

AUTOLOGISTIC MODELS FOR BINARY DATA ON A LATTICE

JOHN HUGHES^{1,*}, MURALI HARAN^{2,**}, AND PETRUȚA C. CARAGEA^{3,***}

^{1,2}DEPARTMENT OF STATISTICS, THE PENNSYLVANIA STATE
UNIVERSITY, UNIVERSITY PARK, PA 16802, USA

³DEPARTMENT OF STATISTICS, IOWA STATE UNIVERSITY, AMES,
IOWA 50011, USA

ABSTRACT. The autologistic model is a Markov random field model for spatial binary data. Because it can account for both statistical dependence among the data and for the effects of potential covariates, the autologistic model is particularly suitable for problems in many fields, including ecology, where binary responses, indicating the presence or absence of a certain plant or animal species, are observed over a two-dimensional lattice. We consider inference and computation for two models: the original autologistic model due to Besag, and the centered autologistic model proposed recently by Caragea and Kaiser. Parameter estimation and inference for these models is a notoriously difficult problem due to the complex form of the likelihood function. We study pseudolikelihood (PL), maximum likelihood (ML), and Bayesian approaches to inference and describe ways to optimize the efficiency of these algorithms and the perfect sampling algorithms upon which they depend, taking advantage of parallel computing when possible. We conduct a simulation study to investigate the effects of spatial dependence and lattice size on parameter inference, and find that inference for regression parameters in the centered model is reliable only for reasonably large lattices ($n > 900$) and no more than moderate spatial dependence. When the lattice is large enough, and the dependence small enough, to permit reliable inference, the three approaches perform comparably, and so we recommend the PL approach for its easier implementation and much faster execution.

1. INTRODUCTION

Spatial binary data appear frequently in several disciplines including ecology, agriculture, epidemiology, geography, and image analysis. In this paper we will focus on binary data on a lattice. (See, e.g.,

Date: September 17, 2010.

Key words and phrases. Bayesian, Markov random field, maximum likelihood, parallel computing, perfect sampling, pseudolikelihood, spatial.

(De Oliveira, 2000) for models for binary data on a continuous spatial domain.) For example, Figure 1 shows the presence of Hermit Thrush (*Catharus guttatus*), a North American songbird, in Pennsylvania (<http://www.carnegiennh.org/atlas/home.htm>). Presence is shown in black.

The autologistic model (Besag, 1972, 1974), which directly imposes a joint Markov random field for the 0-1 data, has proved a very useful model for data of this type. The model has a nearly 40-year history and has been widely used (see, for instance, Gumpertz et al., 1997; Huffer and Wu, 1998; Augustin et al., 1996; Koutsias, 2003; He et al., 2003; Sanderson et al., 2005; Moon and Russell, 2008), perhaps owing to the appeal of the model’s simple and direct specification of dependence and the ease with which the model can be fit by maximum pseudolikelihood estimation.

While pseudolikelihood (Besag, 1975) is a fast and convenient approach to approximate inference for the autologistic model, recent advances in computational methods have allowed for more rigorous Monte Carlo approaches to likelihood (Geyer, 1994) and Bayesian (Møller et al., 2006) inference in the presence of intractable normalizing functions. In addition, Caragea and Kaiser (2009) recently proposed a reparameterization that furnishes the autologistic model with interpretable parameters. However, there is little guidance regarding how well these models and algorithms work in practice as spatial dependence increases or decreases and as the size of the lattice varies.

In this paper we conduct a careful study of autologistic modeling and computation. In Section 2 we review models for binary data on a lattice, with our focus on comparing and contrasting Besag’s traditional autologistic model and the centered autologistic model of Caragea and Kaiser. In Section 3 we describe efficient computing for pseudolikelihood (PL), maximum likelihood (ML), and Bayesian inference for the two models. To our knowledge, this is the first development of MCML and rigorous Bayesian inference for the centered autologistic model. Since the algorithm of Møller et al. requires exact samples from the model of interest, we also develop perfect samplers for both autologistic models, and we use those samplers to parallelize the PL and ML algorithms. Section 4 gives the results of our thorough simulation study, where we investigated the effects on inference of lattice size, degree of dependence, edge percentage, and replication. We provide recommendations regarding lattice sizes for which inference is reliable and computation is feasible. And in Section 5 we present an analysis of

the Hermit Thrush data. In the appendix we derive the joint distribution for the centered autologistic model, which is required for ML and Bayesian inference.

2. MODELS FOR BINARY DATA ON A LATTICE

There are two dominant competing approaches to modeling binary areal data: the logistic spatial generalized linear mixed model (SGLMM) and the autologistic model. Each approach is characterized by its modeling of spatial dependence. The SGLMM models dependence indirectly, by way of a latent Gaussian Markov random field over the lattice in question (Banerjee et al., 2004, Ch. 3). The autologistic model, on the other hand, models dependence directly, through the so-called *autocovariate*, which is a function of the observations themselves. In this paper we will focus on the autologistic model, of which there are two variants.

