

# Bayesian Zero-Inflated Negative Binomial Regression Based on Pólya-Gamma Mixtures

Brian Neelon

**Abstract.** Motivated by a study examining spatiotemporal patterns in inpatient hospitalizations, we propose an efficient Bayesian approach for fitting zero-inflated negative binomial models. To facilitate posterior sampling, we introduce a set of latent variables that are represented as scale mixtures of normals, where the precision terms follow independent Pólya-Gamma distributions. Conditional on the latent variables, inference proceeds via straightforward Gibbs sampling. For fixed-effects models, our approach is comparable to existing methods. However, our model can accommodate more complex data structures, including multivariate and spatiotemporal data, settings in which current approaches often fail due to computational challenges. Using simulation studies, we highlight key features of the method and compare its performance to other estimation procedures. We apply the approach to a spatiotemporal analysis examining the number of annual inpatient admissions among United States veterans with type 2 diabetes.

**Keywords:** zero inflation, zero-inflated negative binomial, Pólya-Gamma distribution, data augmentation, spatiotemporal data.

## 1 Introduction

Count data with an abundance of zeros arise commonly in many scientific fields, including ecology, infectious disease epidemiology, and health services research. Consider, for example, our motivating application, which examines the number of inpatient hospitalizations among United States (US) veterans with type 2 diabetes. The majority of patients had no inpatient admissions, resulting in a count of zero, while some had a handful of admissions and a small fraction had numerous admissions. When the number of zeros is greater than expected under a standard count model, the data are said to be zero inflated relative to the standard model. Zero-inflated count data often require flexible two-part mixture models to address both the excess zeros and the heterogeneous distribution of nonzero counts. A common choice is the zero-inflated model (Lambert, 1992), which is a mixture of a point mass that accounts for the excess zeros and a count distribution for the remaining values. The zero-inflated negative binomial (ZINB) model is a popular choice for modeling zero-inflated data because it simultaneously accommodates zero inflation and overdispersion in the count portion of the model.

Frequentist inference for the ZINB model is carried out using Newton-Raphson routines or the EM algorithm, where the excess zeros are treated as a type of missing data. However, frequentist procedures become computationally challenging for complex data structures, including longitudinal, spatial and spatiotemporal data that incorporate multivariate random effects. This has prompted increased interest in tractable Bayesian

---

Medical University of South Carolina, Charleston, SC, [neelon@musc.edu](mailto:neelon@musc.edu)

approaches to fitting zero-inflated models (Ghosh et al., 2006; Neelon et al., 2010; Zurr et al., 2012). Bayesian inference for the ZINB model is typically implemented in pre-packaged Bayesian software such as WinBUGS (Lunn et al., 2014). While such programs are suitable for relatively simple models, they become computationally infeasible for fitting zero-inflated models with high-dimensional random effects. In our motivating application, for example, WinBUGS was incapable of fitting a ZINB model with spatially correlated random intercepts and slopes.

To allow for additional flexibility, we propose a computationally efficient Bayesian approach to fitting ZINB models that is specifically designed to handle high-dimensional data where existing methods often fail. We augment the data by introducing two latent variables, each following a mixture of normal distributions with independent Pólya-Gamma precision terms (Polson et al., 2013a). As such, our model extends the approach of Polson et al. (2013a) and Pillow and Scott (2012) to the zero-inflated setting. Because the latent variables are conditionally normal, they admit the convenient posterior distributions available under standard Bayesian linear model theory. This leads to efficient Gibbs sampling routines, and enables closed-form updates for various random effect models.

The remainder of the paper is organized into four sections. Section 2 describes the proposed ZINB model, outlines the Bayesian model fitting approach, and discusses extensions to mixed effects models for longitudinal and spatial data. Section 3 presents numerical examples that highlight salient properties of the model and the proposed Gibbs sampler. In Section 4, we apply the model to a study examining spatiotemporal patterns in inpatient admissions among US veterans residing in three southeastern states from 2011–2015. The final section provides a discussion and offers directions for future research.

## 2 Bayesian Zero-Inflated Negative Binomial Model

### 2.1 The Zero-Inflated Negative Binomial Model

Zero-inflated models are mixtures of a point mass at zero, representing the excess zeros, and a count distribution for the remaining values. The term “excess” denotes the fact that the data contain more zeros than expected under a standard count model. By construction, zero-inflated models partition zeros into two types. The first type, typically referred to as a “structural” zero, corresponds to individuals who are not at risk for an event, and therefore have no opportunity for a positive count. The second type, termed the “at-risk” or “chance” zero, applies to a latent class of individuals who are at risk for an event but nevertheless have an observed response of zero. For example, in our application examining the number of inpatient hospitalizations, the structural zeros might represent patients who are in good health or can be treated through outpatient care, and thus have no recorded inpatient days. In contrast, the at-risk zeros might correspond to patients with more serious chronic conditions who, for various reasons, have had no inpatient admissions in a given year. Thus, zero-inflated models can be viewed as latent class models in which the classes are formed by the two types of zeros.

The ZINB model is a common choice for modeling zero-inflated data because it addresses not only zero inflation, but also overdispersion among the counts in the at-risk class. In its generic form, the ZINB model is expressed as

$$y_i \sim (1 - \pi_i)\mathbb{1}_{(w_i=0 \wedge y_i=0)} + \pi_i\text{NB}(\mu_i, r)\mathbb{1}_{(w_i=1)}, \quad i = 1, \dots, n, \quad (1)$$

where  $y_i$  is the count response for individual  $i$ ,  $\mathbb{1}_{(\cdot)}$  is the indicator function, and  $w_i$  is a latent “at-risk” indicator variable such that with probability  $1 - \pi_i$ ,  $w_i = y_i = 0$  implying a structural zero, and with probability  $\pi_i$ ,  $w_i = 1$  and  $y_i$  is in turn drawn from a negative binomial distribution with mean  $\mu_i$  and dispersion parameter  $r > 0$ . Thus,  $\pi_i$  denotes the probability of being in the at-risk class while  $\mu_i$  denotes the mean count among the at-risk population.

The at-risk indicators,  $w_1, \dots, w_n$ , are typically modeled using a logistic model of the form

$$\text{logit}(\pi_i) = \text{logit}[\Pr(w_i = 1 | \beta_1)] = \mathbf{x}_i^T \beta_1 = \eta_{1i}, \quad (2)$$

where  $\mathbf{x}_i$  is a  $p \times 1$  vector of covariates and  $\beta_1$  is a vector of regression parameters. Equation (2) is commonly referred to as the binary or logistic component of the ZINB model, which we denote with the subscript “1” in equation (2). Next, for reasons discussed below, we follow Pillow and Scott (2012) and parameterize the negative binomial component (conditional on  $w_i = 1$ ) as

$$\begin{aligned} p(y_i | r, \beta_2, w_i = 1) &\stackrel{d}{=} \frac{\Gamma(y_i + r)}{\Gamma(r)y_i!} (1 - \psi_i)^r \psi_i^{y_i} \quad \forall i \text{ s.t. } w_i = 1, \quad \text{where} \\ \psi_i &= \frac{\exp(\mathbf{x}_i^T \beta_2)}{1 + \exp(\mathbf{x}_i^T \beta_2)} = \frac{\exp(\eta_{2i})}{1 + \exp(\eta_{2i})}. \end{aligned} \quad (3)$$

Equation (3) is often referred to as the count or negative binomial component of the ZINB model, which we denote with the subscript “2”. The expected value and variance among the counts in the at-risk class are

$$\begin{aligned} E(y_i | r, \beta_2, w_i = 1) &= \frac{r\psi_i}{1 - \psi_i} = r \exp(\eta_{2i}) = \mu_i, \\ \text{Var}(y_i | r, \beta_2, w_i = 1) &= \frac{r\psi_i}{(1 - \psi_i)^2} = r \exp(\eta_{2i}) [1 + \exp(\eta_{2i})] \\ &= \mu_i(1 + \mu_i/r). \end{aligned} \quad (4)$$

The marginal mean, averaged over  $w_i$ , is  $E(y_i) = \pi_i \mu_i$ . The parameter  $\alpha = 1/r$  captures the overdispersion in at-risk class, so that as  $\alpha \rightarrow \infty$ , the at-risk counts become increasingly dispersed relative to the Poisson. Above, we have assumed the same set of covariates  $\mathbf{x}_i$  for both the binary and count components, but in general this is not necessary.

## 2.2 Bayesian Inference for the ZINB Model

We now outline the posterior sampling algorithm for the fixed effects ZINB model. The details are presented in the following sub-sections, but in brief, the algorithm proceeds in four steps:

1. Given current parameter values, update the latent at-risk indicators,  $w_1, \dots, w_n$ , from their discrete full conditional distributions
2. Update  $\beta_1$  using the Gibbs sampler proposed by Polson et al. (2013a) for logistic regression
3. Conditional on  $w_i = 1$ , update  $\beta_2$  using the negative binomial Gibbs sampler proposed by Pillow and Scott (2012)
4. Update  $r$  using either a random-walk Metropolis-Hastings step or the two-stage Gibbs sampler proposed by Zhou and Carin (2015)

Steps 1, 2 and 3 involve Gibbs updates that admit closed-form full conditionals, and Step 4 involves either a straightforward Metropolis-Hastings update or a Gibbs update.

### Step 1: Update the Latent At-Risk Indicators

In Step 1 of the sampler, we update the at-risk indicators,  $w_1, \dots, w_n$ . As outlined in the online [supplement](#) (Neelon, 2018), the full conditional for  $w_i$  is a discrete distribution with probabilities that depend on whether the observed count,  $y_i$ , is zero or non-zero. If  $y_i > 0$ , then subject  $i$  belongs to the at-risk class, and hence by definition,  $w_i = 1$  with probability 1. Conversely, if  $y_i = 0$ , then we observe either a structural zero (implying that  $w_i = 0$ ) or an at-risk zero (implying  $w_i = 1$ ). Here, we draw  $w_i$  from a Bernoulli distribution with probability

$$\begin{aligned} \theta_i &= \Pr(w_i = 1 | y_i = 0, \text{rest}) = \Pr(\text{at-risk zero} | \text{at risk or structural zero}) \\ &= \frac{\pi_i v_i^r}{1 - \pi_i(1 - v_i^r)}, \end{aligned} \quad (5)$$

where, from equation (2),  $\pi_i = \exp(\eta_{1i})/[1 + \exp(\eta_{1i})]$  is the unconditional probability that  $w_i = 1$ , and  $v_i = 1 - \psi_i$ , where  $\psi_i$  is the negative binomial event probability defined in equation (3). The result follows from a direct application of Bayes' Theorem. The proof is presented in Appendix A of the [supplement](#).

### Step 2: Update $\beta_1$

To implement Step 2, we employ the data-augmentation Gibbs sampler proposed by Polson et al. (2013a). The approach introduces a vector of latent variables that are scale mixtures of normals with independent Pólya-Gamma precision terms. A random variable  $\omega$  is said to have a Pólya-Gamma distribution with parameters  $b > 0$  and  $c \in \Re$ , if

$$\omega \sim \text{PG}(b, c) \stackrel{d}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)}, \quad (6)$$

where the  $g_k$ 's are independently distributed according to  $\text{Ga}(b, 1)$ .

