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

Brian Neelon*

Abstract. Motivated by a study examining spatiotemporal patterns in inpatient length of stay, 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 duration of inpatient stays 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 hospitalization days among United States (US) veterans with type 2 diabetes. The majority of patients had no inpatient hospitalizations, resulting in a count of zero, while some had relatively brief stays and a small fraction had lengthy stays. 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

© 0000 International Society for Bayesian Analysis DOI: 0000

imsart-ba ver. 2014/10/16 file: ZINB21.tex date: August 8, 2018
Bayesian Zero-Inflated Negative Binomial Regression

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 the length of inpatient hospital stays 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 days, 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 stays 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

\[ y_i \sim (1 - \pi_i) \mathbb{I}_{(w_i = 0 \land y_i = 0)} + \pi_i \text{NB}(\mu_i, r) \mathbb{I}_{(w_i = 1)}, \quad i = 1, \ldots, n, \quad (2.1) \]

where \( y_i \) is the count response for individual \( i \), \( \mathbb{I} \) is the indicator function, and \( w_i \) is a latent “at-risk” indicator variable such that with probability \( 1 - \pi_i \), \( w_i = 0 \) 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, \( w_1, \ldots, w_n \), are typically modeled using a logistic model of the form

\[ \logit(\pi_i) = \log \left[ \Pr(w_i = 1 | \beta_1) \right] = x_i^T \beta_1 = \eta_{1i}, \quad (2.2) \]

where \( x_i \) is a \( p \times 1 \) vector of covariates and \( \beta_1 \) is a vector of regression parameters. Equation (2.2) is commonly referred to as the binary or logistic component of the ZINB model, which we denote with the subscript “1” in equation (2.2). Next, for reasons discussed below, we follow Pillow and Scott (2012) and parameterize the negative binomial component (conditional on \( w_i = 1 \)) as

\[ p(y_i | r, \beta_2, w_i = 1) = \frac{\Gamma(y_i + r) (1 - \psi_i)^r \psi_i^{y_i}}{\Gamma(r) y_i!} \quad \forall i \text{ s.t. } w_i = 1, \]

where \( \psi_i = \exp(x_i^T \beta_2) \). Equation (2.3) 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

\[ E(y_i | r, \beta_2, w_i = 1) = \frac{r \psi_i}{1 - \psi_i} = r \exp(\eta_{2i}) = \mu_i \]

\[ \text{Var}(y_i | r, \beta_2, w_i = 1) = \frac{r \psi_i}{(1 - \psi_i)^2} = r \exp(\eta_{2i}) \left[ 1 + \exp(\eta_{2i}) \right] = \mu_i (1 + \mu_i / r). \quad (2.4) \]

The marginal mean, averaged over \( w_i \), is \( E(y_i) = \pi_i \mu_i \). The parameter \( \alpha = 1/r \) captures the overdispersion in the data, so that as \( \alpha \to \infty \), the counts become increasingly dispersed relative to the Poisson. Above, we have assumed the same set of covariates \( x_i \) 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:
Bayesian Zero-Inflated Negative Binomial Regression

1. Given current parameter values, update $\beta_1$ using the Gibbs sampler proposed by Polson et al. (2013a) for logistic regression.

2. Conditional on $w_i = 1$, update $\beta_2$ using the negative binomial Gibbs sampler proposed by Pillow and Scott (2012).

3. Update $r$ using a random-walk Metropolis-Hastings step.

4. Update the latent at-risk indicators, $w_1, \ldots, w_n$, from their discrete full conditional distributions.

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

**Step 1: Bayesian Logistic Regression**

To implement Step 1, 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 $c \in \mathbb{R}$, if

$$\omega \sim \text{PG}(b, c) \overset{d}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k - 1/2)^2 + c^2/(4\pi^2)}, \quad (2.5)$$

where the $g_k$’s are independently distributed according to a $\text{Ga}(b, 1)$ distribution. Polson et al. (2013a) establish two important properties of the PG($b, c$) density. First, for $a \in \mathbb{R}$ and $\eta \in \mathbb{R}$, it follows that

$$\frac{(e^{\eta})^a}{(1 + e^{\eta})^b} = 2^{-b} e^{\kappa \eta} \int_0^\infty e^{-\omega \eta^2/2} p(\omega|b, 0) \, d\omega, \quad (2.6)$$

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

$$p(\omega|b, c) = \frac{\exp(-c^2 \omega/2)p(\omega|b, 0)}{E_\omega[\exp(-c^2 \omega/2)]} = \frac{\exp(-c^2 \omega/2)p(\omega|b, 0)}{\int_0^\infty e^{-c^2 \omega/2} p(\omega|b, 0) \, d\omega}. \quad (2.7)$$

Under the logistic model in equation (2.2), the Bernoulli likelihood for the at-risk indicators $\mathbf{w} = (w_1, \ldots, w_n)^T$ is

$$p(\mathbf{w}|\beta_1) = \prod_{i=1}^{n} p(w_i|\beta_1)$$
\[ \prod_{i=1}^{n} \left( \frac{\exp(\eta_{1i})}{1 + \exp(\eta_{1i})} \right)^{w_i} \left( \frac{1}{1 + \exp(\eta_{1i})} \right)^{1-w_i} \]

where \( \eta_{1i} = x_i^T \beta_1 \). The \( i \)-th element of the Bernoulli likelihood has the same form as the left-hand expression in equation (2.6), with \( a_i = w_i \) and \( b = 1 \). Thus, we can re-write the Bernoulli likelihood in terms of the Pólya-Gamma random variables \( \omega_1 = (\omega_{11}, \ldots, \omega_{1n})^T \) according to equation (2.6):

\[ p(w_i|\beta_1) \propto e^{\kappa_i\eta_{1i}} \int_{0}^{\infty} e^{-\omega_{1i}^T\eta_{1i}^2/2} p(\omega_{1i}|1,0) \, d\omega_{1i}, \]