2.1. The Traditional Autologistic Model. In his seminal 1974 paper, Besag proposed the automodels, which model spatial dependence among random variables directly (rather than hierarchically) and conditionally (rather than jointly). Among the automodels formulated by Besag was the autologistic model for spatial data with binary responses. The autologistic model has since found many applications in several fields, particularly ecology and epidemiology (Huffer and Wu, 1998; Gumpertz et al., 1997). We will henceforth refer to Besag’s formulation as the traditional autologistic model, which is defined as follows.

If we let \mathbf{Z} be the random field of interest, where $Z_i \in \{0, 1\}$ represents the observation at the i th lattice point for $i = 1, \dots, n$, the full conditional distributions for the traditional autologistic model are given by

$$(1) \quad \log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = \mathbf{X}_i \boldsymbol{\beta} + \sum_{j \neq i} \eta_{ij} Z_j,$$

where \mathbf{X}_i is the i th row of the design matrix, $\boldsymbol{\beta}$ are the regression parameters, and $\boldsymbol{\eta} = \{\eta_{ij}\}$ are dependence parameters such that $\eta_{ij} \neq 0$ iff Z_i and Z_j are neighbors. The sum in (1) is the autocovariate, which models the dependence between Z_i and the remainder of the field, \mathbf{Z}_{-i} . In an SGLMM, the autocovariate is replaced by the i th element of a latent field of random effects.

In this paper we consider only models for which $\eta_{ij} = \eta \mathbf{1}_{\{i \sim j\}}$ (where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function and \sim denotes the neighbor relation) and $\eta > 0$. We assume pairwise-only dependencies, i.e., we assume that

the underlying graph has clique number 2 (Cressie, 1993). In this case the full parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\beta}', \eta)'$, and Brook's Lemma implies the joint distribution

$$\begin{aligned}\pi(\mathbf{Z} | \boldsymbol{\theta}) &= c(\boldsymbol{\theta})^{-1} \exp \left(\sum_i Z_i \mathbf{X}_i \boldsymbol{\beta} + \frac{\eta}{2} \sum_{i,j} \mathbf{1}_{\{i \sim j\}} Z_i Z_j \right) \\ &= c(\boldsymbol{\theta})^{-1} \exp \left(\mathbf{Z}' \mathbf{X} \boldsymbol{\beta} + \frac{\eta}{2} \mathbf{Z}' \mathbf{A} \mathbf{Z} \right),\end{aligned}$$

where \mathbf{A} is an $n \times n$ adjacency matrix, i.e., $\mathbf{A}_{ij} = \mathbf{1}_{\{i \sim j\}}$, and $c(\boldsymbol{\theta})$ is an intractable normalizing function (Brook, 1964; Cressie, 1993, Ch. 6). The normalizing function makes computation challenging for both maximum likelihood and Bayesian inference.

We note that using $Q(\mathbf{Z} | \boldsymbol{\theta}) = \mathbf{Z}' \mathbf{X} \boldsymbol{\beta} + \frac{\eta}{2} \mathbf{Z}' \mathbf{A} \mathbf{Z}$, which is usually called the negpotential function, allows us to write the joint density as

$$\pi(\mathbf{Z} | \boldsymbol{\theta}) = \frac{\exp(Q(\mathbf{Z} | \boldsymbol{\theta}))}{\sum_{\mathbf{Y} \in \Omega} \exp(Q(\mathbf{Y} | \boldsymbol{\theta}))},$$

where the sample space, Ω , is $\{0, 1\}^n$ for a lattice with n points.

2.2. The Centered Autologistic Model. Caragea and Kaiser (2009) showed that the traditional autologistic model fails to provide meaningful interpretations of the model parameters—informally, the traditional model's non-negative autocovariate biases realizations of the field toward $\mathbf{Z} = \mathbf{1}$. They proposed a centered parameterization—reminiscent of auto-Gaussian models (Cressie, 1993)—that remedies the problem. Consequently, the simulation study described below focused on the centered model, and we recommend that the centered model be used in practice.

The full conditional distributions for the centered autologistic model are

$$(2) \quad \log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = \mathbf{X}_i \boldsymbol{\beta} + \sum_{j \neq i} \eta_{ij} (Z_j - \mu_j),$$

where μ_j is the independence expectation of Z_j :

$$\mu_j = \mathbb{E}(Z_j | \boldsymbol{\eta} = \mathbf{0}) = \frac{\exp(\mathbf{X}_j \boldsymbol{\beta})}{1 + \exp(\mathbf{X}_j \boldsymbol{\beta})}.$$

For the case where $\eta_{ij} = \eta \mathbf{1}_{\{i \sim j\}}$, (2) implies the joint distribution

$$\pi(\mathbf{Z} | \boldsymbol{\theta}) = c(\boldsymbol{\theta})^{-1} \exp \left(\mathbf{Z}' \mathbf{X} \boldsymbol{\beta} - \eta \mathbf{Z}' \mathbf{A} \boldsymbol{\mu} + \frac{\eta}{2} \mathbf{Z}' \mathbf{A} \mathbf{Z} \right),$$

where $\boldsymbol{\mu}$ is the vector of independence expectations. See the appendix for the derivation.

