not provide an adequate fit for this dataset, despite its previous use in similar analyses by [32,33]. In contrast, both the Bell and negative binomial distributions demonstrate a good fit, highlighting their suitability for modeling this data.

As in the Simulation Section, MCMC runs of size 50,000 were generated for each posterior distribution of BRM, PRM and NBRM under the G-prior distribution. To reduce the autocorrelations of the generated chains and obtain better convergence results, every 20th sample was used after discarding the first 10,000 generated samples as burn-in. The posterior means, medians, PSDs, and 95% HPD intervals for both regression models are presented in Table 5. The results indicate that the BRM provides estimations with the lowest PSD and narrowest HPD intervals. Moreover, the HPD intervals of the BRMs indicate that  $x_1$  and  $x_3$  are not significant while only  $x_2$  is significant based on the HPD intervals of PRM and NBRM.

We compared the BRM, PRM, and NBRM using the LMPL, DIC, EAIC, and EBIC criteria in Table 6. The results show that the BRM consistently performs best across all measures. In the real data application (Tables 5 and 6), the BRM with a G-prior not only yields narrower HPD intervals but also achieves stronger model selection scores. While the NBRM offers an improvement over the PRM, the BRM still provides the most reliable fit, even with a small sample size. This highlights the practical usefulness of the Bayesian BRM for applied researchers working with limited or overdispersed count data.

**Table 4.** Goodness-of-fit test for the mine fracture dataset.

Count	Observed	Bell	Poisson	Negative Binomial
0	10	10.149	4.744	6.640
1	7	9.164	10.567	10.752
2	8	8.274	11.767	10.173
3	8	6.226	8.736	7.342
4	4	4.216	4.864	4.475
≥5	7	5.970	3.321	4.618
$\chi^2$		1.216	12.523	4.813
<i>p</i> -value		0.943	0.028	0.439



**Table 5.** Bayesian inference of models for the mine fracture dataset.

Model	Parameter	Mean	Median	PSD	95% HPD	
					Lower	Upper
Bell	$\beta_1$	−3.5991	−3.5864	1.0666	−5.6350	−1.6295
	$\beta_2$	−0.0015	−0.0015	0.0009	−0.0032	0.0002
	$\beta_3$	0.06310	0.0634	0.0127	0.0380	0.0879
	$\beta_4$	−0.0032	−0.0031	0.0055	−0.0148	0.0069
	$\beta_5$	−0.0323	−0.0316	0.0164	−0.0666	−0.0032
Poisson	$\beta_1$	−4.1262	−4.0515	1.4028	−7.0078	−1.7307
	$\beta_2$	−0.0015	−0.0014	0.0011	−0.0035	0.0006
	$\beta_3$	0.0695	0.0689	0.0168	0.0386	0.1031
	$\beta_4$	−0.0031	−0.0027	0.0074	−0.0174	0.0116
	$\beta_5$	−0.0314	−0.0304	0.0227	−0.0764	0.0117
Negative Binomial	$\beta_1$	−3.9884	−3.9600	1.2110	−6.6541	−1.8616
	$\beta_2$	−0.0014	−0.0014	0.0010	−0.0032	0.0005
	$\beta_3$	0.0675	0.0674	0.0148	0.0414	0.0987
	$\beta_4$	−0.0027	−0.0024	0.0060	−0.0139	0.0095
	$\beta_5$	−0.0314	−0.0315	0.0186	−0.0665	0.0059

**Table 6.** Bayesian Criteria for the fitted models to the mine fracture dataset.

Model	LMPL	DIC	EAIC	EBIC
Bell	−72.4172	144.4801	149.3860	158.3070
Poisson	−78.1424	156.8615	161.5732	170.4942
Negative Binomial	−75.7465	149.3917	154.5123	163.4332