(2.9)

where \( \kappa_i = w_i - 1/2 \). Let \( \omega_{1i} (i = 1, \ldots, n) \) be independently distributed according to 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 \( \beta_1 \), given \( w \) and \( \omega_1 \), is

\[ p(\beta_1|w, \omega_1) \propto \pi(\beta_1) \exp \left[ -\frac{1}{2} (z_1 - X\beta_1)^T \Omega_1 (z_1 - X\beta_1) \right], \]

(2.10)

where \( \pi(\beta_1) \) is the prior distribution for \( \beta_1 \); for \( i = 1, \ldots, n \), \( z_i = w_i^{-1/2} \) with \( z_1 = (z_{11}, \ldots, z_{1n})^T \); \( \Omega_1 = \text{diag}(\omega_1) \) is an \( n \times n \) precision matrix; and \( X \) is an \( n \times p \) design matrix. It is clear that, given \( \beta_1 \) and \( \Omega_1 \), \( z_1 \) is normally distributed with mean \( \eta_1 = X\beta_1 \) 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 \( z_1 \) and \( \Omega_1 \) is \( N_p(\mu, \Sigma) \), where

\[ \Sigma = \left( \Sigma_0^{-1} + X^T \Omega_1 X \right)^{-1} \]

\[ \mu = \Sigma \left( \Sigma_0^{-1} \beta_0 + X^T \Omega_1 z_1 \right). \]

(2.11)

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 1 proceeds by selecting initial values for \( \beta_1 \) and \( w \) and iterating through the following steps:

1. For \( i = 1, \ldots, n \), update \( \omega_{1i} \) from a PG(1, \( \eta_{1i} \)) density, where \( \eta_{1i} = x_i^T \beta_1 \)
2. For \( i = 1, \ldots, n \), define \( z_{1i} = w_i^{1/2} \)/\( \omega_{1i} \)
3. Conditional on \( z_1 \), update \( \beta_1 \) from \( N_p(\mu, \Sigma) \), where \( \mu \) and \( \Sigma \) are given in (2.11).

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).
Bayesian Zero-Inflated Negative Binomial Regression

Steps 2 and 3: Bayesian Negative Binomial Regression

The Gibbs sampler for the negative binomial component is similar to the one for the logistic component. Adopting the parameterization of the negative binomial in equation (2.3), the conditional likelihood of \( y_i \) given \( w_i = 1 \) is

\[
p(y_i|r, \beta_2, w_i = 1) \propto \exp(\eta_{2i})^{w_i} \frac{(1 - \psi_i)^r \psi_i^{y_i}}{[1 + \exp(\eta_{2i})]^{r+y_i}},
\]

where \( \eta_{2i} = x_i^T \beta_2 \). Exploiting property 1 of the Pólya-Gamma distribution in equation (2.6), it follows that

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

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

\[
p(\beta_2|y^*, r, w, \omega_2) \propto \pi(\beta_2) \exp \left[ -\frac{1}{2} (z_2 - X^* \beta_2)^T \Omega_2 (z_2 - X^* \beta_2) \right],
\]

where \( y^* \) is the \( n^* \times 1 \) subvector of \( y \) 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 \( \omega_{2i} \); \( z_2 \) is a vector of length \( n^* \) with elements \( z_{2i} = \frac{y_i - r}{\omega_{2i}} \); \( \Omega_2 = \text{diag}(\omega_2) \) is an \( n^* \times n^* \) precision matrix; and \( X^* \) is an \( n^* \times p \) design matrix. From (2.14), it is clear that \( z_2 \) is normally distributed with mean \( \eta_{2i} = X_i^T \beta_2 \) and diagonal covariance \( \Omega_2^{-1} \). Thus, assuming a \( \text{N}_p(\beta_0, \Sigma_0) \) prior for \( \beta_2 \), the conjugate full conditional for \( \beta_2 \) given \( z_2 \) and \( \Omega_2 \) is \( \text{N}_p(\mu, \Sigma) \), where

\[
\Sigma = \left( \Sigma_0^{-1} + X^* \Omega_2 X^* \right)^{-1} \\
\mu = \Sigma \left( \Sigma_0^{-1} \beta_0 + X^* \Omega_2 z_2 \right).
\]

The proof can be found in Pillow and Scott (2012) and is summarized in the context of the ZINB in Appendix A of the online supplement.

Thus, given current values for \( \beta_2, w, \) and \( r \), the Gibbs sampler for Steps 2 and 3 proceeds as follows:

1. For \( w_i = 1 \), draw \( \omega_{2i} \) from its PG\((y_i + r, \eta_{2i})\) distribution, where \( \eta_{2i} = x_i^T \beta_2 \)
2. For \( w_i = 1 \), define \( z_{2i} = \frac{y_i - r}{\omega_{2i}} \)
3. Update \( \beta_2 \) from its \( \text{N}_p(\mu, \Sigma) \) distribution, where \( \mu \) and \( \Sigma \) are given in (2.15)
4. Update \( r \) using a random-walk Metropolis-Hastings step.
We assume a LN(0, 1) prior for \( r \), where \( \text{LN}(\cdot) \) denotes a lognormal density. For the Metropolis-Hastings update, we draw candidate values of \( r \) from a zero-truncated normal proposal centered at the current value of \( r \). Alternatively, one could use the Gibbs sampling approach developed by Zhou et al. (2012), although in our experience the univariate Metropolis-Hastings update is straightforward to implement and works well in practice.

To complete the prior specification for the ZINB model, we assign weakly informative \( N_p(0, 100 I_p) \) priors to \( \beta_1 \) and \( \beta_2 \). 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.