Polson et al. (2013a) establish two important properties of the  $\text{PG}(b, c)$  density. First, for  $a \in \mathfrak{R}$  and  $\eta \in \mathfrak{R}$ , it follows that

$$\frac{(e^\eta)^a}{(1 + e^\eta)^b} = 2^{-b} e^{\kappa\eta} \int_0^\infty e^{-\omega\eta^2/2} p(\omega|b, 0) d\omega, \quad (7)$$

where  $\kappa = a - b/2$  and  $p(\omega|b, 0)$  denotes a  $\text{PG}(b, 0)$  density. Next, the conditional distribution  $p(\omega|b, c) \sim \text{PG}(b, c)$  arises from an “exponential tilting” of the  $\text{PG}(b, 0)$  density:

$$\begin{aligned} p(\omega|b, c) &= \frac{\exp(-c^2\omega/2)p(\omega|b, 0)}{\mathbb{E}_\omega[\exp(-c^2\omega/2)]} \\ &= \frac{\exp(-c^2\omega/2)p(\omega|b, 0)}{\int_0^\infty e^{-c^2\omega/2} p(\omega|b, 0) d\omega}. \end{aligned} \quad (8)$$

Under the logistic model in equation (2), the Bernoulli likelihood for the at-risk indicators  $\mathbf{w} = (w_1, \dots, w_n)^T$  is

$$\begin{aligned} p(\mathbf{w}|\boldsymbol{\beta}_1) &= \prod_{i=1}^n p(w_i|\boldsymbol{\beta}_1) \\ &= \prod_{i=1}^n \left( \frac{\exp(\eta_{1i})}{1 + \exp(\eta_{1i})} \right)^{w_i} \left( \frac{1}{1 + \exp(\eta_{1i})} \right)^{1-w_i} \\ &= \prod_{i=1}^n \frac{(e^{\eta_{1i}})^{w_i}}{1 + e^{\eta_{1i}}}, \end{aligned} \quad (9)$$

where  $\eta_{1i} = \mathbf{x}_i^T \boldsymbol{\beta}_1$ . The  $i$ -th element of the Bernoulli likelihood has the same form as the left-hand expression in equation (7), with  $a_i = w_i$  and  $b = 1$ . Thus, we can re-write the Bernoulli likelihood in terms of the Pólya-Gamma random variables  $\boldsymbol{\omega}_1 = (\omega_{11}, \dots, \omega_{1n})^T$  according to equation (7):

$$p(w_i|\boldsymbol{\beta}_1) \propto e^{\kappa_i \eta_{1i}} \int_0^\infty e^{-\omega_{1i} \eta_{1i}^2/2} p(\omega_{1i}|1, 0) d\omega_{1i}, \quad (10)$$

where  $\kappa_i = w_i - 1/2$ . Let  $\omega_{1i}$  ( $i = 1, \dots, n$ ) be independently distributed according to  $\text{PG}(1, \eta_{1i})$ . By appealing to the above properties the Pólya-Gamma distribution, Polson et al. (2013a) show that the full conditional distribution of  $\boldsymbol{\beta}_1$ , given  $\mathbf{w}$  and  $\boldsymbol{\omega}_1$ , is

$$p(\boldsymbol{\beta}_1|\mathbf{w}, \boldsymbol{\omega}_1) \propto \pi(\boldsymbol{\beta}_1) \exp \left[ -\frac{1}{2} (\mathbf{z}_1 - \mathbf{X}\boldsymbol{\beta}_1)^T \boldsymbol{\Omega}_1 (\mathbf{z}_1 - \mathbf{X}\boldsymbol{\beta}_1) \right], \quad (11)$$

where  $\pi(\boldsymbol{\beta}_1)$  is the prior distribution for  $\boldsymbol{\beta}_1$ ; for  $i = 1, \dots, n$ ,  $z_{1i} = \frac{w_i - 1/2}{\omega_{1i}}$  with  $\mathbf{z}_1 = (z_{11}, \dots, z_{1n})^T$ ;  $\boldsymbol{\Omega}_1 = \text{diag}(\boldsymbol{\omega}_1)$  is an  $n \times n$  precision matrix; and  $\mathbf{X}$  is an  $n \times p$  design matrix. It is clear that, given  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\Omega}_1$ ,  $\mathbf{z}_1$  is normally distributed with mean  $\boldsymbol{\eta}_1 =$

$\mathbf{X}\beta_1$  and diagonal covariance  $\Omega_1^{-1}$ . Thus, assuming a  $N_p(\beta_0, \Sigma_0)$  prior for  $\beta_1$ , the full conditional for  $\beta_1$  given  $\mathbf{z}_1$  and  $\Omega_1$  is  $N_p(\mu, \Sigma)$ , where

$$\begin{aligned}\Sigma &= \left( \Sigma_0^{-1} + \mathbf{X}^T \Omega_1 \mathbf{X} \right)^{-1}, \\ \mu &= \Sigma \left( \Sigma_0^{-1} \beta_0 + \mathbf{X}^T \Omega_1 \mathbf{z}_1 \right).\end{aligned}\quad (12)$$

The derivation can be found in Polson et al. (2013a); for convenience, we provide a summary in Appendix A of the online [supplement](#).

Given these results, the Gibbs sampler for Step 2 proceeds by selecting initial values for  $\beta_1$  and  $\mathbf{w}$  and iterating through the following steps:

1. For  $i = 1, \dots, n$ , update  $\omega_{1i}$  from a  $PG(1, \eta_{1i})$  density, where  $\eta_{1i} = \mathbf{x}_i^T \beta_1$
2. For  $i = 1, \dots, n$ , define  $z_{1i} = \frac{w_i - 1/2}{\omega_{1i}}$
3. Conditional on  $\mathbf{z}_1$ , update  $\beta_1$  from  $N_p(\mu, \Sigma)$ , where  $\mu$  and  $\Sigma$  are given in (12).

An efficient accept-reject algorithm is used to sample from the Pólya-Gamma distribution and can be implemented in the R package `BayesLogit` (Polson et al., 2013b).

### Step 3: Update $\beta_2$

The update for  $\beta_2$  is similar to the one for  $\beta_1$ . Adopting the parameterization of the negative binomial in equation (3), the conditional likelihood of  $y_i$  given  $w_i = 1$  is

$$\begin{aligned}p(y_i | r, \beta_2, w_i = 1) &\propto (1 - \psi_i)^r \psi_i^{y_i} \\ &= \frac{\exp(\eta_{2i})^{y_i}}{[1 + \exp(\eta_{2i})]^{r+y_i}},\end{aligned}\quad (13)$$

where  $\eta_{2i} = \mathbf{x}_i^T \beta_2$ . Exploiting property 1 of the Pólya-Gamma distribution in equation (7), it follows that

$$p(y_i | r, \beta_2, w_i = 1) \propto e^{\kappa_i \eta_{2i}} \int_0^\infty e^{-\omega_{2i} \eta_{2i}^2 / 2} p(\omega_{2i} | r + y_i, 0) d\omega_{2i}, \quad (14)$$

where  $\kappa_i = (y_i - r)/2$ . If we let  $\omega_{2i}$  be distributed according to  $PG(y_i + r, \eta_{2i})$ , then following Pillow and Scott (2012), the full conditional for  $\beta_2$  is

$$p(\beta_2 | \mathbf{y}^*, r, \mathbf{w}, \omega_2) \propto \pi(\beta_2) \exp \left[ -\frac{1}{2} (\mathbf{z}_2 - \mathbf{X}^* \beta_2)^T \Omega_2 (\mathbf{z}_2 - \mathbf{X}^* \beta_2) \right], \quad (15)$$

where  $\mathbf{y}^*$  is the  $n^* \times 1$  subvector of  $\mathbf{y}$  corresponding to  $w_i = 1$ ;  $n^* = \sum_{i=1}^n w_i$  is the number of individuals in the at-risk class (i.e., for whom  $w_i = 1$ );  $\omega_2$  is a vector of length  $n^*$  with elements  $\omega_{2i}$ ;  $\mathbf{z}_2$  is a vector of length  $n^*$  with elements  $z_{2i} = \frac{y_i - r}{2\omega_{2i}}$ ;  $\Omega_2 = \text{diag}(\omega_2)$  is an  $n^* \times n^*$  precision matrix; and  $\mathbf{X}^*$  is an  $n^* \times p$  design matrix. From

(15), it is clear that  $\mathbf{z}_2$  is normally distributed with mean  $\boldsymbol{\eta}_2 = \mathbf{X}^* \boldsymbol{\beta}_2$  and diagonal covariance  $\boldsymbol{\Omega}_2^{-1}$ . Thus, assuming a  $N_p(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$  prior for  $\boldsymbol{\beta}_2$ , the conjugate full conditional for  $\boldsymbol{\beta}_2$  given  $\mathbf{z}_2$  and  $\boldsymbol{\Omega}_2$  is  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , where

$$\begin{aligned}\boldsymbol{\Sigma} &= \left( \boldsymbol{\Sigma}_0^{-1} + \mathbf{X}^{*T} \boldsymbol{\Omega}_2 \mathbf{X}^* \right)^{-1}, \\ \boldsymbol{\mu} &= \boldsymbol{\Sigma} \left( \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\beta}_0 + \mathbf{X}^{*T} \boldsymbol{\Omega}_2 \mathbf{z}_2 \right).\end{aligned}\tag{16}$$

The proof can be found in Pillow and Scott (2012) and is summarized in the context of the ZINB model in Appendix A of the online [supplement](#). Thus, given current values for  $\boldsymbol{\beta}_2$ ,  $\mathbf{w}$ , and  $r$ , the Gibbs sampler for Step 3 proceeds as follows:

1. For  $w_i = 1$ , draw  $\omega_{2i}$  from its  $\text{PG}(y_i + r, \eta_{2i})$  distribution, where  $\eta_{2i} = \mathbf{x}_i^T \boldsymbol{\beta}_2$
2. For  $w_i = 1$ , define  $z_{2i} = \frac{y_i - r}{2\omega_{2i}}$
3. Update  $\boldsymbol{\beta}_2$  from its  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  distribution, where  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are given in (16).

#### Step 4: Update $r$

In the final step, we update  $r$  using either a Metropolis-Hastings step or a conjugate Gibbs update. For the Metropolis update, we select a uniform prior with positive support and draw candidate values of  $r$  from a zero-truncated normal proposal centered at the current value of  $r$ . Alternatively, one can adopt the two-stage Gibbs update proposed by Zhou and Carin (2015) and discussed more recently by Dadaneh et al. (2018). In stage 1, latent counts are introduced according to a Chinese restaurant table distribution; in stage 2,  $r$  is sampled from a conjugate Gamma distribution given the latent counts. Details are provided in the online [supplement](#). In our experience, the Metropolis-Hastings update works well in practice, and we therefore present Metropolis-based results in the sections below.

To complete the prior specification for the ZINB model, we assign weakly informative  $N_p(\mathbf{0}, 100\mathbf{I}_p)$  priors to  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$ . These choices work well for the analyses presented in Sections 3 and 4. More generally, we expect little sensitivity to prior specification, except perhaps in cases where there is an extremely high or low percentage of zeros (e.g.,  $> 95\%$  or  $< 5\%$ ). In the former case, there are relatively few nonzero values, resulting in a small at-risk sample; in the latter, there tend to be very few structural zeros, in which case a standard (non-inflated) negative binomial model provides adequate fit. However, these are instances in which maximum likelihood methods also break down. In general, the proposed sampling algorithm works well for scenarios commonly encountered in practice, as illustrated by the numerical examples presented in Section 3.

The MCMC algorithm cycles through Steps 1–4 until convergence, which can be assessed using standard Markov chain Monte Carlo (MCMC) diagnostics such as trace plots, Geweke  $z$ -statistics (Geweke, 1992) and Monte Carlo standard errors. These diagnostics can be obtained from the R packages `coda` (Plummer et al., 2006) and `mcmcse` (Flegal et al., 2017). In our experience, convergence for fixed effects models is almost immediate with excellent mixing. Even for more complex models, we typically observe rapid convergence, as illustrated by the simulated examples presented in Section 3.