#### 4. Discussion and Conclusions

This paper has presented a comprehensive Bayesian framework for Bell regression models (BRMs). To the best of our knowledge, it is the first study to introduce G-priors in this context, underscoring the critical importance of prior choice when modeling overdispersed count data. Alongside the G-prior, we considered a flat-normal prior, with inference carried out using a tailored MCMC algorithm. Our simulation study demonstrated that the G-prior consistently yields more precise parameter estimates, as evidenced by smaller posterior standard deviations, narrower HPD intervals, and lower MSE and MAE compared to the flat-normal prior. In the real-data application, the BRM with a G-prior also outperformed both Poisson and negative binomial regression models across several Bayesian model selection criteria, including the LMPL, DIC, EAIC and EBIC variants. These results highlight not only the methodological contribution but also the practical utility of the proposed approach. Unlike earlier studies that focused primarily on frequentist estimation, this work provides a Bayesian alternative that performs well even with modest sample sizes. Although a full asymptotic analysis is beyond the scope of this paper, our empirical findings align with the established theoretical advantages of G-priors in the literature [23,24]. Future work could involve a detailed examination of the asymptotic properties and an extension of the method to scenarios such as zero-inflated BRMs or the use of other informative prior families to further enhance the model's flexibility and performance.

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

**Funding:** The authors received no specific funding for this work.

**Data Availability Statement:** The data supporting this paper are from previously reported studies and datasets, which have been cited.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

- Montesinos-López, O.A.; Montesinos-López, A.; Pérez-Rodríguez, P. Genomic Prediction Models for Count Data. *J. Agric. Biol. Environ. Stat.* **2015**, *20*, 533–554. [\[CrossRef\]](#)
- Du, J.; Park, Y.T.; Theera-Ampornpunt, N.; McCullough, J.S.; Speedie, S.M. The use of count data models in biomedical informatics evaluation research. *J. Am. Med. Inform. Assoc.* **2015**, *19*, 39–44. [\[CrossRef\]](#)
- Cameron, A.C.; Trivedi, P.K. *Regression Analysis of Count Data*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2013.
- Frees, E.W. *Regression Modeling with Actuarial and Financial Applications*; Cambridge University Press: Cambridge, UK, 2010.
- Lee, D.; Neocleous, T. Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc. C Appl. Stat.* **2010**, *59*, 905–920. [\[CrossRef\]](#)
- Castellares, F.; Ferrari, S.L.P.; Lemonte, A.J. On the Bell distribution and its associated regression model for count data. *Appl. Math. Model.* **2018**, *56*, 172–185. [\[CrossRef\]](#)
- Abduljabbar, L.A.; Algamal, Z.Y. Jackknifed K–L estimator in Bell regression model. *Math. Stat. Eng. Appl.* **2022**, *71*, 267–278.
- Algamal, Z.Y.; Lukman, A.; Golam, B.M.K.; Taofik, A. Modified Jackknifed Ridge Estimator in Bell Regression Model: Theory, Simulation and Applications. *Iraqi J. Comput. Sci. Math.* **2023**, *4*, 146–154. [\[CrossRef\]](#)
- Amin, M.; Akram, M.N.; Majid, A. On the estimation of Bell regression model using ridge estimator. *Commun. Stat. Simul. Comput.* **2021**, *52*, 854–867. [\[CrossRef\]](#)
- Bulut, Y.M.; Lukman, A.F.; Işilar, M.; Adewuyi, E.T.; Algamal, Z.Y. Modified ridge estimator in the Bell regression model. *J. Inverse Ill-Posed Probl.* **2024**, *32*, 1081–1091. [\[CrossRef\]](#)
- Ertan, E.; Algamal, Z.Y.; Erkoç, A.; Ulaş Akay, K. A new improvement Liu-type estimator for the Bell regression model. *Commun. Stat. B Simul. Comput.* **2023**, *54*, 603–614. [\[CrossRef\]](#)
- Majid, A.; Amin, M.; Akram, M.N. On the Liu estimation of Bell regression model in the presence of multicollinearity. *J. Stat. Comput. Simul.* **2021**, *92*, 262–282. [\[CrossRef\]](#)
- Shewaa, G.A.; Ugwuowo, F.I. Combating the Multicollinearity in Bell Regression Model: Simulation and Application. *J. Niger. Soc. Phys. Sci.* **2022**, *4*, 713. [\[CrossRef\]](#)
- Algamal, Z.Y.; Lukman, A.F.; Abonazel, M.R.; Awwad, F.A. Performance of the Ridge and Liu Estimators in the zero-inflated Bell Regression Model. *J. Math.* **2022**, *2022*, 9503460. [\[CrossRef\]](#)
- Lemonte, A.J.; Moreno-Arenas, G.; Castellares, F. Zero-inflated Bell regression models for count data. *J. Appl. Stat.* **2020**, *47*, 265–286. [\[CrossRef\]](#) [\[PubMed\]](#)
- Seifollahi, S.; Bevrani, H.; Algamal, Z.Y. Improved estimators in Bell regression model with application. *J. Stat. Comput. Simul.* **2024**, *94*, 2710–2726. [\[CrossRef\]](#)
- Seifollahi, S.; Bevrani, H.; Algamal, Z.Y. Shrinkage estimators in zero-inflated Bell regression model with application. *J. Stat. Theory Pract.* **2025**, *19*, 1. [\[CrossRef\]](#)
- Imran Alhaseeni, A.M.; Bevrani, H. Bayesian Bell regression model for fitting of overdispersed count data with application. *arXiv* **2024**, arXiv:2403.07067. [\[CrossRef\]](#)
- Bell, E.T. Exponential numbers. *Am. Math. Mon.* **1934**, *41*, 419. [\[CrossRef\]](#)
- Bell, E.T. Exponential polynomials. *Ann. Math.* **1934**, *35*, 258. [\[CrossRef\]](#)
- Corless, R.M.; Gonnet, G.H.; Hare, D.E.; Jeffrey, D.J.; Knuth, D.E. On the Lambert W function. *Adv. Comput. Math.* **1996**, *5*, 329–359. [\[CrossRef\]](#)
- Zhou, H.; Huang, X. Bayesian beta regression for bounded responses with unknown supports. *Comput. Stat. Data Anal.* **2022**, *167*, 107345. [\[CrossRef\]](#)
- Zellner, A. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*; North-Holland/Elsevier: Amsterdam, The Netherlands, 1986; pp. 233–243.
- Hanson, T.E.; Branscum, A.J.; Johnson, W.O. Informative g-priors for logistic regression. *Bayesian Anal.* **2014**, *9*, 597–612. [\[CrossRef\]](#)
- Gamerman, D.; Lopes, H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*; CRC Press: Boca Raton, FL, USA, 2006.
- Geisser, S.; Eddy, W. A predictive approach to model selection. *J. Am. Stat. Assoc.* **1979**, *74*, 153–160. [\[CrossRef\]](#)
- Gelfand, A.; Dey, D.; Chang, H. *Model Determination Using Predictive Distributions with Implementation via Sampling Based Methods (with Discussion)*; Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M., Eds.; Oxford University Press: Oxford, UK, 1992; Volume 1, pp. 7–167. [\[CrossRef\]](#)

