Bayesian Models for Spatial Count Data with Informative Finite Populations with Application to the American Community Survey

Kai Qu[1], Jonathan R. Bradley[1 2]

Abstract

The American Community Survey (ACS) is an ongoing program conducted by the US Census Bureau that publishes estimates of important demographic statistics over pre-specified administrative areas periodically. ACS provides spatially referenced count-valued outcomes that are often recorded/paired with the finite population. For example, the number of people below the poverty line and the total population for each county is estimated by ACS. One common assumption is that the distribution of the spatially referenced count-valued outcome given the finite population is binomial distributed. This conditionally specified model (i.e., the conditional distribution of the count-valued outcome given the finite population is specified as a binomial distribution) does not explicitly define the joint relationship between the count-valued outcome and the finite population. Thus, we consider a joint model for the count-valued outcome and the finite population. When the cross-dependence in our joint model can be leveraged to "improve spatial prediction" (i.e., obtain smaller mean squared prediction error), we say that the finite population is "informative." We model the count given the finite population as binomial distributed, the finite population as negative binomial distributed, and use the multivariate logit-beta (MLB) for prior distributions. This specification leads to closed form expressions of the full-conditional distributions that are straightforward to simulate from within a Gibbs sampler and do not require subjective tuning steps. We illustrate the proposed model through simulations and our motivating application of ACS poverty estimates at the county level. These empirical analyses show the benefits of using our proposed model over the more traditional conditionally specified binomial model.

[1]Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306-4330, jrbradley@fsu.edu

[2](to whom correspondence should be addressed) Department of Statistics, Florida State University, 117 N. Woodward Ave., Tallahassee, FL 32306-4330, jrbradley@fsu.edu

1 Introduction

The American Community Survey (ACS) is an ongoing program conducted by the US Census Bureau that collect demographic statistics from prespecified areas. The ACS often contains tabulations across categories by region (e.g., the number of males and females in a county). Data such as these, can be modeled with a binomial distribution provided that the finite population is observed at each region. It is fairly common to use a conditionally specified model and ignore the distribution of the finite population to perform inference on count-valued outcome. It often happens that the finite population is not a fixed deterministic quantity, but rather a realization of a random variable that may have a complicated joint relationship with the paired count-valued outcome. In this article, we model outcomes given the finite population as binomial with probability π and the finite population as negative binomial with probability p. Several studies indicate that the population of a region is directly related to the number of individuals below the poverty line (Rodgers (1984), Ahlburg (1996), McNicoll (1997), Scherr (2000), Sinding (2009), among others). It is unclear whether or not this direct relationship is simply due to the fact that a larger population naturally implies a larger mean number of count-valued observation, or if more intricate cross-dependence exists between the count-valued observation and the finite population. By "intricate" dependencies (and joint relationships) we mean that π and p may be correlated. If these correlations are present then one can leverage this dependence to improve spatial prediction. When cross-dependence between π and p can be leveraged to "improve spatial prediction" (i.e., obtain smaller mean squared prediction error), we say that the finite population is "informative."

As an example, consider Figure 1, where we plot five-year period estimates of the number of families below the poverty threshold along with the number of households (population). Here, we see that large and small values of poverty and population appear together. An immediate explanation is that the increase of poverty is due solely to a larger trial number (i.e., the ACS five year period estimate of the population). However, we consider two reasons for this pattern: (1) the

increase of poverty is due solely to a larger population; and (2) the increase of poverty is because both the population has increased *and* π and p are positively correlated. This perspective allows for two types of cross-variable dependence, where Item 1 refers to dependence in the data models, and Item 2 refers to dependence between π and p. Thus, in what follows, we develop a Bayesian statistical model that can be used to leverage both sources of dependence to improve the accuracy of predictions.

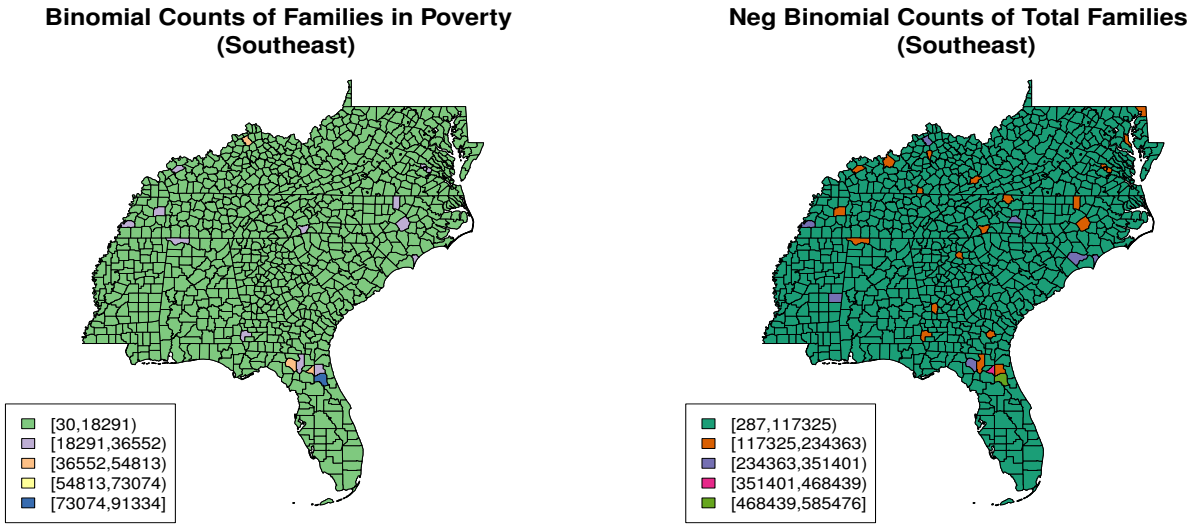

Figure 1: ACS poverty count data: the observed families counts in poverty (left) and the total families number (right).

There are several methods available to model correlated count data. For example, a bivariate Poisson model has been used to define a joint model and account for the dependence between outcomes and finite populations (see Fisher et al., 1943; Terza et al., 1990; Comulada and Weiss, 2007, among others). However, the Poisson distribution is not robust to highly heterogeneous populations (see Fisher et al. (1943), De Oliveira (2013) for further comments), especially when there is significant overdispersion in the outcome. The inflated bivariate Poisson model (Karlis et al., 2005) can accommodate moderate overdispersion by mixing an inflation component to define the joint probability density function. However, maximum likelihood estimation (MLE) through

the expectation maximization (EM) algorithm can be computationally burdensome. To alleviate the overdispersion issue, the negative binomial distribution is more appropriate because it has one more parameter to mitigate issues with heterogeneity among populations (e.g., see Lawless, 1987; Gardner et al., 1995; Hilbe, 2011, among others). Furthermore, quite often inference on π is of primary interest, and bivariate Poisson models are not explicitly parameterized to model π. However, bivariate Poisson models are the predominant choice used to model two dependent count-valued spatial variables. Thus, our proposed model offers an exciting new tool for this literature.