### 2.3 Extensions to Longitudinal and Spatial Data

Equipped with the latent normal variables  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , the ZINB model can easily be extended to accommodate longitudinal, spatial, and time series data — essentially any setting where the model parameters are linear in  $\mathbf{z}_1$  and  $\mathbf{z}_2$ . Suppose, for example, we have count responses,  $y_{ij}$ , measured at occasions  $j = 1, \dots, n_i$  for individual  $i$ . We can model the data using a longitudinal version of the ZINB model, which is expressed as

$$y_{ij} \sim (1 - \pi_{ij}) \mathbb{1}_{(w_{ij}=0 \wedge y_{ij}=0)} + \pi_{ij} \text{NB}(\mu_{ij}, r) \mathbb{1}_{(w_{ij}=1)}, \quad (17)$$

where, for the  $ij$ -th observation,  $w_{ij}$  denotes the at-risk indicator taking value 1 with probability  $\pi_{ij}$ , and  $\mu_{ij}$  is the negative binomial mean analogous to the one presented in line 1 of equation (4). As before, we model  $\pi_{ij}$  using a logit link

$$\text{logit}(\pi_{ij}) = \text{logit}[\Pr(w_{ij} = 1 | \beta_1, \phi_{1i})] = \mathbf{x}_{ij}^T \beta_1 + \mathbf{v}_{ij}^T \phi_{1i} = \eta_{1ij}, \quad (18)$$

where  $\mathbf{x}_{ij}$  is a  $p \times 1$  vector of covariates including appropriate functions of time (e.g., linear or polynomial time trends);  $\beta_1$  is a  $p \times 1$  vector of fixed effect coefficients;  $\mathbf{v}_{ij}^T$  is a  $q \times 1$  random effect design vector that includes functions of time; and  $\phi_{1i} \sim N_q(\mathbf{0}, \mathbf{G}_1)$  is a  $q \times 1$  vector of random effects for the binary component, with  $q \times q$  prior covariance  $\mathbf{G}_1$ . Similarly, the negative binomial component (conditional on  $w_{ij} = 1$ ) is modeled as

$$\begin{aligned} p(y_{ij} | r, \beta_2, \phi_{2i}, w_{ij} = 1) &\stackrel{d}{=} \frac{\Gamma(y_{ij} + r)}{\Gamma(r) y_{ij}!} (1 - \psi_{ij})^r \psi_{ij}^{y_{ij}}, \quad \forall i, j \text{ s.t. } w_{ij} = 1, \\ \psi_{ij} &= \frac{\exp(\mathbf{x}_{ij}^T \beta_2 + \mathbf{v}_{ij}^T \phi_{2i})}{1 + \exp(\mathbf{x}_{ij}^T \beta_2 + \mathbf{v}_{ij}^T \phi_{2i})} = \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}, \end{aligned} \quad (19)$$

where  $\phi_{2i} \sim N_q(\mathbf{0}, \mathbf{G}_2)$  is a  $q \times 1$  vector of random effects for the count component with  $q \times q$  covariance  $\mathbf{G}_2$ . Often it is reasonable to retain the same random effect structure for both components, although in general this is not necessary. For example, we might include only a random intercept in the binary component but a random intercept and slope in the count component. Without loss of generality, we assume throughout that  $q$  is the same for both components; however, the proposed models can be easily modified to accommodate separate dimensions  $q_1$  and  $q_2$  for the two parts of the ZINB model.

To facilitate posterior computation, we again augment the data with latent normal variables for each component. Let  $z_{1ij} = (y_{ij} - 1/2)/\omega_{1ij}$  be the latent normal variable for the binary component of the ZINB at occasion  $ij$ , and let  $z_{2ij} = (y_{ij} - r)/(2\omega_{2ij})$  be the latent normal variable for the count component conditional on  $w_{ij} = 1$ , where  $\omega_{1ij}$  and  $\omega_{2ij}$  follow independent Pólya-Gamma distributions. We model the  $n_i \times 1$  vector  $\mathbf{z}_{1i} = (z_{1i1}, \dots, z_{1in_i})^T$  as

$$\mathbf{z}_{1i} | \boldsymbol{\eta}_{1i}, \boldsymbol{\Omega}_{1i} \sim N_{n_i}(\boldsymbol{\eta}_{1i}, \boldsymbol{\Omega}_{1i}^{-1}), \quad (20)$$

where, for subject  $i$ ,  $\boldsymbol{\eta}_{1i} = (\eta_{1i1}, \dots, \eta_{1in_i})^T = \mathbf{X}_i \beta_1 + \mathbf{V}_i \phi_{1i}$  is the  $n_i \times 1$  linear predictor for the binary component;  $\mathbf{X}_i$  and  $\mathbf{V}_i$  are, respectively,  $n_i \times p$  and  $n_i \times q$  fixed and random effect design matrices;  $\boldsymbol{\Omega}_{1i} = \text{diag}(\omega_{1i})$  is an  $n_i \times n_i$  diagonal precision matrix; and  $\omega_{1i} = (\omega_{1i1}, \dots, \omega_{1in_i})^T$ . Similarly, for the count component, we have

$$\mathbf{z}_{2i} | \boldsymbol{\eta}_{2i}, \boldsymbol{\Omega}_{2i} \sim N_{n_i^*}(\boldsymbol{\eta}_{2i}, \boldsymbol{\Omega}_{2i}^{-1}) \quad (21)$$



Here,  $\mathbf{z}_{2i}$  is a vector of length  $n_i^*$ , where  $n_i^* = \sum_{j=1}^{n_i} w_{ij}$  is the number of at-risk observations for subject  $i$ ;  $\boldsymbol{\eta}_{2i} = \mathbf{X}_i^* \boldsymbol{\beta}_2 + \mathbf{V}_i^* \boldsymbol{\phi}_{2i}$  is the  $n_i^* \times 1$  linear predictor for the count component, where  $\mathbf{X}_i^*$  and  $\mathbf{V}_i^*$  are  $n_i^* \times p$  and  $n_i^* \times q$  fixed and random effect design matrices;  $\boldsymbol{\Omega}_{2i} = \text{diag}(\boldsymbol{\omega}_{2i})$  is an  $n_i^* \times n_i^*$  precision matrix; and  $\boldsymbol{\omega}_{2i}$  is an  $n_i^* \times 1$  vector of PG precisions for the count component.

Note that if  $n_i^* = 0$ , then none of the observations fall into the at risk class – that is,  $w_{ij} = 0$  for all  $j = 1, \dots, n_i$ . This is unlikely to occur unless  $n_i$  is small or  $\phi_{ij}$  is low for  $j = 1, \dots, n_i$ , leading to few at-risk observations for subject  $i$ . When  $n_i^*$  is small, we must rely more heavily on the multivariate normal prior for  $\boldsymbol{\phi}_{2i}$  to shrink the random effects toward a global population mean of zero, thus stabilizing the random effect predictions. As we illustrate in Section 3, the proposed random effect ZINB model performs well even for small  $n_i^*$ .

In many cases, it is reasonable to assume that the binary and count components are correlated, thus allowing for dependence between the at-risk probability and the count distribution among those at risk. In our motivating application, for example, patients who are at high risk for inpatient hospitalizations may also have a greater number of re-admissions compared to those with low risk of inpatient hospitalizations. Recent work suggests that ignoring this association can lead to biased inferences in zero-inflated models (Su et al., 2009). We can accommodate this dependence by allowing  $\phi_{1i}$  and  $\phi_{2i}$  to be correlated according to a multivariate normal distribution. Let  $\boldsymbol{\phi}_i = (\phi_{1i}^T, \phi_{2i}^T)^T$  denote the  $2q \times 1$  vector comprising the random effects for both the binary and count components. We assume the following multivariate normal distribution for  $\boldsymbol{\phi}_i$ :

$$\begin{aligned} \boldsymbol{\phi}_i | \boldsymbol{\Gamma} &\sim \text{N}_{2q}(\mathbf{0}, \boldsymbol{\Gamma}), \\ \text{where } \boldsymbol{\Gamma} &= \begin{pmatrix} \mathbf{G}_1 & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_2 \end{pmatrix} \end{aligned} \quad (22)$$

is a  $2q \times 2q$  positive-definite covariance matrix under the default assumption that  $q$  is the same for the binary and count components. The covariance between the components is captured by the  $q \times q$  off-diagonal elements  $\mathbf{G}_{12} = \mathbf{G}_{21}^T$ .

The correlated random effects model is especially attractive because it allows the level of shrinkage imposed on the random effects to be correlated across components. By applying two related sources of shrinkage to the random effects, the correlated model improves inference in the presence of small  $n_i^*$ . In particular, when there are few at-risk observations, so that  $n_i^*$  is small, the correlated model allows the random effects  $\phi_{2i}$  in the count component to borrow information from  $\phi_{1i}$  in the binary component, which typically has a greater sample size (i.e.,  $n_i \geq n_i^*$ ) and hence more information available for prediction.

The model can further extend to areal spatial and spatiotemporal data by assigning a multivariate conditionally autoregressive (CAR) prior to  $\boldsymbol{\phi}_i$  in equation (22). For example, an intrinsic CAR prior (Banerjee et al., 2014) takes the form

$$\boldsymbol{\phi}_i | \boldsymbol{\phi}_{(-i)}, \boldsymbol{\Gamma} \sim \text{N}_{2q} \left( \frac{1}{m_i} \sum_{l \in \partial_i} \boldsymbol{\phi}_{il}, \frac{1}{m_i} \boldsymbol{\Gamma} \right), \quad (23)$$

where  $m_i$  is the number of neighbors for  $i$ -th areal unit,  $\partial_i$  is the set of neighbors for unit  $i$ , and  $\mathbf{\Gamma}$  is the  $2q \times 2q$  conditional covariance matrix given the remaining spatial random effects,  $\phi_{(-i)}$ . Whereas the multivariate normal prior in equation (22) permits “global” shrinkage to a population mean of zero, the CAR prior borrows information across neighboring spatial regions, resulting in “localized” shrinkage that yields a spatially smoothed map.

As the dimension of  $q$  increases, the joint posterior update for  $\phi_i$  can become unmanageable. Consider, for example, a spatiotemporal intercept and slope model, where each component includes a random intercept and linear time trend. Here, each component includes  $q = 2$  random effects, resulting in a  $4 \times 4$  covariance matrix  $\mathbf{\Gamma}$  in equation (23). Let  $\phi_{11} = (\phi_{111}, \dots, \phi_{1n1})^T$  and  $\phi_{12} = (\phi_{112}, \dots, \phi_{1n2})^T$  denote, respectively, the  $n \times 1$  vectors of random intercepts and slopes for the binary component. Similarly, define  $\phi_{21} = (\phi_{211}, \dots, \phi_{2n1})^T$  and  $\phi_{22} = (\phi_{212}, \dots, \phi_{2n2})^T$  to be the intercept and slope vectors for the count component. Finally, let  $\phi = (\phi_{11}^T, \dots, \phi_{22}^T)^T$ , be the  $4n \times 1$  collection of all random effects. Following Brook’s Lemma (Banerjee et al., 2014), the joint prior for  $\phi$  is given by

$$p(\phi|\mathbf{\Gamma}) \propto \exp \left[ -\frac{1}{2} \phi^T (\mathbf{\Gamma}^{-1} \otimes \mathbf{Q}) \phi \right], \quad (24)$$

where  $\mathbf{Q} = \mathbf{M} - \mathbf{A}$  is an  $n \times n$  “structure” matrix of rank  $n - 1$ ;  $\mathbf{M} = \text{diag}(m_1, \dots, m_n)$  with diagonal elements equal to the number of neighbors for each spatial unit; and  $\mathbf{A}$  is an  $n \times n$  adjacency matrix with  $a_{ii} = 0$ ,  $a_{il} = 1$  if spatial units  $i$  and  $l$  are neighbors, and  $a_{il} = 0$  otherwise. Because  $\mathbf{Q}$  is rank-deficient — and hence  $p(\phi|\mathbf{\Gamma})$  is improper — a sum-to-zero constraint is typically applied to  $\phi$  as part of the MCMC algorithm to ensure an identifiable model (Banerjee et al., 2014).