**Step 4: Updating the Latent At-Risk Indicators**

In the final step of the Gibbs sampler, we update the at-risk indicators, \( w_1, \ldots, w_n \). As outlined in the online supplement, 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

\[
\theta_i = \Pr(w_i = 1 \mid y_i = 0, \text{rest}) = \Pr(\text{at-risk zero} \mid \text{at risk or structural zero})
\]

\[
= \frac{\pi_i \psi_i}{1 - \pi_i (1 - \psi_i)}.
\]

where, from equation (2.2), \( \pi_i = \exp(\eta_i) / [1 + \exp(\eta_i)] \) is the unconditional probability that \( w_i = 1 \), and \( \psi_i = 1 - \psi_i \), where \( \psi_i \) is the negative binomial event probability defined in equation (2.3). The result follows from a direct application of Bayes’ Theorem. The proof is presented in Appendix A of the supplement.

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 and Geweke’s \( z \)-statistic (Geweke, 1992). 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 \( z_1 \) and \( z_2 \), the ZINB model can easily be extended to accommodate longitudinal, spatial, and time series data — essentially any
Bayesian Zero-Inflated Negative Binomial Regression

setting where the model parameters are linear in $z_1$ and $z_2$. Suppose, for example, we have count responses, $y_{ij}$, measured at occasions $j = 1, \ldots, n_i$ for individual $i$. We can model the data using a longitudinal version of the ZINB model, which is expressed as

$$y_{ij} \sim (1 - \pi_{ij}) I(w_{ij} = 0 \land y_{ij} = 0) + \pi_{ij} \text{NB}(\mu_{ij}, r) I(w_{ij} = 1), \quad (2.17)$$

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 (2.4). As before, we model $\pi_{ij}$ using a logit link

$$\text{logit}(\pi_{ij}) = \text{logit}[\Pr(w_{ij} = 1|\beta_1, \phi_{1i})] = x_{ij}^T \beta_1 + w_{ij}^T \phi_1, \quad \pi_{ij}, \quad (2.18)$$

where $x_{ij}$ 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; $w_{ij}$ is a $q \times 1$ random effect design vector that includes functions of time; and $\phi_{1i} \sim N_q(0, G_1)$ is a $q \times 1$ vector of random effects for the binary component, with $q \times q$ prior covariance $G_1$. Similarly, the negative binomial component (conditional on $w_{ij} = 1$) is modeled as

$$p(y_{ij}|r, \beta_2, \phi_{2i}, w_{ij} = 1) = \frac{\Gamma(y_{ij} + r)}{\Gamma(r) \psi_{ij}^r} \frac{(1 - \psi_{ij})^{y_{ij}}}{1 + \exp(x_{ij}^T \beta_2 + w_{ij}^T \phi_2)}, \quad \forall i, j \text{ s.t. } w_{ij} = 1$$

where $\phi_{2i} \sim N_q(0, G_2)$ is a $q \times 1$ vector of random effects for the count component with $q \times q$ covariance $G_2$. 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 $z_{1i} = (z_{1i1}, \ldots, z_{1in_i})^T$ as

$$z_{1i}|\eta_{1i}, \Omega_{1i} \sim N_{n_i}(\eta_{1i}, \Omega_{1i}^{-1}), \quad (2.20)$$

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

$$z_{2i}|\eta_{2i}, \Omega_{2i} \sim N_{n_i^*}(\eta_{2i}, \Omega_{2i}^{-1}) \quad (2.21)$$

Here, $z_{2i}$ 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 $i$; $\eta_{2i} = X_i^* \beta_2 + V_i^* \phi_{2i}$ is the $n_i^* \times 1$ linear predictor for the
count component, where $X_i^*$ and $V_i^*$ are $n_i^* \times p$ and $n_i^* \times q$ fixed and random effect design matrices; $\Omega_{2i} = \text{diag}(\omega_{2i})$ is an $n_i^* \times n_i^*$ precision matrix; and $\omega_{2i}$ is an $n^* \times 1$ 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 $j = 1, \ldots, n_i$. This is unlikely to occur unless $n_i$ is small or $\phi_{ij}$ is low for $j = 1, \ldots, n_i$, 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 $\phi_{2i}$ 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 longer stays 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 $\phi_i = (\phi_{1i}^T, \phi_{2i}^T)^T$ 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 $\phi_i$:

$$
\phi_i | \Gamma \sim N_{2q}(0, \Gamma)
$$

where

$$
\Gamma = \begin{pmatrix}
G_1 & G_{12} \\
G_{21} & G_2
\end{pmatrix}
$$

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 $G_{12} = G_{21}$.

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 (ie., $n_i \geq n_i^*$) 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 $\phi_i$ in equation (2.22). For example, an intrinsic CAR prior (Banerjee et al., 2014) takes the form

$$
\phi_i | \phi_{(-i)}, \Gamma \sim N_{2q} \left( \frac{1}{m_i} \sum_{i \in \partial_i} \phi_{il}, \frac{1}{m_i} \Gamma \right),
$$

where $m_i$ is the number of neighbors for $i$-th areal unit, $\partial_i$ is the set of neighbors for unit $i$, and $\Gamma$ is the $2q \times 2q$ conditional covariance matrix given the remaining spatial random
Bayesian Zero-Inflated Negative Binomial Regression

effects, \( \phi_{(-i)} \). Whereas the multivariate normal prior in equation (2.22) 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 \( \Gamma \) in equation (2.23). Let \( \phi_{11} = (\phi_{111}, \ldots, \phi_{1n1})^T \) and \( \phi_{12} = (\phi_{112}, \ldots, \phi_{1n2})^T \) denote, respectively, the \( n \times 1 \) vectors of random intercepts and slopes for the binary component. Similarly, define \( \phi_{21} = (\phi_{211}, \ldots, \phi_{2n1})^T \) and \( \phi_{22} = (\phi_{212}, \ldots, \phi_{2n2})^T \) to be the intercept and slope vectors for the count component. Finally, let \( \phi = (\phi_{11}, \ldots, \phi_{22})^T \), 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