We use a model similar to latent Gaussian process (LGP) model, which has been widely used to model non-Gaussian data and can be used to do statistical inference and account for uncertainty (see Gelfand and Schliep (2016), Gelman et al. (2013), Cressie and Wikle (2011), among others). Specifically, the count-valued outcomes given the finite populations are modeled with a binomial distribution and finite populations are modeled with a negative binomial distribution at each region. One could assume latent random effects are Gaussian, however, the LGP is not always realistic (De Oliveira, 2013). Furthermore, many non-Gaussian data lead to full-conditional distributions that do not have a closed-form analytical solution. Therefore, the LGP can lead to computational inefficiencies when it requires subjective tuning of Markov chain Monte Carlo (MCMC) algorithms, and can subsequently have convergence challenges especially in high dimensional correlated settings (Garthwaite et al., 2010). One option available to avoid these issues is Polya-Gamma augmentation, which has been used for multinomial data. This model has a conjugate full-conditional distribution where the logit of the proportions can be expressed as a multivariate normal distribution after marginalizing across variances that follow a Polya-Gamma distribution (Polson and Scott (2011)). This method can also suffer from some computation difficulties because an iterative accept/reject sampling algorithm is needed to generate Polya-gamma random variables within the MCMC algorithm.

Another alternative available in the literature is to drop the Gaussian assumption and replace

it with a multivariate logit-beta distribution (MLB) (Gao and Bradley, 2019; Bradley et al., 2019, 2020). The MLB distribution is derived through the linear transformation of univariate independent logit-beta random variables. This MLB distribution is a reasonable alternative because it can be specified arbitrarily close to the default choice of a Gaussian distribution. The full-conditional distribution associated with this distribution are conditional multivariate logit-beta distributions and there are straightforward MCMC sampling schemes that do not require subjective Metroplis-Hastings tuning steps. These tuning steps can be avoided by other algorithms such as Hamiltonian Markov chain Monte Carlo (HMC; Neal, 2011) and integrated nested Laplace approximations (INLA; Rue et al., 2009). However, these models are limited to small parameter space settings, which is not the case for our motivating ACS dataset. The MLB has been used to define a prior on the adjacency matrix in a conditional autoregressive model (Gao and Bradley, 2019), to model spatial-temporal multivariate data (Bradley et al., 2019), and to model multivariate binomial and negative binomial data (Bradley et al., 2020). However, a latent MLB model that jointly models a binomial spatial dataset and a negative binomial spatial dataset does not exist.

We propose a Bayesian hierarchical joint model for spatially referenced binomial counts with informative finite populations. We incorporate dependence directly into the data hierarchically by assuming the finite population in the binomial outcome is distributed as negative binomial. We also incorporate cross-dependence between the (logit) of the probability of outcome and the (logit) of the probability of a trial through a shared spatial basis function expansion (e.g., see Cressie and Wikle, 2015, for a standard reference). We also model within data type (i.e., binomial and negative binomial data) spatial dependence with another spatial basis function expansion. All random and fixed effects are assumed to follow a MLB distribution.

We now outline the remainder of the paper. In Section 2, we introduce our proposed model. A simulation study is given in Section 3, and an analysis of ACS county-level poverty estimates in the Southeast US is given in Section 4. In each of our empirical analyses, we compare to a conditionally specified model, which is the primary competitor for modeling spatially referenced

binomial data. We end this paper with a discussion in Section 5. All technical details and reviews are provided in the appendices.

2 Methodology

We assume spatial binomial data are observed at n regions. Let $\boldsymbol{y}_\pi \equiv (y_{\pi,i} : i = 1,...,n)'$, and $\boldsymbol{y}_p \equiv (y_{p,i} : i = 1,...,n)'$ be n - dimensional vectors consisting of spatial binomial and negative binomial random counts respectively. These vectors are paired, that is, the i-th binomial count $y_{\pi,i}$ is paired with the i-th total $y_{p,i}$. For example, $y_{\pi,i}$ might consist of the number of families below the poverty threshold out of a total of $y_{p,i}$ families in county i. The probabilities associated with the binomial and negative binomial distributions are denoted with $\pi_i \in (0,1)$ and $p_i \in (0,1)$, respectively. Recall, that one of the developments in this article is that we allow π_i and p_i to be correlated. The primary goal is to accurately predict the proportions associated with the binomial distributed observations (i.e., π_i).

We use the MLB distribution for our prior distribution specification. A review of the MLB distribution can be found in Appendix A in the Supplemental Materials. Additional details on the Gibbs sampler are provided in Appendices B and C in the Supplemental Materials.

2.1 Binomial and Negative Binomial Data Models

Count-valued observations $y_{\pi,i}$, and $y_{p,i}$ are assumed to be binomial and negative binomial distribution as follows:

$$\begin{aligned} y_{\pi,i}|\pi_i, y_{p,i} &\sim \mathscr{BN}(\pi_i, y_{p,i}) \\ y_{p,i}|p_i, d_i &\sim \mathscr{NB}(p_i, d_i) \qquad\qquad i = 1,...,n, \end{aligned} \tag{1}$$

where "$\mathscr{BN}$" and "$\mathscr{NB}$" are shorthands for a binomial and negative binomial distribution

and d_i is the dispersion parameter. Here, $logit(\pi_i) \equiv \{\log(\pi_i/(1-\pi_i)\} \equiv \nu_{\pi,i}$ and $logit(p_i) \equiv \{\log(p_i/(1-p_i)\} \equiv \nu_{p,i}$ are considered unknown "processes" to estimate. Equation (1) can be rewritten as (Bradley et al., 2019):

$$
\begin{aligned}
f(y_{\pi,i}|\nu_{\pi,i}, y_{p,i}) &= \binom{y_{p,i}}{y_{\pi,i}} \exp[y_{\pi,i}\nu_{\pi,i} - y_{p,i}\log\{1+\exp(\nu_{\pi,i})\}] \\
f(y_{p,i}|\nu_{p,i}, d_i) &= \binom{y_{p,i}+d_i+1}{y_{p,i}} \exp[y_{p,i}\nu_{p,i} - (y_{p,i}+d_i)\log\{1+\exp(\nu_{p,i})\}] \\
& i = 1, ..., n,
\end{aligned}
\tag{2}
$$

where f denotes a probability density function/ mass function (pdf/pmf). It will be useful to define the n-dimensional vectors $\boldsymbol{y}_\pi \equiv (y_{\pi,1}, ..., y_{\pi,n})'$, $\boldsymbol{y}_p \equiv (y_{p,1}, ..., y_{p,n})'$, $\boldsymbol{\nu}_\pi \equiv (\nu_{\pi,1}, ..., \nu_{\pi,n})'$, $\boldsymbol{\nu}_p \equiv (\nu_{p,1}, ..., \nu_{p,n})'$, and $\boldsymbol{d} \equiv (d_1, ..., d_n)'$. The presence of $y_{p,i}$ in $y_{\pi,i}|\pi_i, y_{p,i}$ induces cross-data-type dependence, which again, is an important component of our model, but this particular implementation has been used by several others (e.g., see for example, Dodd and Dorazio (2004), Kéry et al. (2005), Wu et al. (2015), and Meehan et al. (2017)). When $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$ are assumed independent then the cross-dependence between $\boldsymbol{y}_\pi$ and $\boldsymbol{y}_p$ is determined by the conditional distribution $f(y_{\pi,i}|\nu_{\pi,i}, y_{p,i})$. Hence, when $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$ are assumed independent we call this model the aforementioned conditionally specified model. In what follows, we consider the case where $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$ are dependent.