From expression (24), the joint prior for  $\phi$  is proportional to a multivariate normal density with  $2qn \times 2qn$  precision matrix  $\mathbf{\Gamma}^{-1} \otimes \mathbf{Q}$ , which in many applications is too unwieldy for efficient posterior inference. It is therefore convenient to partition expression (24) into univariate conditional priors for each vector  $\phi_{kk}$  ( $k = 1, \dots, q$ ) given the remaining random effects. This leads to efficient Gibbs sampling by permitting separate updates for each  $n \times 1$  vector  $\phi_{kk}$ . For instance, under the spatiotemporal intercept/slope model described above, the conditional prior for  $\phi_{11}$ , the  $n \times 1$  vector of random intercepts for the binary component, is

$$\begin{aligned} p(\phi_{11}|\phi_{12}, \phi_{21}, \phi_{22}, \mathbf{\Gamma}) &\propto \exp \left[ -\frac{1}{2} (\phi_{11} - \mu_{11})^T \Sigma_{11} (\phi_{11} - \mu_{11}) \right], \quad \text{where} \\ \Sigma_{11} &= \left[ \mathbf{\Gamma}_{11} - \mathbf{\Gamma}_{(1,-1)} \mathbf{\Gamma}_{(-1,-1)}^{-1} \mathbf{\Gamma}_{(-1,1)} \right]^{-1} \mathbf{Q}, \\ \mu_{11} &= \left[ \left( \mathbf{\Gamma}_{(1,-1)} \mathbf{\Gamma}_{(-1,-1)}^{-1} \right) \otimes \mathbf{I}_n \right] \phi_{(-1)}, \end{aligned} \quad (25)$$

where  $\mathbf{\Gamma}$  is the  $4 \times 4$  covariance of  $\phi_i$ ,  $\mathbf{\Gamma}_{11}$  denotes the first element of  $\mathbf{\Gamma}$ ,  $\mathbf{\Gamma}_{(1,-1)}$  is the  $1 \times 3$  vector comprising the first row of  $\mathbf{\Gamma}$  with element 1 removed,  $\mathbf{\Gamma}_{(-1,-1)}$  is the  $3 \times 3$  submatrix of  $\mathbf{\Gamma}$  after removing row 1 and column 1,  $\mathbf{\Gamma}_{(-1,1)}$  is the  $3 \times 1$  vector comprising the first column of  $\mathbf{\Gamma}$  with element 1 removed, and  $\phi_{(-1)} = (\phi_{12}^T, \phi_{21}^T, \phi_{22}^T)^T$  is a  $3n \times 1$  vector of the remaining random effects. Equation (25) follows directly from conditional multivariate normal theory. Similar expressions hold for  $\phi_{12}$ ,  $\phi_{21}$  and  $\phi_{22}$ .

The conditional prior specification in (25) leads to efficient Gibbs updates for the spatial effects. Consider once again the spatial intercept/slope model. As detailed in the Appendix B of the online [supplement](#), the updates for  $\phi_{11}$  and  $\phi_{12}$  in the binary component depend on the likelihood contributions from all  $N = \sum_{i=1}^n n_i$  observations, whereas the updates for  $\phi_{21}$  and  $\phi_{22}$  in the count component rely only on contributions from the  $N^* = \sum_{i=1}^n n_i^* \leq N$  “at-risk” observations for which  $w_{ij} = 1$  ( $i = 1, \dots, n; j = 1, \dots, n_i$ ). This sample-size imbalance prevents a joint Gibbs update for  $\phi = (\phi_{11}^T, \dots, \phi_{22}^T)^T$  based on prior (24). The conditional prior (25) avoids this problem by providing separate univariate updates for  $\phi_{11}$ ,  $\phi_{12}$ ,  $\phi_{21}$  and  $\phi_{22}$ , the first two based on all  $N$  observations and the latter two based on the  $N^*$  at-risk observations. The approach can easily be generalized to  $q > 2$  random effects for each component, as well as to non-spatial longitudinal data, where  $\mathbf{Q} = \mathbf{I}_n$  is of full rank. Additional details on the conditional prior specification can be found in Appendix B of the [supplement](#).

Prior specification for the spatial and non-spatial correlated ZINB models is completed by assigning conditionally conjugate  $N_p(\beta_0, \Sigma_0)$  priors to the fixed effects and an inverse-Wishart( $2q, \Lambda$ ) prior to  $\Gamma$ , where  $\Lambda$  is  $2q \times 2q$  scale matrix. By default, we set  $\beta_0 = \mathbf{0}$ ,  $\Sigma_0 = 100\mathbf{I}_p$ , and  $\Lambda = \mathbf{I}_{2q}$ . To implement the MCMC for random effect ZINB models, we initialize the model parameters and then cycle through the following steps:

1. For all  $i, j$ , update the latent at-risk indicators,  $w_{ij}$ , according to the discrete probability distribution

$$\Pr(w_{ij} = 1 | \text{rest}) = \begin{cases} 1, & \text{if } y_{ij} = 1 \\ \frac{\pi_{ij} v_{ij}^r}{1 - \pi_{ij}(1 - v_{ij}^r)}, & \text{if } y_{ij} = 0 \end{cases}, \quad (26)$$

where  $\pi_{ij} = \exp(\eta_{1ij}) / [1 + \exp(\eta_{1ij})]$  is the unconditional probability that  $w_{ij} = 1$  given in equation (18),  $v_{ij} = 1 - \psi_{ij}$ , and  $\psi_{ij} = \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}$  is the negative binomial event probability defined in equation (19).

2. Update the parameters for the binary component:
  - (a) For all  $i, j$ , sample  $\omega_{1ij}$  from  $\text{PG}(1, \eta_{1ij})$ , where  $\eta_{1ij}$  is defined in equation (18)
  - (b) For all  $i, j$ , define  $z_{1ij} = (y_{ij} - 1/2) / \omega_{1ij}$
  - (c) Update  $\beta_1$  from its normal full conditional
  - (d) For  $k = 1, \dots, q$ , update each  $n \times 1$  vector  $\phi_{1k}$  from its normal full conditional based on the conditional prior specification given in (25); apply sum-to-zero constraints as needed
3. Update the parameters for the count component:
  - (a) For  $w_{ij} = 1$ , sample  $\omega_{2ij}$  from  $\text{PG}(1, \eta_{2ij})$ , where  $\eta_{2ij}$  is defined in equation (19)

- (b) For  $w_{ij} = 1$ , define  $z_{2ij} = (y_{ij} - r)/(2\omega_{2ij})$
  - (c) Update  $\beta_2$  from its normal full conditional
  - (d) For  $k = 1, \dots, q$ , update each  $n \times 1$  vector  $\phi_{2k}$  from its normal full conditional based on the conditional prior specification given in (25); apply sum-to-zero constraints as needed
  - (e) Update  $r$  using a random-walk Metropolis-Hastings step or the two-stage Gibbs sampler analogous to the ones outlined in Section 2.2
4. Update the  $2q \times 2q$  covariance matrix  $\Gamma$  from its conjugate inverse-Wishart full conditional.

Appendix B of the [supplement](#) derives the full conditionals for the spatial intercept/slope ZINB model implemented in Sections 3.3 and 4.

### 3 Simulated Examples

#### 3.1 Simulation 1: Fixed Effects ZINB Model

To illustrate the properties of the model, we conducted a series of simulations of increasing model complexity. First, we generated data from fixed effects ZINB model and compared the results to maximum likelihood estimates (MLEs) obtained using the SAS<sup>®</sup> software procedure NLMIXED (SAS Institute, Cary, North Carolina). The aim was to determine whether the proposed Bayesian approach with weakly informative priors yielded regression estimates and uncertainty intervals similar to those obtained under a classical, frequentist approach. To do so, we generated 1000 observations according to the following ZINB model:

$$\begin{aligned}
 y_i &\sim (1 - \pi_i)\mathbb{1}_{(w_i=0 \wedge y_i=0)} + \pi_i\text{NB}(\mu_i, r)\mathbb{1}_{(w_i=1)}, \\
 \text{logit}(\pi_i) &= \text{logit}[\Pr(w_i = 1|\beta_1)] = \eta_{1i} = \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \beta_{13}x_{i3}, \\
 p(y_i|r, \beta_2, w_i = 1) &\stackrel{d}{=} \frac{\Gamma(y_i + r)}{\Gamma(r)y_i!}(1 - \psi_i)^r \psi_i^{y_i}, \\
 \psi_i &= \frac{\exp(\eta_{2i})}{1 + \exp(\eta_{2i})}, \\
 \eta_{2i} &= \beta_{20} + \beta_{21}x_{i1} + \beta_{22}x_{i2} + \beta_{23}x_{i3},
 \end{aligned} \tag{27}$$

where  $x_{i1}$  was simulated from an  $N(0, 1)$  distribution,  $x_{i2}$  was simulated from a Bernoulli(0.5) distribution,  $x_{i3}$  was simulated from a discrete uniform distribution taking values  $\{0, 1, 2\}$ ,  $\beta_1 = (\beta_{10}, \dots, \beta_{13})^T = (0.5, -0.5, -0.25, 0.25)^T$ ,  $\beta_2 = (\beta_{20}, \dots, \beta_{23})^T = (0.5, -1, 0.75, -0.25)^T$  and  $r = 1$ . These values resulted in 60% zeros, a mean count of 2.8, and the five-number summary (0, 0, 0, 2, 68). Figure S1 in Appendix C of the online [supplement](#) presents a full histogram of the count distribution.

We assigned independent  $N(0, 100)$  priors to the regression coefficients and a  $\text{Unif}(0, 10)$  prior to  $r$ . To update  $r$ , we used a zero-truncated normal proposal with

variance 0.025 centered at the current value, resulting in a Metropolis-Hastings acceptance rate of 36%. Initial values were set at  $\beta_1 = \beta_2 = \mathbf{0}$ , and  $r = 1$ . We ran 50,500 iterations of the MCMC algorithm described in Section 2, discarding the first 500 as burn-in. Figure S2 of the [supplement](#) presents trace plots, Monte Carlo standard errors and p-values from the Geweke diagnostics for selected parameters. Non-significant p-values are indicative of convergence. The p-values for simulation 1 ranged from 0.10 to 0.47, indicating reasonable convergence. The trace plots showed satisfactory mixing. The algorithm took 6.20 minutes to run on a Dell® Precision T3610 workstation, compared to 0.50 seconds for SAS. However, a shorter run of 1500 iterations took approximately 11 seconds to run and produced similar results (Figure S3), indicating that the proposed Bayesian method is comparable to SAS in terms of run time.

Watanabe (Watanabe, 2010) Information Criteria (WAIC) values for the ZINB and negative binomial models were 3011 and 3067, respectively, indicating superior fit for the ZINB model. We based our model comparisons on WAIC rather than the more commonly used Deviance Information Criteria (DIC) because the DIC penalty term can yield negative values for mixture models, such as the ZINB, when the posterior mean deviates from the posterior mode (Celeux et al., 2006; Gelman et al., 2014). For a detailed discussion of WAIC and its comparison to other information criteria, please see Gelman et al. (2014).

Table 1 presents the parameter estimates and 95% intervals for the ZINB model under Bayesian and maximum likelihood estimation. In all cases, the Bayesian estimates were as or more accurate than the maximum likelihood estimates (MLEs). For both methods, the 95% intervals encompassed the simulated values. These results suggest that even for moderate sample sizes with a large percentage of zeros, the proposed Bayesian approach provides a suitable alternative to frequentist estimation for fixed effects ZINB models. The Bayesian approach might prove particularly attractive when prior data can be incorporated into the analysis to improve inferences, as frequentist approaches do not accommodate such information.