Following Caragea and Kaiser, we will henceforth refer to $\boldsymbol{\mu}$, i.e., the mean surface under the assumption of independence, as the *large-scale structure* implied by (2). And by *local structure* or *small-scale structure* we mean clustering induced by the dependence component of the model.

To see why η is difficult to interpret for the traditional model, consider a field under the traditional model having constant independence expectation. Suppose that exactly one of Z_i 's neighbors is equal to 1. Then Z_i 's autocovariate is equal to 1, which implies an increase in $\mathbb{P}(Z_i = 1)$. But any meaningful autocovariate should be negative and increase $\mathbb{P}(Z_i = 0)$ in this situation because the majority of Z_i 's neighbors are 0. Hence, η has no clear interpretation in the traditional model, and so neither does $\boldsymbol{\beta}$.

The centered autocovariate, on the other hand, is signed and measures local structure against large-scale structure (through the independence expectations). This captures a proper notion of dependence and so allows us to interpret the parameters in the obvious way: η represents the “reactivity” of an observation to its neighbors, conditional on the large-scale structure represented by the regression component of the model, $\mathbf{X}\boldsymbol{\beta}$. See Caragea and Kaiser (2009) for a rigorous treatment of interpretability of autologistic models.

3. MONTE CARLO INFERENCE FOR THE AUTOLOGISTIC MODELS

In this section we describe efficient computational approaches to PL, ML, and Bayesian inference for the autologistic models, following Besag (1975), Geyer (1994), and Møller et al. (2006), respectively; Zheng and Zhu (2008) explored these approaches in the context of the traditional autologistic model only.

Since we make frequent use of perfect sampling, we begin with a brief description of efficient perfect sampling for the autologistic models.

3.1. Perfect Sampling. Our algorithms rely heavily on perfect sampling, for three reasons. First, the algorithm of Møller et al. requires that we draw an exact sample during each iteration. Failure to use a perfect sampler results in an algorithm lacking theoretical justification and can affect the accuracy of resulting inference (Murray et al., 2006). Second, perfect samplers for conditionally specified distributions are relatively easy to understand and to implement. And third, a properly optimized perfect sampler is sufficiently fast to allow for the PL analysis of large datasets.

Our perfect samplers are based on coupling from the past (CFTP) (Propp and Wilson, 1996). A CFTP sampler for an autologistic model can be constructed as follows.

We want to simulate exactly from the autologistic model:

$$\mathbb{P}(Z_i = 1 \mid \mathbf{Z}_{-i}, \boldsymbol{\theta}) = p_i = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \sum_{j \sim i} Z_j^*)}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \sum_{j \sim i} Z_j^*)},$$

where $Z_j^* = Z_j$ for the uncentered model and $Z_j^* = Z_j - \mu_j$ for the centered model. The model is said to be attractive if $\eta \geq 0$, which is to say that the cdf, F_i , of Z_i is decreasing in $\sum_{j \sim i} Z_j^*$. Note that

$$F_i(z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 - p_i & \text{if } 0 \leq z < 1 \\ 1 & \text{if } z \geq 1. \end{cases}$$

Since we require non-negative η , p_i is increasing in $\sum_{j \sim i} Z_j^*$, and so F_i is in fact decreasing in $\sum_{j \sim i} Z_j^*$. For such a model, CFTP proceeds as follows (Møller, 1999).

Let $\mathbf{L}_T(t, i)$ and $\mathbf{U}_T(t, i)$ denote the i th observations at time t of the so-called lower and upper chains, respectively, where those chains were started at time T in the past. Fix $T < 0$ and set $\mathbf{L}_T(T, \cdot) = 0$ and $\mathbf{U}_T(T, \cdot) = 1$. Evolve the chains according to

$$\mathbf{L}_T(t, i) = F_i^{-1}(\mathbf{R}(t, i)) \mid \mathbf{L}_T(t, 1:i-1), \mathbf{L}_T(t-1, i+1:n)$$

$$\mathbf{U}_T(t, i) = F_i^{-1}(\mathbf{R}(t, i)) \mid \mathbf{U}_T(t, 1:i-1), \mathbf{U}_T(t-1, i+1:n),$$

where the $\mathbf{R}(t, i)$ are independent standard uniform variates and

$$F_i^{-1}(p) = \begin{cases} 0 & \text{if } p \leq 1 - p_i \\ 1 & \text{if } p > 1 - p_i. \end{cases}$$

If \mathbf{L}_T and \mathbf{U}_T coalesce at time $t_0 \leq 0$, return $\mathbf{L}_T(0, \cdot)$ as a sample from $\pi(\mathbf{Z} \mid \boldsymbol{\theta})$, the joint distribution of the Z_i . Otherwise, double T and start over. Use new uniform variates for $T, T+1, \dots, \frac{T}{2} - 1$, but reuse the previously generated variates for times $\frac{T}{2}, \frac{T}{2} + 1, \dots, -1$.