2.2 Process Models

The unobserved processes $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$ are jointly modeled as

$$
\begin{aligned}
\boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \boldsymbol{\Psi}_{\pi 1}\boldsymbol{\eta}_\pi + \boldsymbol{\Psi}_{\pi 2}\boldsymbol{\eta} + \boldsymbol{\xi}_\pi \\
\boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\Psi}_{p1}\boldsymbol{\eta}_p + \boldsymbol{\Psi}_{p2}\boldsymbol{\eta} + \boldsymbol{\xi}_p,
\end{aligned}
\tag{3}
$$

where $\mathbf{X}_j$ is a known $n \times m$ matrix of covariates, and $\boldsymbol{\beta}_j \in \mathscr{R}^m$ is the associated unknown m-dimensional parameter vector with j=π, p. The m-dimensional within data type fixed effect $\boldsymbol{\beta}_j$, the r-dimensional within data type random effects $\boldsymbol{\eta}_j$, the n-dimensional within data type random effects $\boldsymbol{\xi}_j$, and the r-dimensional between data type random effect of $\boldsymbol{\eta}$ are modeled using MLB distribution. We provide a review of the MLB distribution in Appendix A in the Supplemental Materials. Recall from the Introduction that we use the MLB as a prior because it leads to a Gibbs sampler with full-conditional distributions that one can directly sample from (i.e., no subjective Metroplis-Hastings tuning steps are required). As such, MCMC is considerably easier to implement with the MLB prior than with a more traditional Gaussian prior. Additionally, the MLB prior has the more traditional Gaussian prior as a special limiting case (Bradley et al., 2019), and consequently, can more flexibly model the tails of the distribution.

Here the term $\boldsymbol{\Psi}_k \boldsymbol{\eta}_j$ represents "small-scale" variability and $\boldsymbol{\xi}_j$ represents "fine-scale" variability from a white Gaussian noise process with $k = \pi 1, \pi 2, p1, p2$ and $j = \pi, p$. The term $\boldsymbol{\Psi}_k \boldsymbol{\eta}$ includes cross-data-type dependence between the logit of probabilities, since the covariance between $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$ equals $\boldsymbol{\Psi}_{\pi 2} cov(\boldsymbol{\eta}) \boldsymbol{\Psi}'_{p2}$, which is not necessarily equal to a zero matrix. Recall that allowing for such cross-covariances is a key motivating feature of our model, and when non-zero $\boldsymbol{\Psi}_{\pi 2} cov(\boldsymbol{\eta}) \boldsymbol{\Psi}'_{p2}$ improves prediction we say that the finite population is informative.

For areal data the Moran's I basis function is a useful specification of $\boldsymbol{\Psi}_k$ (see more discussion in Griffith, 2000, 2002, 2004; Tiefelsdorf and Griffith, 2007; Hughes and Haran, 2013; Porter et al., 2014; Bradley et al., 2016). The Moran's I operator is defined as

$$\mathbf{MI}(\mathbf{X}_j, \mathbf{A}) \equiv (\mathbf{I}_n - \mathbf{X}_j(\mathbf{X}'_j\mathbf{X}_j)^{-1}\mathbf{X}'_j)\mathbf{A}(\mathbf{I}_n - \mathbf{X}_j(\mathbf{X}'_j\mathbf{X}_j)^{-1}\mathbf{X}'_j), \tag{4}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix, and $\mathbf{A}$ is the $n \times n$ adjacency matrix associated with the edges formed by all the areal units in the study. Let $\boldsymbol{\Psi}_M \boldsymbol{\Lambda}_M \boldsymbol{\Psi}'_M$ be the spectral decomposition of the $n \times n$ matrix $\mathbf{MI}(\mathbf{X}_j, \mathbf{A})$. The $n \times r$ matrix $\boldsymbol{\Psi}_k$ contains the first r columns of $\boldsymbol{\Psi}_M$. Moran's I basis func-

tions represent a basis for a spatially weighted orthogonal column space of $\mathbf{X}_j$ and consequently this specification of basis functions makes the fixed and random effects not confounded. Additionally, Moran's I basis functions can facilitate low-rank modeling (i.e. choosing $r \ll n$) for high dimensional spatial data (Hughes and Haran (2013), Porter et al. (2014)).

If the regions are equally spaced and "small" then one could treat areal data as point-referenced, and use for example, a radial basis function (Wikle (2002)). In Section 3, we use a thin-plate spline (Wahba (1990)) as below:

$$\psi_{i,b} = \|\boldsymbol{s}_i - \boldsymbol{c}_b\|^2 \log(\|\boldsymbol{s}_i - \boldsymbol{c}_b\|), \tag{5}$$

where $\boldsymbol{s}_i$ is the i-th data points and $\boldsymbol{c}_b$ is a pre-defined knot points. Then $\boldsymbol{\psi}_i = (\psi_{i,1}, \ldots, \psi_{i,r})'$ and the $n \times r$ matrix $\boldsymbol{\Psi}_k = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_r)'$ are the associated basis vector and basis matrix respectively.

The ability to specify any class of basis functions makes our proposed model adaptable to several given correlation specifications. Specifically, suppose you are interested in a covariance matrix of a given form $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}'$, where $\mathbf{V}$ is a $n \times n$ Cholesky decomposition of $\boldsymbol{\Sigma}$. Then, for example, $\boldsymbol{\Phi}_k = \mathbf{V}\mathbf{V}_{\eta,j}^{-1}$ produces $cov(\boldsymbol{\Psi}_k \boldsymbol{\eta}_j) = \mathbf{V}\mathbf{V}_{\eta,j}^{-1} cov(\boldsymbol{\eta}_j)\mathbf{V}_{\eta,j}^{-1\prime}\mathbf{V}' = \boldsymbol{\Sigma}$, where $\mathbf{V}_{\eta,j}$ is the Cholesky decomposition of $cov(\boldsymbol{\eta}_j)$. Thus, one can develop versions of our model that imply a desired given correlation structure.

3 A Simulation Study

In this simulation, we generate binomial and negative binomial data, where the true proportions are specified to be a known function. We fit our model (which is different from the model generating the data) and we verify that our proposed model can recover the true unknown proportion. The following model is used to generate pairs of binomial and negative binomial counts over a fine-one

dimensional grid from zero to 2π (i.e., the spatial domain D=$\{0, \frac{2\pi}{100}, ..., 2\pi\}$):

$$
\begin{aligned}
&y_{\pi,i}|\pi_i, y_{p,i} \sim \mathscr{BN}(\pi_i, y_{p,i}) \\
&y_{p,i}|p_i, d_i \sim \mathscr{NB}(p_i, d_i) \\
&d_i \sim \mathscr{G}(1700, 0.125) \\
&\nu_{\pi,i}|\nu_{p,i} = \begin{cases} -4.64 + 2.32\nu_{p,i} & s_i \in [0, \pi) \\ -7.54 + s_i & s_i \in [\pi, 2\pi] \end{cases} \\
&\nu_{p,i} = 0.01 + 4.00\sin(s_i) \qquad i = 1, ..., 100, s_i \in D,
\end{aligned}
\tag{6}
$$