Model Component	Parameter	Simulated Value	Estimate (95% Interval)	
			Proposed Model <sup>†</sup>	MLE <sup>‡</sup>
Binary	$\beta_{10}$	0.50	0.51 (−0.00, 1.18)	0.39 (−0.12, 0.90)
	$\beta_{11}$	0.50	0.47 (−0.18, 0.85)	0.41 (−0.12, 0.70)
	$\beta_{12}$	−0.25	−0.03 (−0.46, 0.38)	−0.01 (−0.41, 0.38)
	$\beta_{13}$	0.25	0.09 (−0.15, 0.35)	0.09 (−0.15, 0.32)
Count	$\beta_{20}$	0.50	0.37 (−0.04, 0.72)	0.33 (−0.01, 0.67)
	$\beta_{21}$	−1.00	−0.95 (−1.09, −0.82)	−0.95 (−1.08, −0.81)
	$\beta_{22}$	0.75	0.62 (−0.40, 0.84)	0.62 (−0.40, 0.83)
	$\beta_{23}$	−0.25	−0.11 (−0.24, 0.02)	−0.11 (−0.24, 0.02)
	$r$	1.00	1.18 (−0.77, 1.68)	1.25 (−0.78, 1.73)

Table 1: Parameter estimates and 95% intervals for fixed effects ZINB model in simulation study 1. <sup>†</sup>Posterior means and 95% credible intervals for proposed Bayesian ZINB model. <sup>‡</sup>MLEs and 95% confidence intervals obtained using SAS Proc NLMIXED.

### 3.2 Simulation 2: Correlated Random Intercept ZINB Model

For the second simulation study, we generated data for 1000 subjects from the following correlated random intercept model analogous to the one given in equation (17):

$$\begin{aligned}
y_{ij} &\sim (1 - \pi_{ij})\mathbb{1}_{(w_{ij}=0 \wedge y_{ij}=0)} + \pi_{ij}\text{NB}(\mu_{ij}, r)\mathbb{1}_{(w_{ij}=1)}, \\
\text{logit}(\pi_{ij}) &= \text{logit}[\text{Pr}(w_{ij} = 1 | \beta_1, \phi_{1i})] \\
&= \eta_{1ij} = \beta_{10} + \beta_{11}x_i + \beta_{12}t_{ij} + \phi_{1i}, \\
p(y_{ij} | r, \beta_2, \phi_{2i}, w_{ij} = 1) &\stackrel{d}{=} \frac{\Gamma(y_{ij} + r)}{\Gamma(r)y_{ij}!} (1 - \psi_{ij})^r \psi_{ij}^{y_{ij}}, \\
\psi_{ij} &= \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}, \\
\eta_{2ij} &= \beta_{20} + \beta_{21}x_i + \beta_{22}t_{ij} + \phi_{2i}, \\
\phi_i &= (\phi_{1i}, \phi_{2i})^T \sim N_2(\mathbf{0}, \mathbf{\Gamma}),
\end{aligned} \tag{28}$$

where, for  $i = 1, \dots, 1000$  and  $j = 1, \dots, n_i$ ,  $y_{ij}$  denotes the count for individual  $i$  at occasion  $j$ ;  $w_{ij}$  is the “at-risk” indicator for observation  $ij$ ;  $\pi_{ij}$  is the corresponding at-risk probability;  $x_i \sim \text{Bern}(0.5)$  is a time-invariant binary covariate (e.g., gender);  $t_{ij} \sim N(0, 2)$  denotes the timing of observation  $ij$  (after centering, say); and  $\phi_i$  is a bivariate normal vector of random intercepts for the  $i$ -th individual, with mean zero and covariance  $\mathbf{\Gamma}$ . As with simulation 1, the goal was to compare the Bayesian estimates under weakly informative priors to the corresponding MLEs. Maximum likelihood was implemented using SAS Proc NLMIXED, which combines Gaussian quadrature for numerical integration with Newton-Raphson for maximization.

For simulation 2, we generated  $n_i$  according to a discrete uniform distribution ranging from 1 to 10, resulting in a total sample size of  $N = 5585$  with a mean of 5.59 observations per subject. Seven percent of the subjects had no at-risk observations ( $n_i^* = 0$ ), and another 17% had only one at-risk observation. Thus, we were able to evaluate the performance of our model when approximately one quarter of the sample had few (or no) at-risk observations. We assigned the following values to the model parameters:  $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})^T = (0.25, -0.25, 0.25)^T$ ,  $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})^T = (0.50, -0.25, 0.25)^T$ ,  $r = 1.25$ , and  $\mathbf{\Gamma} = \begin{pmatrix} 0.50 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ . These values resulted in 52% zeros, a mean count of 3.34, and a five-number summary of (0, 0, 0, 4, 220). Figure S4 of the online [supplement](#) presents a full histogram of the counts for simulation 2.

As in simulation 1, we assigned independent  $N(0, 100)$  priors to the fixed effects and a  $\text{Unif}(0, 10)$  prior to  $r$ . We reparameterized the bivariate normal prior for  $\phi_i$  using the conditional specification described in equation (25), leading to a conditional prior for  $\phi_{1i}$  of the form

$$\begin{aligned}
\phi_{1i} | \phi_{2i} &\sim N(m, v), \\
v &= \mathbf{\Gamma}_{11} - \frac{\mathbf{\Gamma}_{12}\mathbf{\Gamma}_{12}}{\mathbf{\Gamma}_{22}} = (1 - \rho^2)\mathbf{\Gamma}_{11}, \\
m &= \frac{\mathbf{\Gamma}_{21}}{\mathbf{\Gamma}_{22}}\phi_{2i} = \rho\sqrt{\frac{\mathbf{\Gamma}_{11}}{\mathbf{\Gamma}_{22}}}\phi_{2i},
\end{aligned} \tag{29}$$

where  $\mathbf{\Gamma}_{11}$  and  $\mathbf{\Gamma}_{22}$  are the marginal variances of  $\phi_{1i}$  and  $\phi_{2i}$ , respectively, and  $\rho = \frac{\mathbf{\Gamma}_{12}}{\sqrt{\mathbf{\Gamma}_{11}\mathbf{\Gamma}_{22}}} = \text{Corr}(\phi_{1i}, \phi_{2i})$ . The conditional prior for  $\phi_{2i}$  follows a similar expression. Finally, we assigned an inverse-Wishart(2,  $\mathbf{I}_2$ ) prior to  $\mathbf{\Gamma}$ .

For posterior inference, we implemented the MCMC algorithm described at the end of Section 2. Starting values for  $\beta_1$ ,  $\beta_2$  and  $r$  were identical to those in simulation 1. Initial values for  $\phi_{1i}$  and  $\phi_{2i}$  were drawn from independent standard normal distributions, and  $\mathbf{\Gamma}$  was initialized to  $\mathbf{I}_2$ . To update  $r$ , we used a zero-truncated normal proposal with variance 0.003, resulting in an acceptance rate of 42%. We ran the sampler for 50,500 iterations, discarding the first 500 as burn-in. Figure S5 of the [supplement](#) presents trace plots, Geweke diagnostic p-values and Monte Carlo standard errors for selected parameters. The results indicate excellent mixing. For comparison, we re-ran the algorithm for 2500 iterations (Figure S6 and Table S1), which took 78 seconds to run, compared to 55 seconds for SAS. An even shorter run of 1500 iterations yielded similar results and took only 44 seconds to complete, confirming that the proposed model is competitive with SAS in terms of computation time.

Table 2 presents the parameter estimates and 95% intervals under Bayesian and maximum likelihood estimation. The estimates for the two procedures were nearly identical, with 95% intervals encompassing the true parameter values. These results suggest that the proposed Bayesian approach performs similarly to maximum likelihood for commonly used mixed models with relatively few observations per subject, thus offering an appropriate Bayesian alternative to frequentist estimation in such cases.

Model		Simulated	Estimate (95% Interval)	
Component	Parameter	Value	Proposed Model <sup>†</sup>	MLE <sup>‡</sup>
Binary	$\beta_{10}$	0.25	0.06 (−0.17, 0.30)	0.05 (−0.19, 0.28)
	$\beta_{11}$	−0.25	−0.27 (−0.47, −0.08)	−0.27 (−0.47, −0.08)
	$\beta_{12}$	0.25	0.29 ( 0.21, 0.38)	0.29 ( 0.20, 0.38)
Count	$\beta_{20}$	0.50	0.39 ( 0.20, 0.58)	0.39 ( 0.20, 0.58)
	$\beta_{21}$	−0.25	−0.11 (−0.27, 0.03)	−0.12 (−0.27, 0.03)
	$\beta_{22}$	0.25	0.26 ( 0.21, 0.31)	0.26 ( 0.21, 0.30)
	$r$	1.25	1.32 ( 1.17, 1.54)	1.35 ( 1.19, 1.53)
Random	$\mathbf{\Gamma}_{11} = \text{Var}(\phi_{1i})$	0.50	0.54 ( 0.33, 0.79)	0.52 ( 0.28, 0.76)
Effects	$\mathbf{\Gamma}_{22} = \text{Var}(\phi_{2i})$	0.75	0.77 ( 0.64, 0.90)	0.76 ( 0.63, 0.89)
	$\mathbf{\Gamma}_{12} = \text{Cov}(\phi_{1i}, \phi_{2i})$	0.25	0.30 ( 0.18, 0.41)	0.31 ( 0.19, 0.42)

Table 2: Parameter estimates and 95% intervals for random effects ZINB model in simulation study 2. <sup>†</sup>Posterior means and 95% credible intervals for proposed Bayesian ZINB model. <sup>‡</sup>MLEs and 95% confidence intervals obtained using SAS Proc NLMIXED.

### 3.3 Simulation 3: Spatiotemporal ZINB Model

For the final simulation, we generated data from a spatiotemporal intercept/slope model analogous to the one described in Section 2. To emulate the spatial layout of our application, we used the US Census county-level adjacency matrix for South Carolina, Georgia, and Alabama U.S. Census Bureau (2014). This matrix contains  $n = 272$  counties and

1528 pairwise adjacencies. We simulated 50 observations per county over five years — the study time frame for our application — for a total of  $N = 50 \times 272 \times 5 = 68,000$  observations. We simulated the data from the following spatiotemporal ZINB model:

$$\begin{aligned}
y_{ij} &\sim (1 - \pi_{ij})\mathbb{1}_{(w_{ij}=0 \wedge y_{ij}=0)} + \pi_{ij}\text{NB}(\mu_{ij}, r)\mathbb{1}_{(w_{ij}=1)}, \\
\text{logit}(\pi_{ij}) &= \text{logit}[\Pr(w_{ij} = 1 | \beta_1, \phi_{1i})] \\
&= \eta_{1ij} = \beta_{10} + \beta_{11}t_{ij} + \phi_{1i1} + \phi_{1i2}t_{ij}, \\
p(y_{ij} | r, \beta_2, \phi_{2i}, w_{ij} = 1) &\stackrel{d}{=} \frac{\Gamma(y_{ij} + r)}{\Gamma(r)y_{ij}!} (1 - \psi_{ij})^r \psi_{ij}^{y_{ij}}, \\
\psi_{ij} &= \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}, \\
\eta_{2ij} &= \beta_{20} + \beta_{21}t_{ij} + \phi_{2i1} + \phi_{2i2}t_{ij}, \\
\phi_i | \phi_{(-i)}, \mathbf{\Gamma} &\sim N_4 \left( \frac{1}{m_i} \sum_{l \in \partial_i} \phi_{il}, \frac{1}{m_i} \mathbf{\Gamma} \right)
\end{aligned} \tag{30}$$

where  $t_{ij} \in \{0, 1, 2, 3, 4\}$  denotes study year, with 0 as baseline;  $\phi_{1i} = (\phi_{1i1}, \phi_{1i2})^T$  is a vector comprising the  $i$ -th random intercept ( $\phi_{1i1}$ ) and slope ( $\phi_{1i2}$ ) for the binary component;  $\phi_{2i} = (\phi_{2i1}, \phi_{2i2})^T$  is the corresponding vector of random effects for the count component;  $\phi_i = (\phi_{1i1}, \phi_{1i2}, \phi_{2i1}, \phi_{2i2})^T$  is modeled as multivariate ICAR distribution with conditional covariance  $\mathbf{\Gamma}$ ; and  $m_i$  denotes the number of counties adjacent to county  $i$ . The true parameter values are given in Table 3. These values resulted in a count distribution containing 70% zeros with a five number summary of (0, 0, 0, 1, 158). Figure S7 presents a full histogram of the counts. Unlike the models in simulations 1 and 2, the correlated spatial model (30) cannot be readily fit using existing frequentist or Bayesian software. In WinBUGS, for example, we immediately encountered unavoidable “trap” errors when fitting the model. This may be due to the fact that the ZINB model relies on the so-called “zeros trick” for implementation in WinBUGS, which may contribute to numerical instability. Thus, the proposed approach offers a convenient method for fitting complex ZINB models that cannot be easily accommodated by other means.