Our code was written in R (Ihaka and Gentleman, 1996). We note that replacing loops with vector and matrix operations, which are carried out by fast compiled code, reduces execution time dramatically. Our optimized implementation generates a sample in approximately 1/6 the time required by our first, naive, implementation.

Note that the adjacency matrix, \mathbf{A} , is sparse, but we cannot exploit this fact to speed up the sampler because \mathbf{A} must be accessed row by row in order to update \mathbf{L}_T and \mathbf{U}_T element by element.

3.2. Maximum Pseudolikelihood. The maximum pseudolikelihood estimate (MPLE), $\tilde{\boldsymbol{\theta}}$, of the parameter vector is the value of $\boldsymbol{\theta}$ that maximizes the product of the conditional likelihoods. That is, $\tilde{\boldsymbol{\theta}} = \arg \max \ell_{\text{PL}}(\boldsymbol{\theta})$, where

$$(3) \quad \ell_{\text{PL}}(\boldsymbol{\theta}) = \sum_i \log \frac{\exp(Z_i(\mathbf{X}_i\boldsymbol{\beta} + \eta \sum_{j \sim i} Z_j^*))}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + \eta \sum_{j \sim i} Z_j^*)}.$$

Although (3) is not the true log-likelihood except in the trivial case of independence, Besag (1975) showed that the MPLE converges almost surely to the MLE as the lattice size goes to ∞ (in density, not extent). For the centered autologistic model, (3) is given by

$$\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \eta\mathbf{A}(\mathbf{Z} - \boldsymbol{\mu})) - \sum_i \log(1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + \eta\mathbf{A}_i(\mathbf{Z} - \boldsymbol{\mu}))).$$

Armed with $\tilde{\boldsymbol{\theta}}$, we do inference by way of a parallel parametric bootstrap. That is, we generate, in parallel, B samples from $\pi(\mathbf{Z} | \tilde{\boldsymbol{\theta}})$ and compute the MPLE for each sample, thus obtaining the bootstrap sample $\tilde{\boldsymbol{\theta}}^{(1)}, \dots, \tilde{\boldsymbol{\theta}}^{(B)}$. Appropriate quantiles of the bootstrap sample are then used to construct approximate confidence intervals for the elements of $\boldsymbol{\theta}$. We took $B = 2,000$ and attained Monte Carlo standard errors on the order of 0.001, which indicates a very good approximation.

Because the bootstrap sample can be generated in parallel and little subsequent processing is required, this approach to inference is very efficient computationally, even for large lattices. For example, we found that PL analysis of a 40×40 dataset requires only a few minutes while MCML requires a few hours and the Bayesian procedure requires days (perhaps weeks, if the dependence is strong enough). A reasonable upper limit on the lattice size for a PL analysis is $n = 40,000$, for we were able to process a dataset of this size in approximately five days on a 64-node cluster with 3 GHz processors. We used R's `snow` (Tierney et al., 2009) package to parallelize our implementation.

3.3. Monte Carlo Maximum Likelihood. Although MPLE is straightforward and computationally efficient, it is known to be statistically inefficient, and so we appeal now to the statistical efficiency (but more demanding computation) of maximum likelihood estimation.

Here we derive the Monte Carlo maximum likelihood estimate (MCMLE) for the centered autologistic model. Although both MPLE and MCMLE are based on approximating $\pi(\mathbf{Z} | \boldsymbol{\theta})$, the former sidesteps the intractable normalizing constant while the latter approximates the ratio $c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})$.

More specifically, MCMLE employs the approximate log-likelihood

$$(4) \quad \ell_M(\boldsymbol{\theta}) = \log \frac{h(\mathbf{Z} | \boldsymbol{\theta})}{h(\mathbf{Z} | \tilde{\boldsymbol{\theta}})} - \log \left(\underbrace{\frac{1}{M} \sum_{i=1}^M \frac{h(\mathbf{Y}_i | \boldsymbol{\theta})}{h(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})}}_{\dagger} \right),$$

where h is the unnormalized likelihood, \mathbf{Z} are the data, $\tilde{\boldsymbol{\theta}}$ is the MPLE of $\boldsymbol{\theta}$, and $\mathbf{Y}_1, \dots, \mathbf{Y}_M$ are perfect samples from the model at $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$. Note that (\dagger) approximates $c(\boldsymbol{\theta})/c(\tilde{\boldsymbol{\theta}})$.