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

where \( Q = M - A \) is an \( n \times n \) matrix of rank \( n - 1 \); \( M = \text{diag}(m_1, \ldots, m_n) \) with diagonal elements equal to the number of neighbors for each spatial unit; and \( A \) 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 \( Q \) is rank-deficient — and hence \( p(\phi|\Gamma) \) 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 (2.24), we see that the joint prior for \( \phi \) is proportional to a multivariate normal density with \( 2mn \times 2qn \) precision matrix \( \Gamma^{-1} \otimes Q \), which is generally too unwieldy for efficient posterior inference. A convenient alternative is to derive univariate conditional priors for \( \phi_{kk} \) \( (k = 1, \ldots, q) \) 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

\[
p(\phi_{11} | \phi_{12}, \phi_{21}, \phi_{22}, \Gamma) \propto \exp \left[ -\frac{1}{2} (\phi_{11} - \mu_{11})^T \Sigma_{11} (\phi_{11} - \mu_{11}) \right],
\]

where

\[
\Sigma_{11} = \left[ \Gamma_{11} - \Gamma_{(1,-1)} \Gamma_{(-1,1)}^{-1} \Gamma_{(-1,1)} \right]^{-1} Q
\]

and

\[
\mu_{11} = \left[ (\Gamma_{(1,-1)} \Gamma_{(-1,1)}^{-1}) \otimes I_n \right] \phi_{(-1)},
\]

where \( \Gamma \) is the \( 4 \times 4 \) covariance of \( \phi_i \); \( \Gamma_{11} \) denotes the first element of \( \Gamma \); \( \Gamma_{(1,-1)} \) is the \( 1 \times 3 \) vector comprising the first row of \( \Gamma \) with element 1 removed; \( \Gamma_{(-1,1)} \) is the \( 3 \times 3 \) submatrix of \( \Gamma \) after removing row 1 and column 1; \( \Gamma_{(-1,1)} \) is the \( 3 \times 1 \) vector comprising the first column of \( \Gamma \) with element 1 removed; and \( \phi_{(-1)} = (\phi_{12}, \phi_{21}, \phi_{22})^T \) is a \( 3n \times 1 \) vector of the remaining random effects. Prior (2.25) follows from conditional multivariate normal theory. Similar expressions hold for \( \phi_{12}, \phi_{21} \) and \( \phi_{22} \).
The conditional prior specification in (2.25) 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 $w_{ij} = 1$ ($i = 1, \ldots, n; j = 1, \ldots, n_i$). This sample-size imbalance prevents a joint Gibbs update for $\phi = (\phi_{11}^T, \ldots, \phi_{22}^T)^T$ based on prior (2.24). The conditional prior (2.25) 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 $Q = I_n$ 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 a $2q \times 2q$ scale matrix. By default, we set $\beta_0 = 0$, $\Sigma_0 = I_p$, and $\Lambda = I_{2q}$. To implement the MCMC for random effect ZINB models, we initialize the model parameters and then cycle through the following steps:

1. Update the parameters for the binary component:
   (a) For all $i, j$, sample $\omega_{1ij}$ from PG$(1, \eta_{1ij})$, where $\eta_{1ij}$ is defined in equation (2.18)
   (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 $k = 1, \ldots, q$, update each $n \times 1$ vector $\phi_{1k}$ from its normal full conditional based on the conditional prior specification given in (2.25); apply sum-to-zero constraints as needed

2. Update the parameters for the count component:
   (a) For $w_{ij} = 1$, sample $\omega_{2ij}$ from PG$(1, \eta_{2ij})$ where $\eta_{2ij}$ is defined in equation (2.19)
   (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 $k = 1, \ldots, q$, update each $n \times 1$ vector $\phi_{2k}$ from its normal full conditional based on the conditional prior specification given in (2.25); apply sum-to-zero constraints as needed
   (e) Update $r$ using a random-walk Metropolis-Hastings step analogous to the one outlined in Section 2.2

3. Update the $2q \times 2q$ covariance matrix $\Gamma$ from its conjugate inverse-Wishart full conditional
4. For all $i,j$, update the latent at-risk indicators, $w_{ij}$, according to the discrete probability distribution

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

where $\pi_{ij} = \exp(\eta_{1ij})/(1 + \exp(\eta_{1ij}))$ is the unconditional probability that $w_{ij} = 1$ given in equation (2.18), $\nu_{ij} = 1 - \psi_{ij}$, and $\psi_i = \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}$ is the negative binomial event probability defined in equation (2.19).

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 a fixed effects ZINB model and compared the results to maximum likelihood estimates (MLEs) obtained using the SAS® 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:

$$y_i \sim (1 - \pi_i) \mathbb{I}_{(w_i = 0 \land y_i = 0)} + \pi_i \text{NB}(\mu_i, r) \mathbb{I}_{(w_i = 1)}$$

$$\logit(\pi_i) = \logit[\Pr(w_i = 1| \beta_1)] = \beta_{i0} + \beta_{i1}x_{i1} + \beta_{i2}x_{i2} + \beta_{i3}x_{i3}$$

$$p(y_i|r, \beta_2, w_i = 1) \propto \frac{\Gamma(y_i + r)}{\Gamma(r)y_i!}(1 - \psi_i)^r \psi_i^{y_i}$$