As before, we assumed independent  $N(0, 100)$  priors for the fixed effects and a  $\text{Unif}(0, 10)$  prior for  $r$ . Following equation (25), we partitioned the multivariate intrinsic CAR prior for  $\phi_i$  into separate univariate priors. We completed the prior specification by assigning an inverse-Wishart( $4, \mathbf{I}_4$ ) prior to  $\mathbf{\Gamma}$ . Starting values for  $\beta_1$ ,  $\beta_2$  and  $r$  were identical to those in simulations 1 and 2. Initial values for the random effects were drawn from independent standard normal distributions, and  $\mathbf{\Gamma}$  was initialized to  $\mathbf{I}_4$ . To update  $r$ , we used a zero-truncated normal proposal with variance 0.0002, resulting in an acceptance rate of 43%. To improve efficiency of the algorithm, we used the R package `spam` (Furrer and Sain, 2010; Gerber and Furrer, 2015) to convert the CAR structure matrix  $\mathbf{Q}$  in equation (24) to a sparse matrix object. This avoids computationally intensive matrix operations designed for dense matrices. We ran the sampler for 50,500 iterations with a burn-in of 500. Trace plots, Geweke diagnostics and Monte Carlo standard errors were indicative of convergence and showed reasonable mixing for a range of model parameters (Figure S8). For comparison, we re-ran the analysis for



2,500 iterations (Figure S9 and Table S2), as well as for 1500 iterations with a run time of 9 minutes. We obtained similar results for all scenarios.

Table 3 presents the posterior means and 95% credible intervals (CrIs) for the model parameters. Overall, the estimates were close to the true value with 95% intervals that overlapped the simulated value. Figure 1 presents maps of the true values and posterior mean predictions for  $\phi_{1i1}$  and  $\phi_{1i2}$  ( $i = 1, \dots, n$ ), the random intercepts and slopes for the binary component of the spatiotemporal ZINB model. In each case, the spatial pattern for the estimated effects closely mirrored the true spatial distribution, suggesting the proposed model accurately recovered the underlying spatial pattern in the data. Figure 2 presents the corresponding maps for the count component. Again, the predicted spatial pattern is similar to the true pattern, but with increased smoothing. The additional smoothing is not surprising, as the count component conditions on the at-risk class, and therefore has fewer observations available for spatial prediction than the binomial component. As a result, it relies more heavily on the CAR smoothing prior for prediction.

Model Component	Parameter	Simulated Value	Posterior Mean (95% CrI)
Binary	$\beta_{10}$	-0.25	-0.27 ( -0.33, -0.19)
	$\beta_{11}$	0.25	0.22 ( 0.19, 0.26)
Count	$\beta_{20}$	0.50	0.49 ( 0.43, 0.54)
	$\beta_{21}$	-0.25	-0.24 ( -0.26, -0.22)
	$r$	1.00	1.03 ( 0.96, 1.09)
Random Effects	$\Gamma_{11} = \text{Var}(\phi_{1i1})$	0.50	0.61 ( 0.42, 0.84)
	$\Gamma_{12} = \text{Cov}(\phi_{1i1}, \phi_{1i2})$	0.10	0.03 ( -0.05, 0.11)
	$\Gamma_{13} = \text{Cov}(\phi_{1i1}, \phi_{2i1})$	0.10	0.09 ( -0.02, 0.20)
	$\Gamma_{14} = \text{Cov}(\phi_{1i1}, \phi_{2i2})$	-0.10	-0.07 ( -0.13, -0.01)
	$\Gamma_{22} = \text{Var}(\phi_{1i2})$	0.15	0.21 ( 0.14, 0.29)
	$\Gamma_{23} = \text{Cov}(\phi_{1i2}, \phi_{2i1})$	0.10	0.12 ( 0.05, 0.20)
	$\Gamma_{24} = \text{Cov}(\phi_{1i2}, \phi_{2i2})$	0.10	0.08 ( 0.04, 0.12)
	$\Gamma_{33} = \text{Var}(\phi_{2i1})$	0.50	0.55 ( 0.42, 0.71)
	$\Gamma_{34} = \text{Cov}(\phi_{2i1}, \phi_{2i2})$	0.10	0.06 ( 0.01, 0.11)
	$\Gamma_{44} = \text{Var}(\phi_{2i2})$	0.15	0.17 ( 0.13, 0.22)

Table 3: Parameter estimates and 95% credible intervals (CrIs) for the spatiotemporal ZINB model in simulation study 3.

As a final comparison, we fitted two additional models. First, we fit a “partially correlated” model that assumed independence between the binary and count components — that is, we assumed  $\mathbf{\Gamma}$  in equation (22) to be block diagonal with  $\mathbf{G}_{12} = \mathbf{G}_{21} = \mathbf{0}$ . For this model, we assigned independent inverse-Wishart( $2, \mathbf{I}_2$ ) priors to  $\mathbf{G}_1$  and  $\mathbf{G}_2$ . Second, we fit an uncorrelated model that assumed no correlation among any of the random effects — i.e., we assumed  $\mathbf{\Gamma}$  to be strictly diagonal. Here we assigned independent inverse-Gamma(0.001, 0.001) priors to the random effect variances. The WAIC values were 143,456 for the fully correlated model, 143,480 for the partially correlated model, and 143,653 for the uncorrelated model, indicating superior predictive accuracy for the fully correlated model. Because WAIC relies on a conditionally independent partitioning of the data, which can be problematic for spatially correlated data (Gelman et al.,

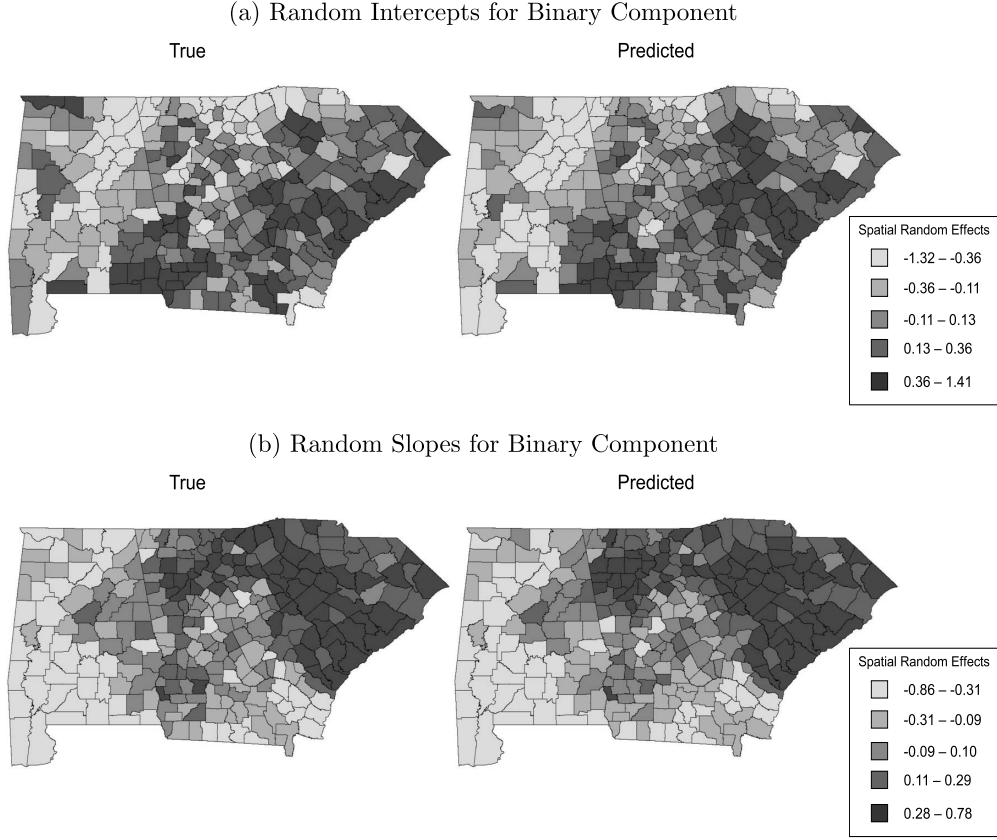


Figure 1: Simulated and predicted (a) random intercepts and (b) random slopes for the binary component of the ZINB model in simulation study 3. Top Left: Simulated random intercepts. Top Right: Predicted random intercepts. Bottom Left: Simulated random slopes. Bottom Right: Predicted random slopes.

2014), we additionally compared models based the root mean square predictive errors (RMSPEs) for each random effect vector,  $\phi_{11}$ ,  $\phi_{12}$ ,  $\phi_{21}$ ,  $\phi_{22}$ , using the expression

$$\text{RMSPE}_{kk} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\phi_{kik} - \hat{\phi}_{kik})^2}, \quad k = 1, 2, \quad (31)$$

where  $\phi_{kik}$  denotes the simulated value of random effect  $kk$  for subject  $i$ , and  $\hat{\phi}_{kik}$  is the corresponding predicted value. Table S3 in the [supplement](#) presents the posterior mean RMSPEs for each random effect under the three models. In all cases, the fully correlated model produced the smallest RMSPEs, particularly in contrast to the uncorrelated model. In addition, the parameter estimates for the uncorrelated model showed extreme bias (Table S4). These results suggest that ignoring even modest association among the

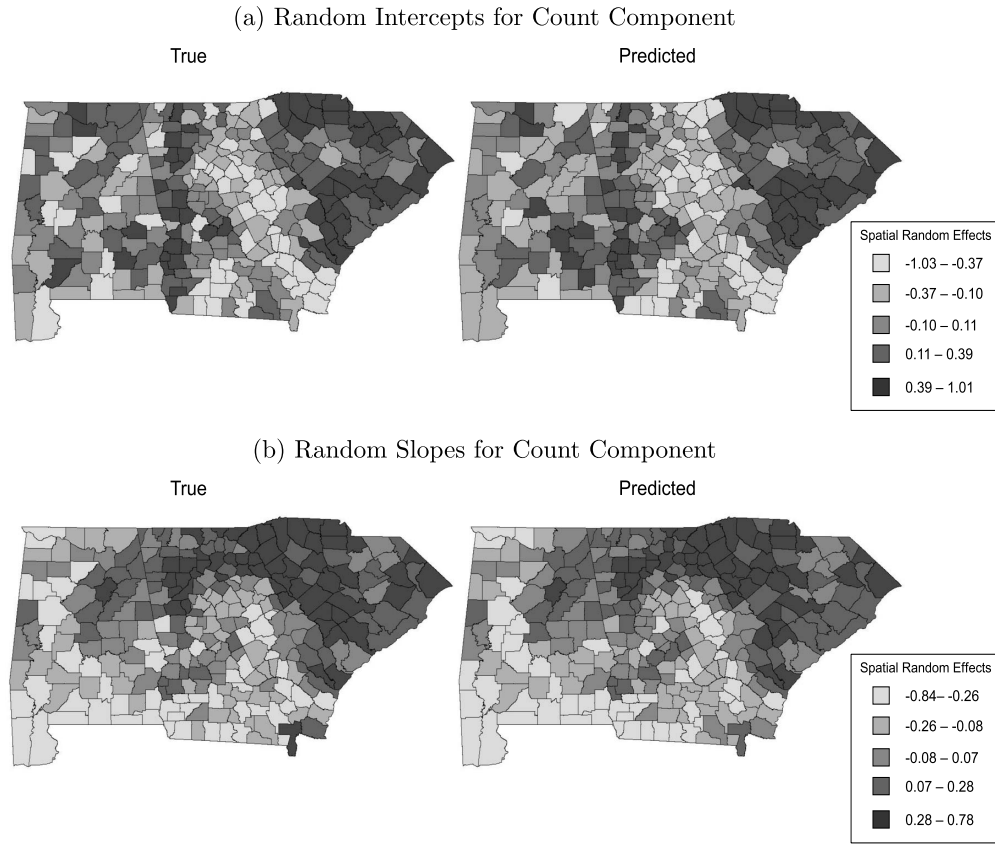


Figure 2: Simulated and predicted (a) random intercepts and (b) random slopes for the count component of the ZINB model in simulation study 3. Top Left: Simulated random intercepts. Top Right: Predicted random intercepts. Bottom Left: Simulated random slopes. Bottom Right: Predicted random slopes.

random effects can result in diminished predictive performance and imprecise estimates. These model comparison methods could also be used to test for other types of model misspecification, such as the choice of link function for the binary component. Taken together, the three simulations suggest that the proposed approach is comparable to maximum likelihood for relatively simple models, but can accommodate more complex scenarios where existing methods are impractical.