We maximize (4) to obtain $\hat{\boldsymbol{\theta}}_M$, the Monte Carlo maximum likelihood estimate of $\boldsymbol{\theta}$. According to Geyer (1994), $\hat{\boldsymbol{\theta}}_M$ converges almost surely to the maximum likelihood estimate, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ as $M \rightarrow \infty$. We found that taking $M = 10,000$ gives approximately the same Monte Carlo standard errors as for the bootstrap inference described in the previous section.

For the centered autologistic model,

$$h(\mathbf{Z} | \boldsymbol{\theta}) = \exp \left(\mathbf{Z}' \mathbf{X} \boldsymbol{\beta} - \eta \mathbf{Z}' \mathbf{A} \boldsymbol{\mu} + \frac{\eta}{2} \mathbf{Z}' \mathbf{A} \mathbf{Z} \right),$$

and so (4) becomes

$$(5) \quad \ell_M(\boldsymbol{\theta}) = Q(\mathbf{Z} | \boldsymbol{\theta}) - Q(\mathbf{Z} | \tilde{\boldsymbol{\theta}}) - \log \left(\frac{1}{M} \sum_{i=1}^M \exp(Q(\mathbf{Y}_i | \boldsymbol{\theta}) - Q(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})) \right),$$

where Q is the negpotential function. We obtain approximate sampling variances by taking the diagonal elements of $\hat{\boldsymbol{\mathcal{I}}}_M^{-1} = [-\nabla^2 \ell_M(\hat{\boldsymbol{\theta}}_M)]^{-1}$, which converges to $\hat{\boldsymbol{\mathcal{I}}}^{-1}$, the observed information matrix.

We obtain Monte Carlo standard errors as follows. Geyer (1994) showed that $\sqrt{M}(\hat{\boldsymbol{\theta}}_M - \hat{\boldsymbol{\theta}}) \Rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{I}}^{-1} \boldsymbol{\Sigma} \boldsymbol{\mathcal{I}}^{-1})$, and so we require an estimator of $\boldsymbol{\Sigma}$, which we now develop. Let

$$\begin{aligned} \mathbf{T}_i(\boldsymbol{\theta}) &= \left(\frac{\nabla h(\mathbf{Z} | \boldsymbol{\theta})}{h(\mathbf{Z} | \boldsymbol{\theta})} - \frac{\nabla h(\mathbf{Y}_i | \boldsymbol{\theta})}{h(\mathbf{Y}_i | \boldsymbol{\theta})} \right) \frac{h(\mathbf{Y}_i | \boldsymbol{\theta})}{h(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})} \\ &= \{ (\mathbf{Z} - \mathbf{Y}_i)' \mathbf{X} - \eta (\mathbf{Z} - \mathbf{Y}_i)' \mathbf{A} (\mathbf{X} \bullet \boldsymbol{\mu} \bullet (\mathbf{1} - \boldsymbol{\mu})) , \\ &\quad 1/2 (\mathbf{Z} - \mathbf{Y}_i)' \mathbf{A} (\mathbf{Z} - \mathbf{Y}_i) - (\mathbf{Z} - \mathbf{Y}_i)' \mathbf{A} \boldsymbol{\mu} \}' \\ &\quad \exp(Q(\mathbf{Y}_i | \boldsymbol{\theta}) - Q(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})), \end{aligned}$$

where \bullet denotes row wise multiplication (vector-vector or matrix-vector). Then

$$\frac{1}{\sqrt{M}} \sum_{i=1}^M \mathbf{T}_i(\boldsymbol{\theta}) \Rightarrow \frac{c(\boldsymbol{\theta})}{c(\tilde{\boldsymbol{\theta}})} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Since the sample mean from (5) estimates the ratio of normalizing constants, we have

$$(6) \quad \left(\frac{1}{M} \sum_{i=1}^M \exp(Q(\mathbf{Y}_i | \boldsymbol{\theta}) - Q(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})) \right)^{-1} \frac{1}{M} \sum_{i=1}^M \mathbf{T}_i(\boldsymbol{\theta}) \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{M} \boldsymbol{\Sigma}\right).$$