where "$\mathscr{G}(\alpha, \kappa)$" stands for "Gamma" distribution with shape $\alpha > 0$, and scale $\kappa > 0$. We interpret D=$\{0, \frac{2\pi}{100}, ..., 2\pi\}$ as a fine-one dimensional grid over 0 to 2π (i.e., in steps of 1/100), however ultimately, the length of the grid is an arbitrary choice that serves as illustration. The values of the slope and intercept in $\nu_{\pi,i}$, and the values 0.01 and 4.00 in $\nu_{p,i}$ are chosen to define the range of values of $\nu_{\pi,i}$ and $\nu_{p,i}$. The $logit(\mathbf{p}) \equiv \{\log(\mathbf{p}/(1-\mathbf{p})\} \equiv \boldsymbol{\nu}_p$ is a continuous, symmetric and smooth function with the approximate range of [-4.00, 4.00]. The $logit(\boldsymbol{\pi}) \equiv \{\log(\boldsymbol{\pi}/(1-\boldsymbol{\pi})\} \equiv \boldsymbol{\nu}_\pi$ is a non-continuous, asymmetrical function of $\nu_{p,i}$ with the approximate range of [-4.75, 4.25]. However, there are several values that would give the same responses, and thus, our choices are somewhat subjective. To add complexity to the true function, we include a change-point to $\nu_{\pi,i}|\nu_{p,i}$ at π, where we switch from a linear function of $\nu_{p,i}$ to a linear function of s_i. Also, the presence of $\nu_{p,i}$ in the definition of $\nu_{\pi,i}$ suggests a functional relationship between $\nu_{\pi,i}$ and $\nu_{p,i}$. See simulated dataset and the true proportions in Figure 2. To see if our method was robust to the relative magnitude of the negative binomial counts, we specified the prior on d_i to produce negative binomial counts that could covered a wide range of magnitudes. The simulated negative binomial count quantiles (25%, 50% and 75%) based on our choice of prior on d_i are: 15, 215, 3600. We interpret the quantiles 5, 215, 3600 as small to large count-values.

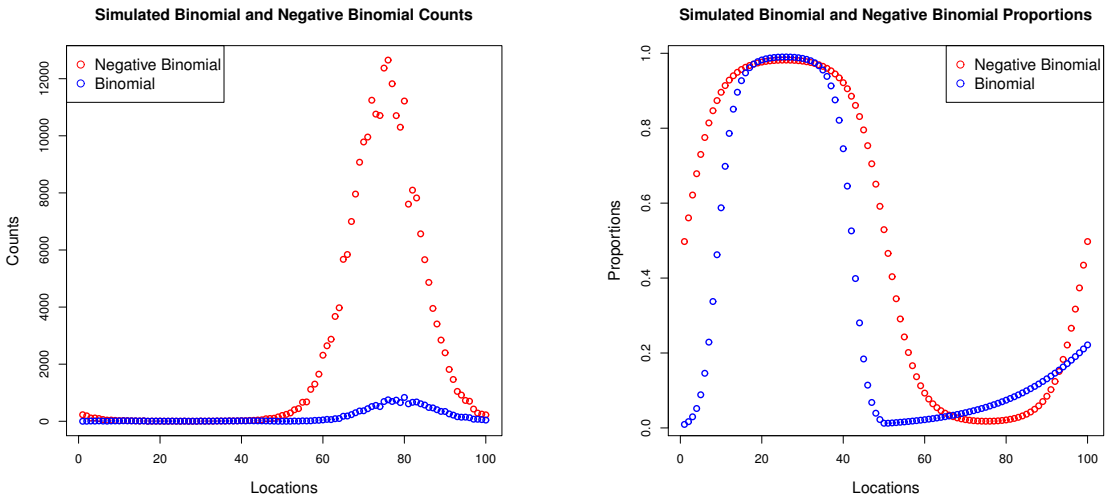

Figure 2: Simulated binomial and negative binomial counts and true proportions.

We use our proposed model to predict the unknown function in the right panel of Figure 2 (using the data displayed in the left panel). To use our model, one has to specify the basis matrices, covariates and several hyperparameters. We assume an intercept only model for $\nu_{\pi,i}$ and $\nu_{p,i}$ respectively. We consider two choices for basis matrices $\mathbf{\Psi}_{\pi j}$ and $\mathbf{\Psi}_{pj}$: j=1,2. The first is setting the elements equal to evaluations of thin-plate splines with r=25 knots. This use of thin-plate splines results in a 100×25 matrix that we denote with $\mathbf{\Psi}^*$. The second choice is to set $\mathbf{\Psi}_{kj}$ equal to a 100×25 matrix of zeros; k=π, p, j=1,2.

We consider six special cases of the model in (3):

- Joint model with between-type and within-type (JBW) random effects:

$$\begin{aligned} \boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \mathbf{\Psi}^* \boldsymbol{\eta}_\pi + \mathbf{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_\pi \\ \boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \mathbf{\Psi}^* \boldsymbol{\eta}_p + \mathbf{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_p, \end{aligned} \tag{7}$$

Here we set $\mathbf{\Psi}_{\pi 1}$= $\mathbf{\Psi}_{p1}$= $\mathbf{\Psi}_{\pi 2}$= $\mathbf{\Psi}_{p2}$= $\mathbf{\Psi}^*$.

- Joint model with between-type (JB) random effects:

$$\begin{aligned} \boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_\pi \\ \boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_p, \end{aligned} \tag{8}$$

  Here we set $\boldsymbol{\Psi}_{\pi 1}$= $\boldsymbol{\Psi}_{p1}$=$\mathbf{0}_{100,25}$ and $\boldsymbol{\Psi}_{\pi 2}$= $\boldsymbol{\Psi}_{p2}$= $\boldsymbol{\Psi}^*$, where $\mathbf{0}_{100,25}$ is a 100×25 matrix of zeros.

- Joint model with between-type random effects and within-type random effects for ν_p (JB-WNEG):

$$\begin{aligned} \boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_\pi \\ \boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\Psi}^* \boldsymbol{\eta}_p + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_p, \end{aligned} \tag{9}$$

  Here we set $\boldsymbol{\Psi}_{\pi 1}$=$\mathbf{0}_{100,25}$ and = $\boldsymbol{\Psi}_{p1}$=$\boldsymbol{\Psi}_{\pi 2}$= $\boldsymbol{\Psi}_{p2}$= $\boldsymbol{\Psi}^*$.

- Joint model with between-type random effects and within-type random effects for ν_π (JBW-BIN):

$$\begin{aligned} \boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \boldsymbol{\Psi}^* \boldsymbol{\eta}_\pi + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_\pi \\ \boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\Psi}^* \boldsymbol{\eta} + \boldsymbol{\xi}_p, \end{aligned} \tag{10}$$

  Here we set $\boldsymbol{\Psi}_{p1}$=$\mathbf{0}_{100,25}$ and = $\boldsymbol{\Psi}_{\pi 1}$=$\boldsymbol{\Psi}_{\pi 2}$= $\boldsymbol{\Psi}_{p2}$= $\boldsymbol{\Psi}^*$.