## 4 Analysis of Inpatient Admissions

We applied the spatiotemporal ZINB model to an analysis of inpatient hospital stays among 23,533 diabetic veterans residing in Alabama, Georgia, and South Carolina from 2011 to 2015. We modeled the annual number of inpatient admissions using a model

analogous to equation (30). Covariates included patient age (centered), sex, and race (non-Hispanic white vs other); Elixhauser score (Quan et al., 2005), a measure of comorbidity burden; and an indicator for service connected disability medical coverage, with 1 implying full medical coverage and 0 implying partial coverage. The sample comprised 68% zeros, an average of 1.24 admissions annually, and a five-number summary of (0, 0, 0, 1, 235). Figure S10 in Appendix D of the [supplement](#) provides a histogram of the counts, and Table S5 presents sample summary statistics. Prior distributions and MCMC specifications were identical to those for simulation 3. We ran the algorithm for a 50,000 iterations with a conservative burn-in of 25,000. We additionally thinned the chain by 25 to conserve disc space on the VA central server. Trace plots and Geweke statistics suggested MCMC convergence with adequate mixing (Figure S11). Generally speaking, shorter MCMC runs and burn-ins should be adequate for most applications. For example, a run of 5500 iterations with a burn-in of 500 yielded nearly identical results in the current case study (Figure S12 and Table S6). Table 4 presents the posterior means and 95% CrIs for the model parameters. The negative fixed effects estimates for year ( $\beta_{11}$  and  $\beta_{21}$ ) suggest that there was a general decline in admissions

Model Component	Parameter	Variable	Posterior Mean (95% CrI)
Binary	$\beta_{10}$	Intercept	0.11 (−0.10, 0.34)
	$\beta_{11}$	Year	−0.07 (−0.09, −0.06)
	$\beta_{12}$	Age	0.00 (−0.002, 0.002)
	$\beta_{12}$	Male Gender	−0.02 (−0.12, 0.09)
	$\beta_{12}$	NHW Race	0.03 (−0.01, 0.08)
	$\beta_{12}$	Full Disability Coverage	−0.12 (−0.17, −0.07)
	$\beta_{12}$	Elixhauser Score	0.43 (0.41, 0.45)
Count	$\beta_{20}$	Intercept	0.86 (0.72, 1.00)
	$\beta_{21}$	Year	−0.04 (−0.05, −0.03)
	$\beta_{22}$	Age	0.003 (0.002, 0.005)
	$\beta_{22}$	Male Gender	0.05 (−0.01, 0.12)
	$\beta_{22}$	NHW Race	−0.04 (−0.07, −0.01)
	$\beta_{22}$	Full Disability Coverage	−0.04 (−0.09, −0.01)
	$\beta_{22}$	Elixhauser Score	0.16 (0.15, 0.17)
	$r$	Dispersion	0.77 (0.72, 0.83)
Random Effects	$\Gamma_{11}$	$\text{Var}(\phi_{1i1})$	0.04 (0.03, 0.05)
	$\Gamma_{12}$	$\text{Cov}(\phi_{1i1}, \phi_{1i2})$	0.01 (0.001, 0.01)
	$\Gamma_{13}$	$\text{Cov}(\phi_{1i1}, \phi_{2i1})$	0.01 (0.002, 0.02)
	$\Gamma_{14}$	$\text{Cov}(\phi_{1i1}, \phi_{2i2})$	0.01 (0.001, 0.01)
	$\Gamma_{22}$	$\text{Var}(\phi_{1i2})$	0.02 (0.01, 0.02)
	$\Gamma_{23}$	$\text{Cov}(\phi_{1i2}, \phi_{2i1})$	0.01 (0.001, 0.01)
	$\Gamma_{24}$	$\text{Cov}(\phi_{1i2}, \phi_{2i2})$	0.003 (0.001, 0.01)
	$\Gamma_{33}$	$\text{Var}(\phi_{2i1})$	0.06 (0.05, 0.09)
	$\Gamma_{34}$	$\text{Cov}(\phi_{2i1}, \phi_{2i2})$	0.01 (0.01, 0.02)
	$\Gamma_{44}$	$\text{Var}(\phi_{2i2})$	0.02 (0.01, 0.02)

Table 4: Parameter estimates and 95% credible intervals (CrIs) for the spatiotemporal ZINB model in the VA inpatient study. NHW: Non-Hispanic white.

over time. This finding is consistent with recent VA efforts to reduce inpatient admissions through improved outpatient services (Kaboli et al., 2012). Additionally, patients with full disability coverage had fewer admissions. This finding supports recent studies showing that patients with full disability coverage are more likely to seek outpatient care because their copays are fully covered (Chuan-Fen et al., 2012). Not surprisingly, higher comorbidity scores were associated with increased admissions. The random effect covariances showed modest heterogeneity across counties in both components of the model.

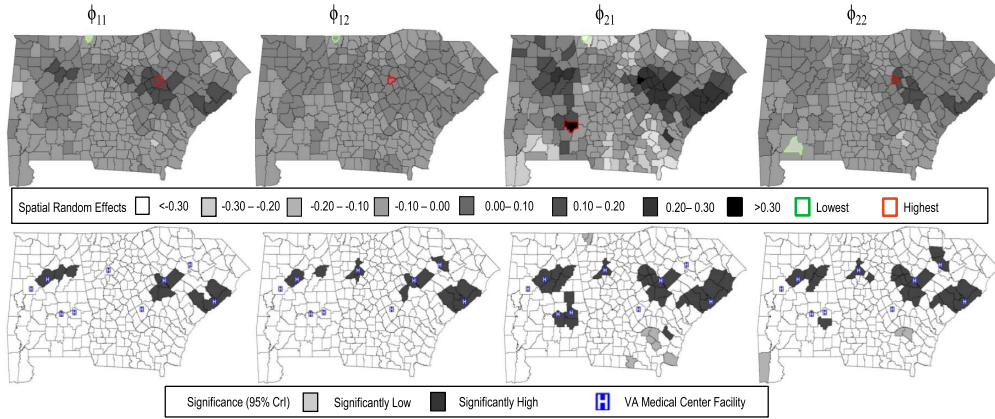


Figure 3: County maps of predicted spatial random effects (top panel) and posterior significance (lower panel) for the VA inpatient study. From left to right, the maps correspond to the random intercepts and slopes for the binary component, and the random intercepts and slopes for the count component. For lower panel, dark gray denotes significantly high random effect, light gray denotes significantly low, and white denotes non-significant.

Figure 3 maps the spatial random effects for each component. The upper panels show the predicted random effect values, while the lower panels map the posterior significance, with the white shade representing non-significant effects (i.e., a 95% CrI overlapping zero), the dark shade corresponding to positive significance (95% CrI  $> 0$ ), and light shade denoting counties with significantly negative effects (95% CrI  $< 0$ ). VA medical centers are superimposed on the maps. The random intercept for the count component (upper panel, map 3) showed the greatest variability, confirming the result found in Table 4 for  $\Gamma_{33}$ . In general, the maps show a band of elevated spatial effects extending from southeast South Carolina through central Georgia and Alabama, with several hotspots of elevated random effects in urban areas such as Charleston and Columbia, SC; Augusta and Atlanta, GA; and Birmingham, AL. These areas are home to large VA facilities. Thus, after controlling for other factors, including comorbidity burden, patients residing near urban VA facilities tend to have more annual admissions compared to those in more rural areas. Recent studies have shown that urban medical facilities typically have larger bed capacities compared to rural facilities; as a result, increased

admissions may be a byproduct of “discretionary” factors such as hospital capacity rather than clinical factors such as severity of illness (Fisher et al., 2000). Our findings appear to support this conclusion.

Table 5 presents the predicted marginal mean number of admissions,  $E(y_{ij}) = \pi_{ij}\mu_{ij}$ , for patients residing in three hypothetical counties in years 2011 and 2015, along with accompanying multiplicative ratios. All three patients were from the reference covariate population. The first county corresponded to an “average” county in which spatial random effects were set to zero. The spatial random effects for the remaining two counties were one standard deviation above and one standard deviation below average, respectively. As Table 5 indicates, there was a decrease in expected counts for patients in the average and below-average counties. This is consistent with the negative fixed effect coefficients for year ( $\beta_{11}$  and  $\beta_{21}$ ) found in Table 4. In contrast, for the above-average county, there was an increase over time, reflecting the fact that the positive random slope standard deviations ( $\sqrt{0.02} = 0.14$ ) were larger than the negative fixed effect coefficients for year, resulting in a net increase over time. The multiplicative ratios comparing an average county to a below-average county were 1.42 (1.35, 1.46) in 2011 and 3.37 (3.05, 3.77) in 2015. Thus, in 2015, patients in the average county had 3.37 times more admissions on average than patients in the below-average county. The multiplicative ratios comparing above- and below-average counties were 1.98 (1.83, 2.18) in 2011 and an impressive 10.17 (8.35, 12.46) in 2015. These results suggest that while there is an overall decline in admissions over time for patients residing in average or below-average counties, there appears to be substantial spatial heterogeneity in the magnitude of the trend across counties.

Mean No. of Admissions	Year	
	2011	2015
Average County	0.97 (0.85, 1.10)	0.71 (0.62, 0.81)
1 SD Above Average	1.36 (1.19, 1.54)	2.12 (1.82, 2.46)
1 SD Below Average	0.68 (0.59, 0.78)	0.21 (0.17, 0.25)
Multiplicative Ratios		
Average vs. Below-Average	1.42 (1.35, 1.48)	3.37 (3.05, 3.77)
Above-Average vs. Average	1.40 (1.34, 1.46)	3.00 (2.74, 3.31)
Above-Average vs. Below-Average	1.98 (1.83, 2.18)	10.17 (8.35, 12.46)

Table 5: Mean number of annual admissions per patient and corresponding multiplicative ratios for patients residing in 3 hypothetical counties. 95% credible intervals are given in parentheses. Estimates are for the reference covariate group. Random effects for the average county were set to 0. Random effects for the remaining counties were set to 1 standard deviation (SD) above and 1 SD below average.

The space-time interaction is highlighted more prominently in Figures 4(a) and 4(b). Figure 4(a) presents the mean number of admissions per patient for each county in 2011 and 2015, while Figure 4(b) displays the net change over time in expected admissions per patient for each county. Estimates correspond to a patient in the reference covariate group. Approximately 12% of the counties had increasing trends over time. These coun-

ties were concentrated in urban areas such as Charleston, Augusta and Birmingham, which again are home to large VA medical centers. These counties could be targeted for policy initiatives, such as improved outpatient services, to reduce inpatient admissions. Such efforts also have important cost-saving implications: a recent VA report estimates the per-patient daily cost of inpatient care to be \$3300 (Health Economic Resource Center, 2017). By pinpointing facilities associated with frequent inpatient admissions, the VA can help manage overhead costs while minimizing the burden imposed on both patients and hospital staff.