Since the $\mathbf{T}_i(\boldsymbol{\theta})$ are iid, (6) implies that we can estimate $\boldsymbol{\Sigma}$ using the sample covariance matrix of

$$\left(\frac{1}{M} \sum_{i=1}^M \exp(Q(\mathbf{Y}_i | \boldsymbol{\theta}) - Q(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})) \right)^{-1} \{\mathbf{T}_i(\boldsymbol{\theta})\}_{i=1}^M,$$

i.e.,

$$\hat{\boldsymbol{\Sigma}}_M = \widehat{Cov} \left(\left(\frac{1}{M} \sum_{i=1}^M \exp(Q(\mathbf{Y}_i | \hat{\boldsymbol{\theta}}_M) - Q(\mathbf{Y}_i | \tilde{\boldsymbol{\theta}})) \right)^{-1} \{\mathbf{T}_i(\hat{\boldsymbol{\theta}}_M)\}_{i=1}^M \right).$$

Thus the estimated Monte Carlo variances are the diagonal elements of $\frac{1}{M} \hat{\boldsymbol{\Sigma}}_M^{-1} \hat{\boldsymbol{\Sigma}}_M \hat{\boldsymbol{\Sigma}}_M^{-1}$.

MCMLE is considerably more intensive computationally than MPLE because MCMLE employs five times as many Monte Carlo samples to attain the same Monte Carlo standard errors. Generating those samples in parallel of course reduces the running time dramatically, and the appropriate use of vector and matrix operations greatly reduces the time required to optimize the approximate likelihood and calculate Monte Carlo standard errors.

3.4. MCMC for Bayesian Inference. Until recently, intractable normalizing functions made rigorous Bayesian analyses impossible for the autologistic and other models. But Møller et al. (2006) presented an auxiliary-variable MCMC algorithm that allows us to construct a proposal distribution so that the normalizing constant cancels out of the Metropolis-Hastings ratio. Employing their method requires only that we can draw independent realizations from the unnormalized density for any value of $\boldsymbol{\theta}$, which we do by means of perfect sampling. Here we describe their method for the centered autologistic model.

Let h again denote the unnormalized density, let Q denote the neg-potential function, let $\mathbf{Y} \in \Omega$ denote the auxiliary variable, and let $p(\cdot)$ denote a prior distribution. Then the Metropolis-Hastings random walk acceptance probability for the algorithm of Møller et al. is given by

$$\begin{aligned} \alpha &= \frac{h(\mathbf{Y}^* | \tilde{\boldsymbol{\theta}})h(\mathbf{Z} | \boldsymbol{\theta}^*)h(\mathbf{Y} | \boldsymbol{\theta})p(\boldsymbol{\theta}^*)}{h(\mathbf{Y} | \tilde{\boldsymbol{\theta}})h(\mathbf{Z} | \boldsymbol{\theta})h(\mathbf{Y}^* | \boldsymbol{\theta}^*)p(\boldsymbol{\theta})} \frac{c(\tilde{\boldsymbol{\theta}})}{c(\boldsymbol{\theta})} \frac{c(\boldsymbol{\theta})}{c(\boldsymbol{\theta}^*)} \frac{c(\boldsymbol{\theta}^*)}{c(\boldsymbol{\theta})} \\ &= \frac{h(\mathbf{Y}^* | \tilde{\boldsymbol{\theta}})h(\mathbf{Z} | \boldsymbol{\theta}^*)h(\mathbf{Y} | \boldsymbol{\theta})p(\boldsymbol{\theta}^*)}{h(\mathbf{Y} | \tilde{\boldsymbol{\theta}})h(\mathbf{Z} | \boldsymbol{\theta})h(\mathbf{Y}^* | \boldsymbol{\theta}^*)p(\boldsymbol{\theta})}, \end{aligned}$$

where \mathbf{Y}^* is the proposed auxiliary variable, $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^{*'}, \eta^*)'$ is the proposed $\boldsymbol{\theta}$, and $\tilde{\boldsymbol{\theta}}$ is the maximum pseudolikelihood estimate of $\boldsymbol{\theta}$. Taking logarithms and rearranging gives

$$\begin{aligned} \log \alpha &= Q(\mathbf{Y}^* | \tilde{\boldsymbol{\theta}}) - Q(\mathbf{Y}^* | \boldsymbol{\theta}^*) \\ &\quad + Q(\mathbf{Z} | \boldsymbol{\theta}^*) - Q(\mathbf{Z} | \boldsymbol{\theta}) \\ &\quad + Q(\mathbf{Y} | \boldsymbol{\theta}) - Q(\mathbf{Y} | \tilde{\boldsymbol{\theta}}) \\ &\quad + \log p(\boldsymbol{\theta}^*) - \log p(\boldsymbol{\theta}). \end{aligned}$$