$$\psi_i = \frac{\exp(\eta_{2i})}{1 + \exp(\eta_{2i})}$$

$$\eta_{2i} = \beta_{20} + \beta_{21}x_{i1} + \beta_{22}x_{i2} + \beta_{23}x_{i3},$$

where $\psi_i$ is defined in equation (2.3), $x_{i1}$ 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\}, $\beta_1 = (\beta_{10}, \ldots, \beta_{13})^T = (0.5, -0.5, -0.25, 0.25)^T$, $\beta_2 = (\beta_{20}, \ldots, \beta_{23})^T = (0.5, -1, 0.75, -0.25)^T$ 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. 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 assigned independent \( N(0, 100) \) priors to the regression coefficients and a \( \text{LN}(0, 1) \) 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 \( \beta_1 = \beta_2 = 0 \), and \( r = 1 \). We ran 2500 iterations of the MCMC algorithm described in Section 2, discarding the first 500 as burn-in. Figure S2 of the supplement presents trace plots 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 0.04 to 0.94, indicating reasonable convergence. The trace plots showed satisfactory mixing. The algorithm took 17 seconds to run on a Dell® Precision T3610 workstation, compared to 0.50 seconds for SAS. However, a shorter run of 1500 iterations took approximately 7 seconds to run and produced similar results, indicating that the proposed Bayesian method is comparable to SAS in terms of run time. As an additional comparison, we re-ran the model for 20,000 iterations and obtained nearly identical results (Figure S3 and Table S1).

Table 1 presents the parameter estimates and 95% intervals under Bayesian and maximum likelihood estimation. In all cases, the Bayesian estimates were as or more accurate than the 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.

<table>
<thead>
<tr>
<th>Model Component</th>
<th>Parameter</th>
<th>Simulated Value</th>
<th>Estimate (95% Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta_{10} )</td>
<td>0.50</td>
<td>0.52 (0.05, 1.26)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{11} )</td>
<td>0.50</td>
<td>0.50 (0.21, 0.89)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{12} )</td>
<td>–0.25</td>
<td>–0.06 (–0.52, 0.36)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{13} )</td>
<td>0.25</td>
<td>0.11 (–0.14, 0.38)</td>
</tr>
<tr>
<td>Count</td>
<td>( \beta_{20} )</td>
<td>0.50</td>
<td>0.41 (0.06, 0.74)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{21} )</td>
<td>–1.00</td>
<td>–0.96 (–1.09, –0.83)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{22} )</td>
<td>0.75</td>
<td>0.63 (0.40, 0.85)</td>
</tr>
<tr>
<td></td>
<td>( \beta_{23} )</td>
<td>–0.25</td>
<td>–0.11 (–0.25, 0.02)</td>
</tr>
<tr>
<td></td>
<td>( r )</td>
<td>1.00</td>
<td>1.12 (0.78, 1.61)</td>
</tr>
</tbody>
</table>

Table 1: Parameter estimates and 95% intervals for fixed effects ZINB model in simulation study 1. †Posterior means and 95% credible intervals for proposed Bayesian ZINB model. ‡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 (2.17):

\[
y_{ij} \sim (1 - \pi_{ij}) I_{(w_{ij}=0 \land y_{ij}=0)} + \pi_{ij} \text{NB}(\mu_{ij}, r) I_{(w_{ij}=1)}
\]
Bayesian Zero-Inflated Negative Binomial Regression

\[
\logit(\pi_{ij}) = \log \left[ \Pr(w_{ij} = 1|\eta_{ij}) \right] = \beta_{10} + \beta_{11}x_i + \beta_{12}t_{ij} + \phi_i
\]

\[
p(y_{ij} | r, \beta_2, \phi_{2i}, w_{ij} = 1) = \frac{\Gamma(y_{ij} + r)}{\Gamma(r) y_{ij}!} (1 - \psi_{ij})^{r} \psi_{ij}^{y_{ij}}
\]

\[
\psi_{ij} = \frac{\exp(\eta_{2ij})}{1 + \exp(\eta_{2ij})}
\]

\[
\eta_{2ij} = \beta_0 + \beta_{21}x_i + \beta_{22}t_{ij} + \phi_i
\]

\[
\phi_i = (\phi_{1i}, \phi_{2i})^T \sim N_2(0, \Gamma), \tag{3.2}
\]

where, for \(i = 1, \ldots, 1000\) and \(j = 1, \ldots, n_i\), \(y_{ij}\) 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); \(t_{ij} \sim \text{N}(0, 2)\) 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 \(\Gamma\). 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: \(\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})^T = (0.25, -0.25, 0.25)^T\), \(\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})^T = (0.50, -.25, 0.25)^T\), \(r = 1.25\), and \(\Gamma = \begin{pmatrix} 0.50 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}\). 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 \(\text{N}(0, 100)\) priors to the fixed effects and a LN\((0, 1)\) prior to \(r\). We reparameterized the bivariate normal prior for \(\phi_i\) using the conditional specification described in equation (2.25), leading to a conditional prior for \(\phi_{1i}\) of the form

\[
\phi_{1i} | \phi_{2i} \sim N(m, v)
\]

\[
v = \frac{\Gamma_{11} - \Gamma_{12} \Gamma_{12}}{\Gamma_{22}} = (1 - \rho^2) \Gamma_{11}
\]

\[
m = \frac{\Gamma_{21}}{\Gamma_{22}} \phi_{2i} = \rho \sqrt{\frac{\Gamma_{11}}{\Gamma_{22}}} \phi_{2i}, \tag{3.3}
\]

where \(\Gamma_{11}\) and \(\Gamma_{22}\) are the marginal variances of \(\phi_{1i}\) and \(\phi_{2i}\), respectively, and \(\rho = \frac{\Gamma_{12}}{\sqrt{\Gamma_{11} \Gamma_{22}}} = \text{Corr}(\phi_{1i}, \phi_{2i})\). The conditional prior for \(\phi_{2i}\) follows a similar expression. Finally, we assigned an inverse-Wishart\((2, I_2)\) prior to \(\Gamma\).
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 $\Gamma$ was initialized to $I_2$. To update $r$, we used a zero-truncated normal proposal with variance 0.003, resulting in an acceptance rate of 41%. We ran the sampler for 2500 iterations, discarding the first 500 as burn-in; the algorithm took 78 seconds to run, compared to 55 seconds for SAS. However, a 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. Figure S5 of the supplement presents trace plots and Geweke diagnostic p-values for selected parameters. The results indicate reasonable mixing. A longer run of 50,000 iterations produced nearly identical results, suggesting that the estimates stabilize quickly (Figure S6 and Table S2).