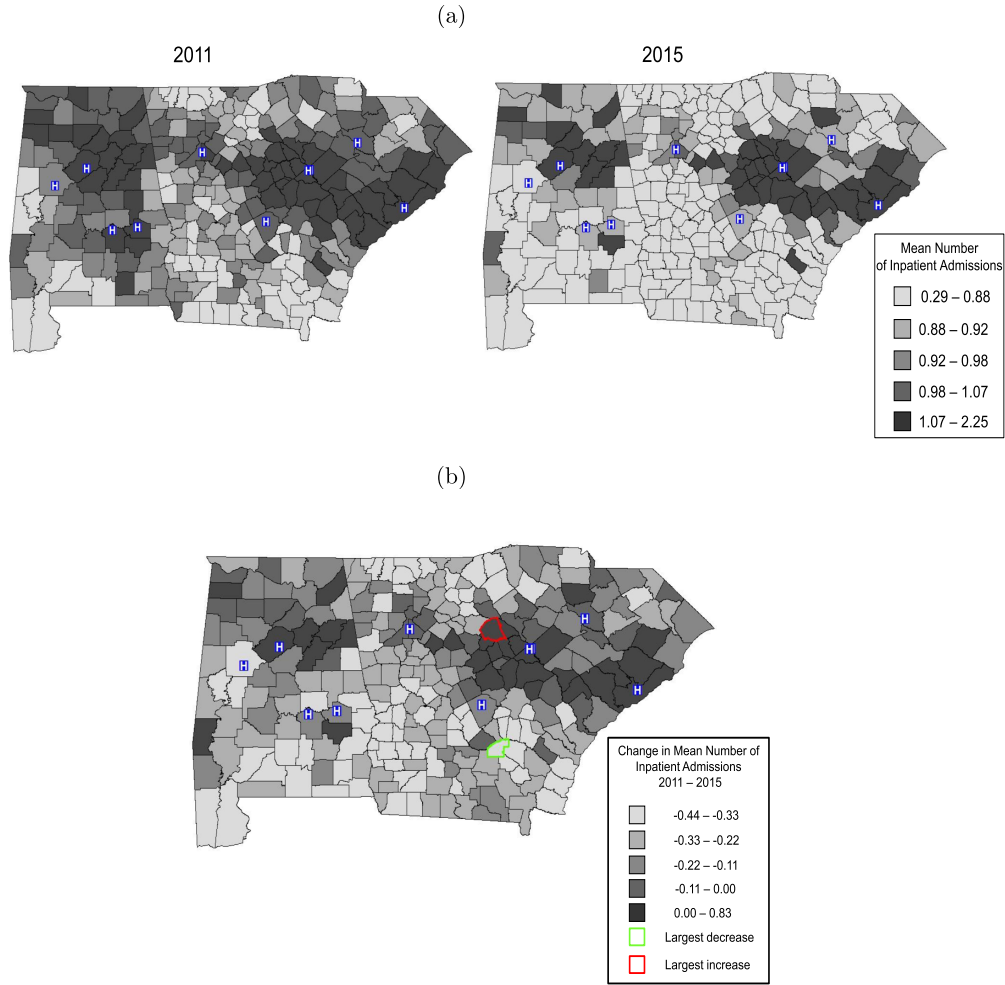


Figure 4: Panel (a): Predicted per-patient admissions by county. Panel (b): Change in mean number of admissions per patient from 2011–2015. Both figures correspond to a patient in the reference covariate group. In Figure (b), the darkest shade shows counties with net increase in admissions. “H” denotes VA medical center.



## 5 Conclusion

We have proposed an efficient Bayesian approach to fitting ZINB models. The proposed data-augmented Gibbs sampler makes use of easily sampled Pólya-Gamma random variables; conditional on these latent variables, inference proceeds via straightforward Bayesian inference for linear models. As such, the model can be easily extended to more complex settings, including those involving multivariate, longitudinal and spatiotemporal data. Our simulations showed that the approach performs well across a range of scenarios, even in the case of few at-risk observations. For simpler models, the approach yields estimates similar to maximum likelihood, but can accommodate more complex data that are not amenable to current methods. In terms of computation time, our simulations suggest that the approach is comparable to existing software when such comparisons are available.

There are a number of potential areas for future work. Although the ZINB is among the most common choices for modeling zero-inflated data, it cannot accommodate underdispersion, which occurs when there are fewer counts than expected under a standard count model. Future work might consider alternative count distributions that permit underdispersion, such as the generalized Poisson (Consul, 1989), while preserving the convenient Gibbs updates presented here. The model could also be extended to accommodate high-dimensional geostatistical data through the use of reduced rank and predictive process models (Banerjee, 2017). Restricted spatial regression could further be used to address spatial confounding due to collinearity between spatial random effects and spatially varying, cluster-level covariates (Hodges and Reich, 2010). Other extensions include finite mixture ZINB models to study underlying subgroups in the population, and shrinkage priors for high-dimensional predictors. More generally, the proposed method should prove useful in settings where interest lies in modeling zero-inflated count data within a Bayesian inferential framework.

## Supplementary Material

Supplementary material for “Bayesian Zero-Inflated Negative Binomial Regression Based on Pólya-Gamma Mixtures” (DOI: [10.1214/18-BA1132SUPP](https://doi.org/10.1214/18-BA1132SUPP); .pdf). This supplement contains derivations of the full conditionals discussed in Section 2 (Appendices A and B), additional tables and figures for the simulation studies presented in Section 3 (Appendix C), and additional tables and figures for case study presented in Section 4 (Appendix D).

## References

- Banerjee, S. (2017). “High-Dimensional Bayesian Geostatistics.” *Bayesian Analysis*, 12(2): 583–614. [MR3654826](#). doi: <https://doi.org/10.1214/17-BA1056R>. 24
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC, second edition. [MR3362184](#). 9, 10



- Celeux, G., Forbes, F., Robert, C. P., and Titterton, D. M. (2006). “Deviance information criteria for missing data models.” *Bayesian Analysis*, (4): 651–673. MR2282197. doi: <https://doi.org/10.1214/06-BA122>. 13
- Chuan-Fen, L., Bryson, C. L., Burgess, J. F., Sharp, N., Perkins, M., and Maciejewski, M. (2012). “Use of outpatient care in VA and Medicare among disability-eligible and age-eligible veteran patients.” *BMC Health Services Research*, 12(51). 21
- Consul, P. (1989). *Generalized Poisson Distributions: Properties and Applications*. New York: Marcel Dekker. MR0974108. 24
- Dadaneh, S. Z., Zhou, M., and Qian, X. (2018). “Bayesian negative binomial regression for differential expression with confounding factors.” *Bioinformatics*, 34(19): 3349–3356. 7
- Fisher, E. S., Wennberg, J. E., Stukel, T. A., Skinner, J. S., Sharp, S. M., Freeman, J. L., and Gittelsohn, A. M. (2000). “Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors.” *BMC Health Services Research*, 34(6): 1351. 22
- Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2017). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN. R package version 1.3-2. 7
- Furrer, R. and Sain, S. (2010). “spam: A Sparse Matrix R Package with Emphasis on MCMC Methods for Gaussian Markov Random Fields.” *Journal of Statistical Software, Articles*, 36(10): 1–25. 16
- Gelman, A., Hwang, J., and Vehtari, A. (2014). “Understanding Predictive Information Criteria for Bayesian Models.” *Statistics and Computing*, 24(6): 997–1016. MR3253850. doi: <https://doi.org/10.1007/s11222-013-9416-2>. 13, 17
- Gerber, F. and Furrer, R. (2015). “Pitfalls in the implementation of Bayesian hierarchical modeling of areal count data: An illustration using BYM and Leroux Models.” *Journal of Statistical Software, Code Snippets*, 63(1): 1–32. 16
- Geweke, J. (1992). “Evaluating the accuracy of sampling-based approaches to calculating posterior moments.” In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, 169–193. Oxford: Clarendon Press. MR1380276. 7
- Ghosh, S. K., Mukhopadhyay, P., and Lu, J.-C. (2006). “Bayesian analysis of zero-inflated regression models.” *Journal of Statistical Planning and Inference*, 136(4): 1360–1375. MR2253768. doi: <https://doi.org/10.1016/j.jspi.2004.10.008>. 2
- Health Economic Resource Center (2017). “Inpatient Average Cost Data Table, 2000–2016.” Technical report, US Department of Veterans Affairs, Washington, DC. 23
- Hodges, J. S. and Reich, B. J. (2010). “Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love.” *The American Statistician*, 64(4): 325–334. MR2758564. doi: <https://doi.org/10.1198/tast.2010.10052>. 24

- Kaboli, P., Go, J., Hockenberry, J., and et al. (2012). “Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 veterans affairs hospitals.” *Annals of Internal Medicine*, 157(12): 837–845. 21
- Lambert, D. (1992). “Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing.” *Technometrics*, 34(1): 1–14. 1
- Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2014). *The BUGS Book: A practical introduction to Bayesian analysis*. Boca Raton: Chapman & Hall/CRC. 2
- Neelon, B. (2018). “Supplementary material for “Bayesian Zero-Inflated Negative Binomial Regression Based on Pólya-Gamma Mixtures””. *Bayesian Analysis*. doi: <https://doi.org/10.1214/18-BA1132SUPP>. 4
- Neelon, B. H., O’Malley, A. J., and Normand, S.-L. T. (2010). “A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use.” *Statistical Modelling*, 10(4): 421–439. MR2797247. doi: <https://doi.org/10.1177/1471082X0901000404>. 2
- Pillow, J. and Scott, J. (2012). “Fully Bayesian inference for neural models with negative-binomial spiking.” In Bartlett, P., Pereira, F., Burges, C., Bottou, L., and Weinberger, K. (eds.), *Advances in Neural Information Processing Systems 25*, 1907–1915. MIT Press. 2, 3, 4, 6, 7
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). “CODA: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, 6(1): 7–11. URL <https://journal.r-project.org/archive/> 7
- Polson, N. G., Scott, J. G., and Windle, J. (2013a). “Bayesian inference for logistic models using Pólya-Gamma latent variables.” *Journal of the American Statistical Association*, 108(504): 1339–1349. MR3174712. doi: <https://doi.org/10.1080/01621459.2013.829001>. 2, 4, 5, 6
- Polson, N. G., Scott, J. G., and Windle, J. (2013b). “Bayesian inference for logistic models using Pólya-Gamma latent variables.” Most recent version: Feb. 2013. URL <http://arxiv.org/abs/1205.0310> 6
- Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Duncan Saunders, L., Beck, C., Feasby, T., and A Ghali, W. (2005). “Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data.” *Medical care*, 43: 1130–1139. 20
- Su, L., Tom, B. D. M., and Farewell, V. T. (2009). “Bias in 2-part mixed models for longitudinal semicontinuous data.” *Biostatistics*, 10(2): 374–389. 9
- U.S. Census Bureau (2014). “TIGER/Line Shapefiles.” Suitland, MD. 15
- Watanabe, S. (2010). “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory.” *Journal of Machine Learning Research*, 11: 3571–3594. MR2756194. 13

Zhou, M. and Carin, L. (2015). “Negative Binomial Process Count and Mixture Modeling.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37: 307–320.

[4](#), [7](#)

Zurr, A. F., Saveliev, A. A., and Ieno, E. N. (2012). *Zero Inflated Models and Generalized Linear Mixed Models with R*. Newburgh: Highland Statistics Ltd. [2](#)

### **Acknowledgments**

This work was supported in part by Merit Award HX002299-01A2 from the U.S. Department of Veterans Affairs Health Services Research and Development Program. The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government. Special thanks to Melanie Davis for her assistance with this manuscript.