Explicitly, for the centered model, we have

$$\begin{aligned} \log \alpha &= \mathbf{Y}^{*'}\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + \frac{\tilde{\eta} - \eta^*}{2}\mathbf{Y}^{*'}\mathbf{A}\mathbf{Y}^* + \eta^*\mathbf{Y}^{*'}\mathbf{A}\boldsymbol{\mu}^* - \tilde{\eta}\mathbf{Y}^{*'}\mathbf{A}\tilde{\boldsymbol{\mu}} \\ &\quad + \mathbf{Z}'\mathbf{X}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) + \frac{\eta^* - \eta}{2}\mathbf{Z}'\mathbf{A}\mathbf{Z} + \eta\mathbf{Z}'\mathbf{A}\boldsymbol{\mu} - \eta^*\mathbf{Z}'\mathbf{A}\boldsymbol{\mu}^* \\ &\quad + \mathbf{Y}'\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \frac{\eta - \tilde{\eta}}{2}\mathbf{Y}'\mathbf{A}\mathbf{Y} + \tilde{\eta}\mathbf{Y}'\mathbf{A}\tilde{\boldsymbol{\mu}} - \eta\mathbf{Y}'\mathbf{A}\boldsymbol{\mu} \\ &\quad + \log p(\boldsymbol{\theta}^*) - \log p(\boldsymbol{\theta}). \end{aligned}$$

Because the auxiliary proposals cannot be generated in parallel, this Bayesian analysis is by far the most computationally expensive of the three. Aside from optimizing the perfect sampler, we were able to achieve a small gain in efficiency as follows.

We used a normal random walk Metropolis-Hastings algorithm, and so our proposal for $\boldsymbol{\theta}$ was trivariate normal, i.e., $\boldsymbol{\theta}^{*(k+1)} | \boldsymbol{\theta}^{*(k)} \sim \mathcal{N}_3(\boldsymbol{\theta}^{*(k)}, \mathbf{V})$. Rather than take $\mathbf{V} = \mathbf{I}$ throughout, we let $\mathbf{V} = \mathbf{I}$ only for a training run of 100,000 iterations. Then we used the posterior sample covariance matrix from the training run as the proposal covariance matrix for a subsequent run of 300,000 iterations. We used the latter sample to do inference. The training run resulted in a much better mixing chain, which reduced the total number of iterations from $> 500,000$ to 400,000. Still, theoretically sound Bayesian inference (following Møller

et al.) is impractical for large lattices because the perfect sampler is in $O(n^2)$. Faster approximate Bayesian approaches exist but lack theoretical justification (Hoeting et al., 2000).

4. SIMULATION STUDY

The aim of our simulation study was to thoroughly evaluate the performance of the centered autologistic model in a regression setting.

We note that Dormann (2007) conducted a simulation study to assess the performance of autologistic regression. But Dormann’s study employed the traditional autologistic model, which has serious shortcomings and, we believe, should be supplanted by the centered autologistic model. Moreover, the lattice size in Dormann’s study was 1,108, which is considerably larger than the typical binary areal dataset, and Dormann evaluated the performance of the autologistic model based on MPLE alone.

Our study applied the above mentioned MPLE, MCMLE, and Bayesian procedures to datasets obtained from our perfect sampler for a range of lattice sizes and values of η . We simulated each dataset from the centered model; over square lattices of dimension 10, 20, and 40; for η equal to 0.2, 0.6, and 1; and with $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (1, 1)'$ and the spatial coordinates of the lattice points as covariates. This implies the conditional distributions

$$(7) \quad \log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = x_i + y_i + \eta \sum_{j \sim i} (Z_j - \mu_j).$$

The coordinates were restricted to the unit square, and so the regression component of the model imposes a probability gradient that increases from the origin toward $(1, 1)$. This large-scale structure is shown in Figure 2.

We chose the usual four-nearest-neighbor edge set for the underlying graph, which conforms to our restriction to graphs with clique number 2.

It is left to specify priors for the Bayesian analysis. Since $\eta \geq 0$ and 1 was the largest value of η used in our study, we assumed $\eta \sim \mathcal{U}(0, 2)$. And we chose $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 100 \mathbf{I})$.

The results of the study are shown in Figures 3, 4, and 5. Tabulated results are given in the appendix.

Evidently, for smaller lattices, the data contain too little information to infer η unless η is large, and the data contain too little information to reliably infer $\boldsymbol{\beta}$ irrespective of the degree of dependence (by ‘reliable’ we mean that inference is correct, i.e., a confidence interval almost

certainly contains the true value of the parameter and rarely contains 0 unless the true value of the parameter is 0). For larger lattices, inference for β is reliable so long as η is not too large, and inference for η is reliable when the strength of dependence is at least moderate.

Our desire to offer guidance regarding lattice size led us to conduct two followup studies on a 30×30 lattice. The first study included two additional values for η , 0.4 and 0.8. The table of results appears in the appendix. The second study fixed η at 0.6 and computed bootstrap confidence intervals for 1,000 simulated datasets. The resulting coverage rates appear in Table 1. These followup studies suggest that $n = 900$ is a threshold beyond which inference for the regression parameters should be valid provided that dependence is not too strong.