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.

<table>
<thead>
<tr>
<th>Model Component</th>
<th>Parameter</th>
<th>Simulated Value</th>
<th>Estimate (95% Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Proposed Model$^\dagger$</td>
<td>MLE$^\dagger$</td>
</tr>
<tr>
<td>Binary</td>
<td>$\beta_{10}$</td>
<td>0.25</td>
<td>0.06 (−0.17, 0.29)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>−0.25</td>
<td>−0.28 (−0.49, −0.21)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>0.25</td>
<td>0.29 (0.21, 0.38)</td>
</tr>
<tr>
<td>Count</td>
<td>$\beta_{20}$</td>
<td>0.50</td>
<td>0.41 (0.22, 0.59)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>−0.25</td>
<td>−0.11 (−0.26, 0.04)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>0.25</td>
<td>0.26 (0.21, 0.31)</td>
</tr>
<tr>
<td></td>
<td>$r$</td>
<td>1.25</td>
<td>1.32 (1.17, 1.51)</td>
</tr>
<tr>
<td>Random Effects</td>
<td>$\Gamma_{11} = \text{Var}(\phi_{1i})$</td>
<td>0.50</td>
<td>0.54 (0.32, 0.87)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{22} = \text{Var}(\phi_{2i})$</td>
<td>0.75</td>
<td>0.76 (0.64, 0.90)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{12} = \text{Cov}(\phi_{1i}, \phi_{2i})$</td>
<td>0.25</td>
<td>0.30 (0.20, 0.41)</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimates and 95% intervals for random effects ZINB model in simulation study 2. $^\dagger$Posterior means and 95% credible intervals obtained for proposed Bayesian ZINB model. $^\dagger$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:

\[
y_{ij} \sim (1 - \pi_{ij}) \mathbb{I}(w_{ij} = 0 \land y_{ij} = 0) + \pi_{ij} \text{NB}(\mu_{ij}, r) \mathbb{I}(w_{ij} = 1)
\]
\[
\text{logit}(\pi_{ij}) = \text{logit}[\Pr(w_{ij} = 1 | \beta_1, \phi_1)] = \beta_{10} + \beta_{11} t_{ij} + \phi_{11} + \phi_{12} t_{ij}
\]
\[
p(y_{ij}|\beta_2, \phi_{2i}, w_{ij} = 1) \propto \frac{\Gamma(\eta_{2i})}{\Gamma(r) y_{ij}!} (1 - \psi_{ij})^{\psi_{ij}}
\]
\[
\psi_{ij} = \frac{\exp(\eta_{2i})}{1 + \exp(\eta_{2i})}
\]
\[
\eta_{2i} = \beta_{20} + \beta_{21} t_{ij} + \phi_{21} + \phi_{22} t_{ij}
\]
\[
\phi_i | \phi_{(-i)}, \Gamma \sim \text{N}_4 \left( \frac{1}{m_i} \sum_{l \in \partial i} \phi_l, \frac{1}{m_i} \Gamma \right)
\]

(3.4)

where \( t_{ij} \in \{0, 1, 2, 3, 4\} \) denotes study year, with 0 as baseline; \( \phi_i = (\phi_{1i}, \phi_{12})^T \) is a vector comprising the \( i \)-th random intercept \((\phi_{1i})\) and slope \((\phi_{12})\) for the binary component; \( \phi_{2i} = (\phi_{21i}, \phi_{22})^T \) is the corresponding vector of random effects for the count component; \( \phi_i = (\phi_{1i}, \phi_{12}, \phi_{21i}, \phi_{22})^T \) is modeled as multivariate ICAR distribution with conditional covariance \( \Gamma \); 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 (3.4) 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 (Lunn et al., 2014), which may have contributed 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 \( \text{N}(0, 100) \) priors for the fixed effects and a \( \text{LN}(0, 1) \) prior for \( r \). Following equation (2.25), we partitioned the multivariate ICAR prior for \( \phi_i \) into separate univariate priors. We completed the prior specification by assigning an inverse-Wishart(4, \( I_4 \)) prior to \( \Gamma \). 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 \( \Gamma \) was initialized to \( I_4 \). To update \( r \), we used a zero-truncated normal proposal with variance 0.0002, resulting in an acceptance rate of 45%. To improve efficiency of the algorithm, we used the R package \texttt{spam} (Furrer and Sain, 2010; Gerber and Furrer, 2015) to convert the MCAR structure matrix \( Q \) in equation (2.24) to sparse matrix object. This avoids computationally intensive matrix operations designed for dense matrices. We ran the sampler for 2500 iterations with a burn-in of 500; the algorithm took 19 minutes to run. Trace plots and Geweke diagnostics were indicative of convergence and showed reasonable mixing for a range of model parameters (Figure S8). For comparison, we re-ran the analysis 1500 iterations, which took only 9 minutes to run, as well as for 50,000 iterations with a burn-in of 25,000 and a thinning interval of 25 (Figure S9 and Table S3). 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 $\phi_{1i2}$ ($i = 1, \ldots, n$), 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 mirror the true spatial distribution, suggesting the proposed model can accurately recover 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.