- Conditionally Specified (CS) with only within-type random effects:

$$\begin{aligned} \boldsymbol{\nu}_\pi &= \mathbf{X}_\pi \boldsymbol{\beta}_\pi + \boldsymbol{\Psi}^* \boldsymbol{\eta}_\pi + \boldsymbol{\xi}_\pi \\ \boldsymbol{\nu}_p &= \mathbf{X}_p \boldsymbol{\beta}_p + \boldsymbol{\Psi}^* \boldsymbol{\eta}_p + \boldsymbol{\xi}_p, \end{aligned} \tag{11}$$

  Here we set $\boldsymbol{\Psi}_{p1}$=$\boldsymbol{\Psi}_{p2}$=$\mathbf{0}_{100,25}$ and $\boldsymbol{\Psi}_{\pi 1}$= $\boldsymbol{\Psi}_{\pi 2}$= $\boldsymbol{\Psi}^*$.Here, there is no cross-dependence assumed between $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$, and hence, the cross-dependence between $\{y_{\pi,i}\}$ and $\{y_{p,i}\}$ is completely determined by the conditional distribution of $y_{\pi,i}|\pi_i, y_{p,i}$.

These six models are evaluated through the credible interval coverage of $\{\pi_i\}$, and the mean

squared predicted errors (MSPE). Let

$$\text{MSPE} = \sum_i \left\{ \pi_i - E(\pi_i | \mathbf{y}_\pi, \mathbf{y}_p) \right\}^2 / n.$$

The mean squared prediction error is the average squared difference between the predicted value and unobserved truth used to generate the data. By credible interval coverage we are specifically computing the proportion of times (over simulated replicates of the data) the credible intervals for $\{\pi_i\}$ correctly contain the true proportion. The CS model produces the aforementioned conditionally specified model for $\{y_{\pi,i}\}$, and the five remaining models are all considered different "joint models." Thus, since the conditionally specified model is our primary competitor, we are particularly interested in observing whether or not any of the five joint models outperform the CS model in terms of coverage and/or MSPE.

The Gibbs sampler was implemented with 2,000 iterations, and has a burn-in of 1,000 iterations. The MCMC chain convergence is checked based on trace plot, and no convergence issues were found. Directly sampling from full-conditional distributions allows partially allowed for a shorter MCMC chain than more traditional approaches, which is one of the advantages of our algorithm. The hyperparameters of the MLB distribution, in practice, are chosen by minimizing a criterion. In this illustration we set them as ε=0.05 and ρ=0.965 for binomial model; ε=0.1 and ρ=0.99 for negative binomial model; and σ=0.000001 for both binomial and negative binomial model (see these hyperparameters definition detailed in Appendix B in the Supplemental Materials). The comparison of these 6 models are repeated with 30 replicate of datasets $\{\boldsymbol{y}_\pi, \boldsymbol{y}_p\}$.

Table 1, Table 2 and the left panel of Figure 3 show that all six special cases yield reasonable predictions (i.e., close to the true and small MSPE) and credible interval coverage for $\boldsymbol{\nu}_\pi$. In this simulation, JB performs very poorly and JBWNEG tends to perform the best in terms of MSPE. In terms of coverage (i.e., the proportion out of 30 replicates that the truth is contained within the 95% credible interval), all methods appear to have similar (high) coverage for binomial data, with

Table 1: Mean squared predicted errors by rank and type for simulated data, average over 100 locations and 30 replicates. The rank is the basis matrix column number. Minimum MSPE by rank and type are bolded. The 95% confidence intervals for the MSPE over the 30 replicate (i.e., mean MSPE plus or minus 1.96 times the standard deviation) are in parentheses.

Rank	JB	JBWNEG	JBWBIN	JBW	CS
5	1.4213	1.0129	**0.9277**	0.9674	0.9711
	(1.3542, 1.4883)	(0.9119, 1.1139)	(0.8037, 1.0518)	(0.8356, 1.0993)	(0.8463, 1.0959)
10	1.4011	**0.9432**	1.0140	1.1004	1.0445
	(1.3026,1.4997)	(0.8452, 1.0412)	(0.8874, 1.1405)	(0.9599, 1.2409)	(0.9234, 1.1656)
15	1.4237	**1.0541**	1.2337	1.3777	1.2868
	(1.3669, 1.4804)	(0.9155, 1.1928)	(1.0442, 1.4232)	(1.1660, 1.5894)	(1.0934, 1.4802)
25	1.3555	**1.0404**	1.3012	1.4077	1.3159
	(1.2691, 1.4420)	(0.8991, 1.1817)	(1.1196, 1.4829)	(1.2113, 1.6041)	(1.1343, 1.4975)
50	1.4028	**1.1856**	1.4196	1.5164	1.4260
	(1.2925, 1.5131)	(0.9859, 1.3853)	(1.1741, 1.6651)	(1.2607, 1.7721)	(1.1780, 1.6740)

Table 2: Estimated credible interval coverage by rank for the simulated data. Maximum coverage by rank and type are bolded. The coverage of π_i over all i and averaged over the 30 replicates. The 95% confidence intervals for the proportion of credible interval completely covering the true proportion over the 30 replicate are in parentheses.

Rank	JB	JBWNEG	JBWBIN	JBW	CS
5	0.9933	0.9913	**0.9993**	**0.9993**	0.9987
	(0.9899, 0.9968)	(0.9869, 0.9957)	(0.9984, 1.0000)	(0.9984, 1.0000)	(0.9966, 1.0000)
10	0.9940	0.9953	0.9987	0.9977	**0.9990**
	(0.9912, 0.9968)	(0.9931, 0.9976)	(0.9974, 0.9999)	(0.9959, 0.9995)	(0.9979, 1.0000)
15	0.9957	0.9966	**0.9973**	**0.9973**	0.9963
	(0.9932, 0.9981)	(0.9942, 0.9990)	(0.9957, 0.9989)	(0.9955, 0.9992)	(0.9937, 0.9989)
25	0.9960	0.9956	0.9970	0.9973	**0.9977**
	(0.9936, 0.9984)	(0.9934, 0.9979)	(0.9951, 0.9989)	(0.9955, 0.9992)	(0.9959, 0.9995)
50	0.9957	**0.9973**	0.9967	**0.9973**	**0.9973**
	(0.9932, 0.9981)	(0.9957, 0.9989)	(0.9949, 0.9984)	(0.9957, 0.9989)	(0.9957,0.9989)

JBWBIN, JBW and Uni tending to perform (marginally) the best.

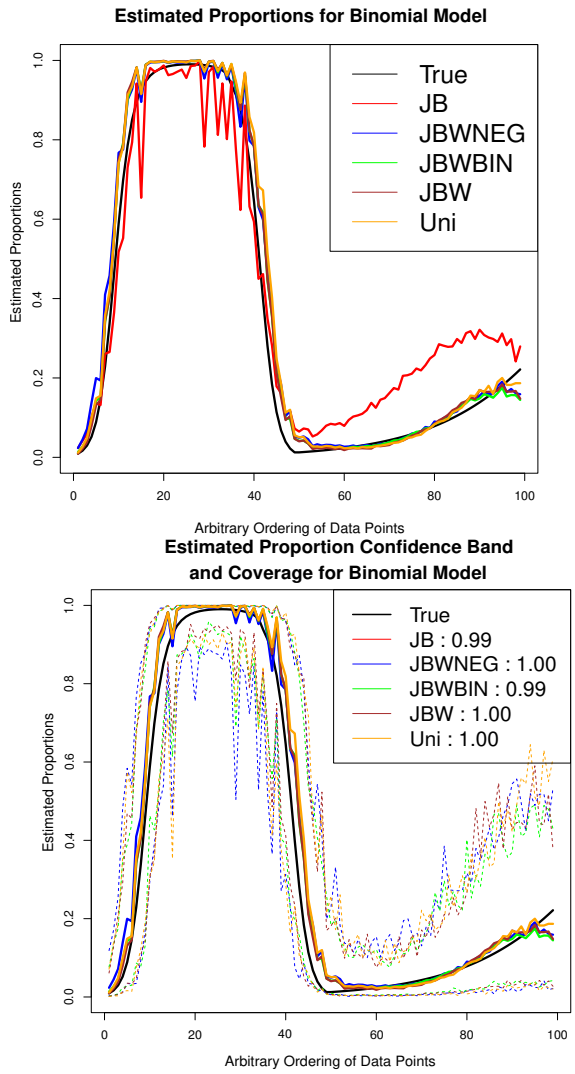

Figure 3: The first panel presents the true values and estimated values of $\{\pi_i\}$ by method with i on the x-axis. Similarly, the right panel presents the credible bands for $\{\pi_i\}$ by method.