- 
28. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B Stat. Methodol.* **2002**, *64*, 583–639. Available online: <http://www.jstor.org/stable/3088806> (accessed on 23 October 2002). [[CrossRef](#)]
  29. Brooks, S.P.; Smith, J.; Vehtari, A.; Plummer, M.; Stone, M.; Robert, C.P.; Titterton, D.M.; Nelder, J.A.; Atkinson, A.; Dawid, A.P.; et al. Discussion on the paper by Spiegelhalter, Best, Carlin, and van der Linde. *J. R. Stat. Soc. B Stat. Methodol.* **2002**, *64*, 616–639.
  30. Carlin, B.; Louis, T. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Chapman & Hall/CRC: Boca Raton, FL, USA, 2001.
  31. Chen, M.; Shao, Q.; Ibrahim, J. *Monte Carlo Methods in Bayesian Computation*; Springer: New York, NY, USA, 2000.
  32. Myers, R.H.; Montgomery, D.C.; Vining, G.G.; Robinson, T.J. *Generalized Linear Models: With Applications in Engineering and the Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
  33. Lukman, A.F.; Arashi, M.; Prokaj, V. Robust biased estimators for Poisson regression model: Simulation and applications. *Concurr. Comput. Pract. Exp.* **2023**, *35*, e7594. [[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.