<table>
<thead>
<tr>
<th>Model Component</th>
<th>Parameter</th>
<th>Simulated Value</th>
<th>Posterior Mean (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>$\beta_{10}$</td>
<td>$-0.25$</td>
<td>$-0.28 (-0.34, -0.22)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>$0.25$</td>
<td>$0.22 (0.19, 0.25)$</td>
</tr>
<tr>
<td>Count</td>
<td>$\beta_{20}$</td>
<td>$0.50$</td>
<td>$0.47 (0.43, 0.53)$</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>$-0.25$</td>
<td>$-0.24 (-0.26, -0.22)$</td>
</tr>
<tr>
<td></td>
<td>$r$</td>
<td>$1.00$</td>
<td>$1.05 (0.98, 1.09)$</td>
</tr>
<tr>
<td>Random Effects</td>
<td>$\Gamma_{11} = \text{Var}(\phi_{1i1})$</td>
<td>$0.50$</td>
<td>$0.59 (0.42, 0.82)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{12} = \text{Cov}(\phi_{1i1}, \phi_{1i2})$</td>
<td>$0.10$</td>
<td>$0.03 (-0.05, 0.11)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{13} = \text{Cov}(\phi_{1i1}, \phi_{2i1})$</td>
<td>$0.10$</td>
<td>$0.10 (-0.01, 0.21)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{14} = \text{Cov}(\phi_{1i1}, \phi_{2i2})$</td>
<td>$-0.10$</td>
<td>$-0.07 (-0.13, -0.01)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{22} = \text{Var}(\phi_{1i2})$</td>
<td>$0.15$</td>
<td>$0.20 (0.14, 0.26)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{23} = \text{Cov}(\phi_{1i2}, \phi_{2i1})$</td>
<td>$0.10$</td>
<td>$0.12 (0.05, 0.19)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{24} = \text{Cov}(\phi_{1i2}, \phi_{2i2})$</td>
<td>$0.10$</td>
<td>$0.08 (0.04, 0.11)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{33} = \text{Var}(\phi_{2i1})$</td>
<td>$0.50$</td>
<td>$0.55 (0.42, 0.71)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{34} = \text{Cov}(\phi_{2i1}, \phi_{2i2})$</td>
<td>$0.10$</td>
<td>$0.06 (0.01, 0.11)$</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{44} = \text{Var}(\phi_{2i2})$</td>
<td>$0.15$</td>
<td>$0.17 (0.13, 0.22)$</td>
</tr>
</tbody>
</table>

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 $\Gamma$ in equation (2.22) to be block diagonal with $G_{12} = G_{21} = 0$. For this model, we assigned independent inverse-Wishart($2, I_2$) priors to $G_1$ and $G_2$. Second, we fit an uncorrelated model that assumed no correlation among any of the random effects — i.e., we assumed $\Gamma$ to be strictly diagonal. Here we assigned independent inverse-Gamma(0.001,0.001) priors to the random effect variances. The WAIC values were 143,462 for the fully correlated model, 143,482 for the partially correlated model, and 143,661 for the uncorrelated model, indicating superior predictive accuracy for the fully correlated model. We also computed the root mean square predictive errors.
Bayesian Zero-Inflated Negative Binomial Regression

(a) Random Intercepts for Binary Component

![Simulated vs Predicted Random Intercepts](image1)

(b) Random Slopes for Binary Component

![Simulated vs Predicted Random Slopes](image2)

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.

(RMSPEs) for each random effect vector, $\phi_{11}$, $\phi_{12}$, $\phi_{21}$, $\phi_{22}$, using the expression

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

where $\phi_{kik}$ denotes the simulated value of random effect $kk$ for subject $i$, and $\hat{\phi}_{kik}$ is the corresponding predicted value. Table S4 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 S5). These results suggest that ignoring even modest association among the 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 break down.
4 Analysis of Inpatient Length of Stay

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 days using a model analogous to equation (3.4). 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, a mean length of stay (LOS) of 1.24 days, and a five-number summary of (0, 0, 0, 1, 235). Figure S10 in the supplement provides a histogram of the counts, and Table S6 presents sample summary statistics. Prior distributions and MCMC specifications were identical to those for simulation 3. We ran the algorithm for a conservative 5000 iterations with a burn-in of 500. Trace plots and Geweke statistics suggested MCMC convergence with adequate mixing (Figure S11). A longer run of 50,000 iterations yielded nearly identical results (Figure S12 and Table S7).

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 LOS over time. This finding is consistent with recent VA efforts to reduce length of patient stays through improved outpatient services (Kaboli et al., 2012). Additionally, patients with full disability coverage had shorter inpatient stays. 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 longer inpatient stays. The random effect covariances showed modest heterogeneity across counties in both components of the model.