4 Application: Analysis of ACS Period Estimates of Poverty Rates in Southeast United States

The number of families that fall below the poverty threshold can be modeled using a conditionally specified binomial distribution, however the population in a county may express spatial dependence and is observed with uncertainty. This suggests an opportunity to leverage possible cross-dependence between the probability of being below the poverty line and the mean of the population

to produce more accurate predictions of poverty rates.

The primary goal of this analysis is to predict the family poverty rate for the Southeast region of US at county level using the proposed Bayesian hierarchical model. The five-year period estimates (US Census Bureau, 2013-2017 American Community Survey 5-Year Cumulative Estimates) of incidence of poverty and the population are available at 948 counties in the Southeast US region, which we focus our analysis on. The left and right panels in Figure 1 show the incidence of poverty and the population by county in the Southeast US. We subjectively chose roughly 95% of the observations to be training. That is, we randomly select 898 out of 948 counties to be the training data set and the 50 counties are treated as a holdout to the model fit.

The Moran's I (MI) basis functions are used, and a second order of adjacency matrix is chosen using cross validation. The MSPE (between the sample proportion and the model estimated proportion) using the first and second order adjacency matrix are 0.01351 and 0.01308, respectively. Consequently, we use the second order adjacency matrix. For covariates, we use the logit transformation of the sample proportion of each race, age, education and single households at the county level, which are all five-year ACS period estimates that are common factors in socioeconomic research (see Scherr (2000), Sinding (2009), among others). These same covariates are used in the model for $\boldsymbol{\nu}_\pi$ and $\boldsymbol{\nu}_p$. The majority of predictors are significant (i.e., zero is not in the credible interval) across both binomial and negative binomial models. The model is evaluated based on the coverage of the credible intervals (i.e., whether or not the credible interval for $y_{p,i}$ and $y_{\pi,i}$ contains the hold-out observations) and mean squared predicted errors (MSPE) of the holdout data. The hyper-parameters are chosen to minimize (training) MSPE, and are ε=1.5 and ρ=0.995 for the binomial model; and ε=0.1 and ρ=0.99 for the negative binomial model (See Table 3 and Figure 4 for MSPE versus prior specification). In general, the estimate of these hyperparameters change for each dataset. The Gibbs sampler is implemented with runs of 5,000 iterations, and has a burn-in of 3,000 iterations. No convergence issues were found. Specifically, trace plots (see Figure 6) were informally checked and the associated Gelman-Rubin's diagnostics (Gelman and

Table 3: MSPE by different priors specifications for ACS binomial data mixed effects model.

$Prior_{index}$	ε_{Neg}	ρ_{Neg}	ε_{Bin}	ρ_{Bin}	MSPE
1	0.2500	1.00	3.25	2.250	1.2118
2	0.2150	1.00	2.75	1.750	0.5221
3	0.1950	1.00	2.50	1.500	0.2645
4	0.1750	1.00	2.25	1.250	0.0891
5	0.1500	1.00	2.00	1.000	0.0134
6	**0.1000**	**0.99**	**1.50**	**0.995**	**0.0129**
7	0.0750	0.85	1.00	0.875	0.0359
8	0.0500	0.65	0.80	0.600	0.3444
9	0.0250	0.45	0.60	0.400	1.0445
10	0.0125	0.25	0.40	0.200	3.0420
11	0.0100	0.05	0.20	0.050	9.9736

Rubin, 1992) were found to be between 1.0 and 1.1, respectively.

In Figure 5 and Table 4, we see that all versions of our model (we dropped JB due to its poor performance in the simulation) produce very small out-of-sample (i.e., on the order of 0.13 to 0.16) prediction errors. For ranks above 750, we roughly reach the nominal 95% coverage. The marginally best performing model is JBW at rank 948. The CS model is fairly competitive in terms of MSPE. However, in terms of coverage each joint model is preferable than the CS model, with JBWBIN and JBW performing comparably. Hence in terms of coverage and (marginally) MSPE it appears that the finite population is informative.

5 Discussion

In this article, we develop a Bayesian approach to jointly model a spatially referenced count-valued outcome and an associated finite population. We are motivated by estimates made available by ACS. We use the MLB latent process (Bradley et al. (2020), Gao and Bradley (2019)), which allows us to draw directly from the full-conditional distributions and makes Gibbs sampling more computationally efficient than other choices. The primary motivation for the proposed methodology is that complex spatial dependencies can exist within and between the two probabilities

Table 4: Posterior summary by model and rank for ACS binomial testing data (50).

(Estimated credible interval coverage)

Rank	JBWNEG	JBWBIN	JBW	CS
948	0.9600	**0.9800**	**0.9800**	0.8400
750	0.8600	**0.9400**	0.9200	0.7000
500	**0.8400**	0.8200	0.8200	0.6400
250	0.7000	**0.7800**	0.7600	0.5000
100	0.5000	**0.6200**	**0.6200**	0.4600
50	0.4200	0.5000	0.5600	0.3400

(Mean squared predicted errors)

Rank	JBWNEG	JBWBIN	JBW	CS
948	0.1338	0.1343	**0.1329**	0.1347
750	0.1402	0.1488	0.1395	**0.1384**
500	0.1496	0.1590	0.1536	**0.1405**
250	0.1615	0.1676	0.1655	**0.1364**
100	0.1475	0.1497	0.1477	**0.1362**
50	0.1363	0.1387	0.1395	**0.1362**

associated with binomial and negative binomial data. In our proposed joint model, these complex spatial dependencies are explicitly defined with basis functions. Furthermore, the specification of low-rank basis functions provides the flexibility of incorporating dimension reduction.

We tested versions of proposed model via simulations. The simulated studies show that the joint model based on the MLB latent processes achieves reasonable credible interval coverage and accurate predicted proportions. The motivating analysis of ACS poverty data in the Southeast US illustrates that the proposed Bayesian hierarchical joint model yields precise and accurate predictions of the probability that a family falls below the poverty threshold. Furthermore, there is evidence that the finite population is informative for predicting US poverty rates.