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 block groups with significantly negative effects (95% CrI < 0). VA medical centers are superimposed on the map. 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 longer stays 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, longer hospital stays 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>
<thead>
<tr>
<th>Model Component</th>
<th>Parameter</th>
<th>Variable</th>
<th>Posterior Mean (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary</td>
<td>$\beta_{10}$</td>
<td>Intercept</td>
<td>0.11 (−0.10, 0.33)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>Year</td>
<td>−0.07 (−0.09, −0.06)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{12}$</td>
<td>Age</td>
<td>0.00 (−0.002, 0.002)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{13}$</td>
<td>Male Gender</td>
<td>−0.01 (−0.12, 0.09)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{14}$</td>
<td>NHW Race</td>
<td>0.03 (−0.01, 0.07)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{15}$</td>
<td>Full Disability Coverage</td>
<td>−0.12 (−0.17, −0.08)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{16}$</td>
<td>Elixhauser Score</td>
<td>0.43 (0.41, 0.45)</td>
</tr>
<tr>
<td>Count</td>
<td>$\beta_{20}$</td>
<td>Intercept</td>
<td>0.86 (0.72, 1.00)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>Year</td>
<td>−0.04 (−0.05, −0.03)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{22}$</td>
<td>Age</td>
<td>0.003 (0.002, 0.005)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{23}$</td>
<td>Male Gender</td>
<td>0.05 (−0.01, 0.12)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{24}$</td>
<td>NHW Race</td>
<td>−0.04 (−0.07, −0.01)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{25}$</td>
<td>Full Disability Coverage</td>
<td>−0.04 (−0.07, −0.02)</td>
</tr>
<tr>
<td></td>
<td>$\beta_{26}$</td>
<td>Elixhauser Score</td>
<td>0.16 (0.15, 0.17)</td>
</tr>
<tr>
<td>$r$</td>
<td></td>
<td>Dispersion</td>
<td>0.77 (0.72, 0.83)</td>
</tr>
<tr>
<td>Random Effects</td>
<td>$\Gamma_{11}$</td>
<td>Var($\phi_{111}$)</td>
<td>0.04 (0.03, 0.05)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{12}$</td>
<td>Cov($\phi_{111}, \phi_{121}$)</td>
<td>0.01 (0.001, 0.01)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{13}$</td>
<td>Cov($\phi_{111}, \phi_{211}$)</td>
<td>0.01 (0.002, 0.02)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{14}$</td>
<td>Cov($\phi_{111}, \phi_{221}$)</td>
<td>0.01 (0.001, 0.01)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{22}$</td>
<td>Var($\phi_{112}$)</td>
<td>0.02 (0.01, 0.02)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{23}$</td>
<td>Cov($\phi_{112}, \phi_{211}$)</td>
<td>0.01 (0.001, 0.01)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{24}$</td>
<td>Cov($\phi_{112}, \phi_{221}$)</td>
<td>0.003 (0.001, 0.01)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{33}$</td>
<td>Var($\phi_{211}$)</td>
<td>0.06 (0.05, 0.09)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{34}$</td>
<td>Cov($\phi_{211}, \phi_{221}$)</td>
<td>0.01 (0.01, 0.02)</td>
</tr>
<tr>
<td></td>
<td>$\Gamma_{44}$</td>
<td>Var($\phi_{221}$)</td>
<td>0.02 (0.01, 0.02)</td>
</tr>
</tbody>
</table>

Table 4: Parameter estimates and 95% credible intervals (CrIs) for the spatiotemporal ZINB model in length of stay study. NHW: Non-Hispanic white.
Bayesian Zero-Inflated Negative Binomial Regression

Figure 3: County maps of predicted spatial random effects (top panel) and posterior significance (lower panel) for length of stay 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.

Table 5 presents the predicted marginal mean LOS, $E(y_{ij}) = \pi_{ij} \mu_{ij}$, for patients residing in three hypothetical counties in years 2011 and 2015, along with accompanying risk ratios. All three patients were from the reference covariate population. The first patient was assumed to reside in an “average” county in which spatial random effects were set to zero. The remaining two patients were from counties whose spatial random effects were, respectively, one standard deviation above and one standard deviation below average. 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 risk 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, while the risk 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 LOS 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.

Table 5: Mean LOS in days and corresponding risk 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.

<table>
<thead>
<tr>
<th>Mean LOS (Days)</th>
<th></th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average County</td>
<td>0.97 (0.85, 1.10)</td>
<td>0.71 (0.62, 0.81)</td>
</tr>
<tr>
<td>1 SD Above Average</td>
<td>1.36 (1.19, 1.54)</td>
<td>2.12 (1.82, 2.46)</td>
</tr>
<tr>
<td>1 SD Below Average</td>
<td>0.68 (0.59, 0.78)</td>
<td>0.21 (0.17, 0.25)</td>
</tr>
<tr>
<td>Risk Ratios</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average vs. Below-Average</td>
<td>1.42 (1.35, 1.48)</td>
<td>3.37 (3.05, 3.77)</td>
</tr>
<tr>
<td>Above-Average vs. Average</td>
<td>1.40 (1.34, 1.46)</td>
<td>3.00 (2.74, 3.31)</td>
</tr>
<tr>
<td>Above-Average vs. Below-Average</td>
<td>1.98 (1.83, 2.18)</td>
<td>10.17 (8.35, 12.46)</td>
</tr>
</tbody>
</table>

The space-time interaction is highlighted more prominently in Figures 4(a) and 4(b). Figure 4(a) presents the mean LOS for each county in 2011 and 2015, while Figure 4(b) displays the net change over time in expected LOS for each county. Estimates correspond to a patient in the reference covariate group. Approximately 12% of the counties had increasing trends over time. These counties 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 LOS. 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 long patient stays, the VA can help manage overhead costs while minimizing the burden imposed on both patients and hospital staff.

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
Bayesian Zero-Inflated Negative Binomial Regression

Figure 4: Panel (a): Predicted mean LOS in days by county. Panel (b): Change in mean LOS in days from 2011–2015. Both figures correspond to a patient in the reference covariate group. In Figure (b), darkest shade shows counties with net increase in LOS.

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” can be found at (http://www.url.edu/download/0000.zip). This supplement contains derivations of the full conditionals discussed in Section 2, additional tables and figures for the simulation studies presented in Section 3, and additional tables and figures for the length of stay study presented in Section 4.

**References**


Bayesian Zero-Inflated Negative Binomial Regression


U.S. Census Bureau (2014). “TIGER/Line Shapefiles.” Suitland, MD.


**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.