There are exciting avenues to extend the proposed joint model. First, it is natural to extend this

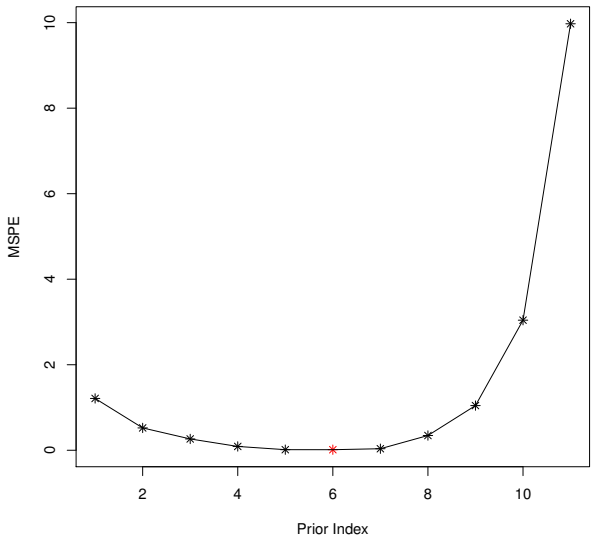

Figure 4: MSPE versus prior specification for ACS binomial data mixed effects model.

model to the spatio-temporal case, where the time indexed random effects are used to incorporate temporal dependencies. Additionally, the multiscale nature of time in ACS may be modeled using the techniques in the spatio-temporal change of support literature (e.g., see Bradley et al., 2015). There are limitations to our method that could also lead to future developments. In particular, the specification of MLB hyperparameters require small sensitivity studies. One could instead develop a hyperprior distribution for these quantities. Similarly, the choice of the rank r requires a sensitivity student, where one could instead place a prior distribution for certain classes of basis functions.

Acknowledgments

Jonathan R. Bradley's research was partially supported by the U.S. National Science Foundation (NSF) under NSF grant SES-1853099 and the National Institute of Health (NIH) under grant 1R03AG070669-01.

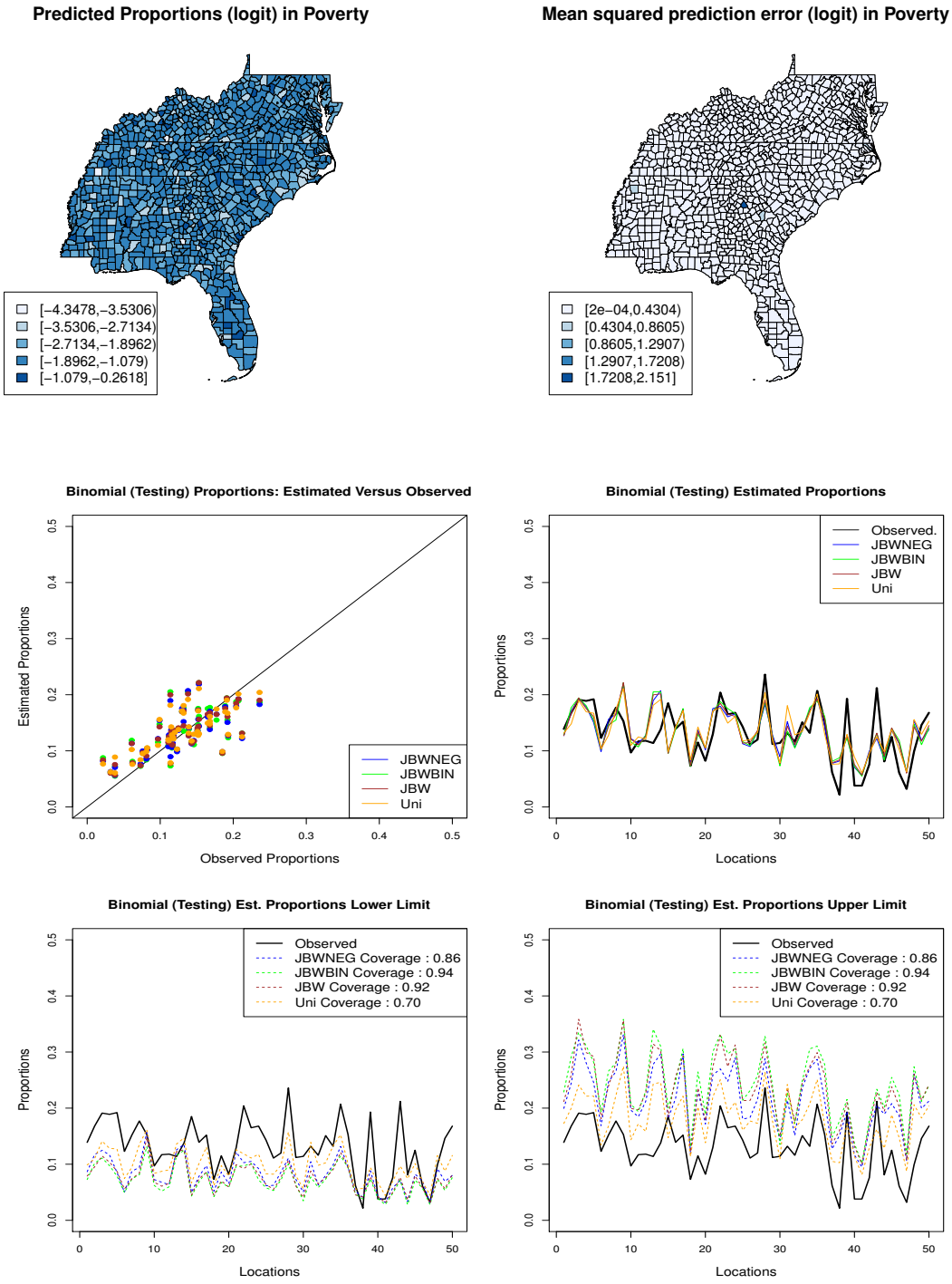

Figure 5: Testing (50) models for binomial data: the first row displays the predicted logit of poverty proportions and their associated MSPE (on the logit-scale for visualization purposes); the second row displays the estimated proportions versus the observed proportions; and the third row displays the 95% lower and upper credible limits for estimated proportions.

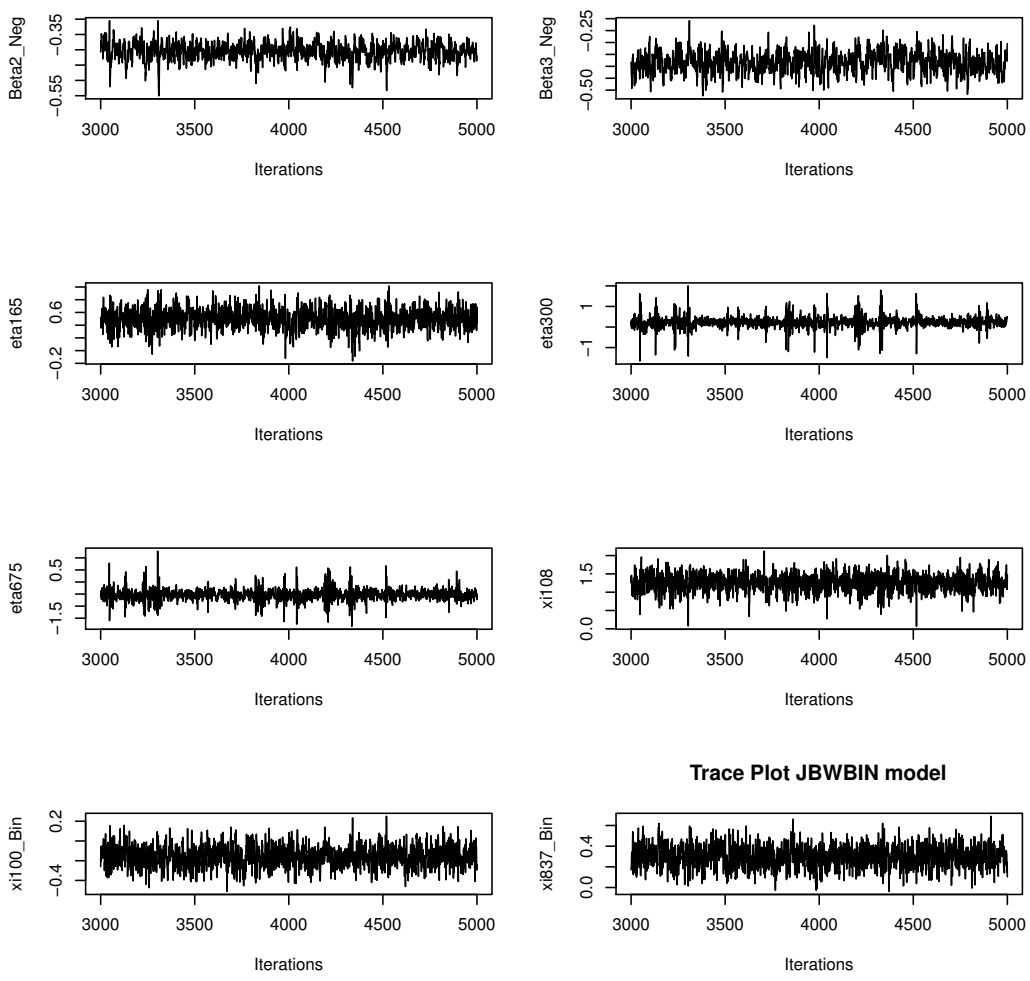

Figure 6: Trace Plots of JBWBIN model for ACS binomial training data (898).

References

Ahlburg, D. A. (1996). "Population growth and poverty." In *the impact of population growth on well-being in developing countries*, 219–258. Springer.

Bradley, J., Wikle, C., and Holan, S. (2015). "Spatio-temporal change of support with application to American Community Survey multi-year period estimates." *Stat*, 4, 255 – 270.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2020). "Bayesian Hierarchical Models with Conju-

gate Full-Conditional Distributions for Dependent Data from the Natural Exponential Family." *Journal of the American Statistical Association*, 115, 2037–2052.

Bradley, J. R., Wikle, C. K., and Holan, S. H. (2016). "Bayesian spatial change of support for count-valued survey data with application to the american community survey." *Journal of the American Statistical Association*, 111, 514, 472–487.

— (2019). "Spatio-temporal models for big multinomial data using the conditional multivariate logit-beta distribution." *Journal of Time Series Analysis*, 40, 3, 363–382.

Comulada, W. S. and Weiss, R. E. (2007). "On models for binomial data with random numbers of trials." *Biometrics*, 63, 2, 610–617.

Cressie, N. and Wikle, C. K. (2011). "Statistics for spatio-temporal data. Hoboken."

— (2015). *Statistics for spatio-temporal data*. John Wiley & Sons.

De Oliveira, V. (2013). "Hierarchical Poisson models for spatial count data." *Journal of Multivariate Analysis*, 122, 393–408.

Dodd, C. K. and Dorazio, R. M. (2004). "Using counts to simultaneously estimate abundance and detection probabilities in a salamander community." *Herpetologica*, 60, 4, 468–478.

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). "The relation between the number of species and the number of individuals in a random sample of an animal population." *The Journal of Animal Ecology*, 42–58.

Gao, H. and Bradley, J. R. (2019). "Bayesian analysis of areal data with unknown adjacencies using the stochastic edge mixed effects model." *Spatial Statistics*, 31, 100357.

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995). "Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models." *Psychological bulletin*, 118, 3, 392.

Garthwaite, P. H., Fan, Y., and Sisson, S. A. (2010). "Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process." *arXiv preprint arXiv:1006.3690*.

Gelfand, A. E. and Schliep, E. M. (2016). "Spatial statistics and Gaussian processes: A beautiful marriage." *Spatial Statistics*, 18, 86–104.

Gelman, A. and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." *Statistical science*, 7, 4, 457–472.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Griffith, D. A. (2000). "A linear regression solution to the spatial autocorrelation problem." *Journal of Geographical Systems*, 2, 2, 141–156.

— (2002). "A spatial filtering specification for the auto-Poisson model." *Statistics & probability letters*, 58, 3, 245–251.

— (2004). "A spatial filtering specification for the autologistic model." *Environment and Planning A*, 36, 10, 1791–1811.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Hughes, J. and Haran, M. (2013). "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 1, 139–159.

Karlis, D., Ntzoufras, I., et al. (2005). "Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R." *Journal of Statistical Software*, 14, 10, 1–36.

Kéry, M., Royle, J. A., and Schmid, H. (2005). "Modeling avian abundance from replicated counts using binomial mixture models." *Ecological applications*, 15, 4, 1450–1461.

Lawless, J. F. (1987). "Negative binomial and mixed Poisson regression." *Canadian Journal of Statistics*, 15, 3, 209–225.

McNicoll, G. (1997). "Population and poverty: A review and restatement."

Meehan, T. D., Michel, N. L., and Rue, H. (2017). "Estimating animal abundance with N-mixture models using the R-INLA package for R." *arXiv preprint arXiv:1705.01581*.

Neal, R. M. (2011). "MCMC using Hamiltonian dynamics." *Handbook of markov chain monte carlo*, 2, 11, 2.

Polson, N. G. and Scott, J. G. (2011). "Default Bayesian analysis for multi-way tables: a data-augmentation approach." *arXiv preprint arXiv:1109.4180*.

Porter, A. T., Holan, S. H., Wikle, C. K., and Cressie, N. (2014). "Spatial Fay–Herriot models for small area estimation with functional covariates." *Spatial Statistics*, 10, 27–42.

Rodgers, G. (1984). "Poverty and population: Approaches and evidence."

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the royal statistical society: Series b (statistical methodology)*, 71, 2, 319–392.

Scherr, S. J. (2000). "A downward spiral? Research evidence on the relationship between poverty and natural resource degradation." *Food policy*, 25, 4, 479–498.

Sinding, S. W. (2009). "Population, poverty and economic development." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1532, 3023–3030.

Terza, J. V., Wilson, P. W., et al. (1990). "Analyzing frequencies of several types of events: A mixed multinomial-Poisson approach." *The Review of Economics and Statistics*, 72, 1, 108–115.

Tiefelsdorf, M. and Griffith, D. A. (2007). "Semiparametric filtering of spatial autocorrelation: the eigenvector approach." *Environment and Planning A*, 39, 5, 1193–1221.

Wahba, G. (1990). *Spline models for observational data*, vol. 59. Siam.

Wikle, C. K. (2002). "A kernel-based spectral model for non-Gaussian spatio-temporal processes." *Statistical Modelling*, 2, 4, 299–314.

Wu, G., Holan, S. H., Nilon, C. H., Wikle, C. K., et al. (2015). "Bayesian binomial mixture models for estimating abundance in ecological monitoring studies." *The Annals of Applied Statistics*, 9, 1, 1–26.