This finding is further supported by the plots in Figure 6. The three plots show maximum pseudolikelihood estimates of β for square lattices of dimension 20, 30, and 40, respectively. Each plot displays 1,000 estimates for $\eta = 0.2$ (white), 1,000 estimates for $\eta = 0.6$ (black), and 1,000 estimates for $\eta = 1$ (gray). We see that point estimation is poor for the 20×20 lattice. For the 30×30 lattice, point estimation is good when η equals 0.2 or 0.6 but begins to suffer when $\eta = 1$. And even $\eta = 1$ is not too strong to prevent good point estimation for the 40×40 lattice.

Note the emergence of a second mode (when $\eta = 1$) as the lattice size increases. The distribution is bimodal, with a smaller mode at the origin, when 1) dependence can be strong enough to allow small-scale structure to obscure large-scale structure, and 2) dependence is not strong enough to make for poor point estimation on the whole (as we see in the 20×20 case). This is not surprising; if large-scale structure is not evident in a given dataset, the data can be explained by dependence alone, resulting in $\tilde{\beta} \approx \mathbf{0}$. We have found that the mass of the smaller mode can be reduced by using the estimate of β obtained from a standard logistic regression as the starting value for pseudolikelihood optimization.

4.0.1. Autologistic Regression Versus Standard Logistic Regression. It is well known that failure to account for consequential dependence can lead to biased estimators and erroneous confidence intervals. The plot in Figure 7 shows the results of a comparison of autologistic regression and standard logistic regression for 1,000 realizations of the centered autologistic model on a 30×30 lattice and with $\eta = 0.7$. We see that standard logistic regression biased the estimator of β toward the origin and gave confidence intervals that were considerably too small.

4.0.2. *The Effect of Edge Percentage on Autologistic Regression.* Since edge observations have fewer neighbors than their interior counterparts, one might conjecture that, *ceteris paribus*, inference should become less reliable as the percentage of edge observations increases. The plots in Figure 8 show 1,000 values of $\tilde{\beta}$ and $\tilde{\eta}$, respectively, for each of three edge percentages, where $n = 1,600$ and $\eta = 0.7$: 10% (40×40), 20% (10×160), and 40% (5×320). Edge percentage evidently has little or no effect on inference.

4.0.3. *The Effect of Replication on Autologistic Regression.* Although the spatial analyst generally sees only one realization of a given spatial process, we thought it prudent to investigate the effect of replication on inference so that we might offer guidance to those fortunate enough to have replicates. The plots in Figure 9 show 1,000 values of $\tilde{\beta}$ for two, five, and ten replicates, where the lattice was 20×20 and $\eta = 0.7$. We see that even two replicates allow for reliable inference in this scenario. Further simulation studies indicated that a sensible threshold beyond which replication permits reliable inference is five replicates for a 10×10 lattice.

5. APPLICATION TO THE HERMIT THRUSH DATA

Our simulation study suggests that this dataset is large enough to permit sound inference (if the dependence is not too strong). Since pseudolikelihood inference is statistically efficient for large lattices and is far more efficient computationally than likelihood or Bayesian inference, we perform a pseudolikelihood analysis on the dataset.

The Hermit Thrush data were compiled by the 2nd Pennsylvania Breeding Bird Atlas, a joint project of the Carnegie Museum of Natural History and the Pennsylvania Game Commission. The dataset covers $n = 4,937$ locations surveyed between 2004 and 2009. The Hermit Thrush is a forest songbird associated primarily with mixed and coniferous forests at higher elevations (Jones and Donovan, 1996). We fit the following centered autologistic model to the data:

$$(8) \quad \log \frac{\mathbb{P}(Z_i = 1)}{\mathbb{P}(Z_i = 0)} = \beta_0 + \beta_1 c_i + \beta_2 e_i + \eta \sum_{j \sim i} (Z_j - \mu_j),$$

where c_i denotes the percentage of the i th areal unit having conifer or mixed cover and e_i denotes the average elevation of the i th unit. The estimate of $\theta = (\beta_0, \beta_1, \beta_2, \eta)'$ is given in Table 2 along with bootstrap 95% confidence intervals. We see that elevation is a highly significant predictor of Hermit Thrush presence while conifer/mixed woodland cover is not significant (at the 0.05 level), probably because mixed

and coniferous forests are found mainly at higher elevations in Pennsylvania. We note that the two predictors are only weakly correlated ($\hat{\rho}(\mathbf{c}, \mathbf{e}) \approx 0.2$).

Since the dataset is large, the above mentioned inference should be valid if the dependence is not too strong. Caragea and Kaiser (2009) claim that, given a regular lattice and four-nearest-neighbor edge set, large-scale structure should dominate small-scale structure when $\eta \leq 1$, and they suggest $\eta = 1$ as a conservative upper limit. Our simulation study corroborates this recommendation (so long as the lattice is large enough) but also shows that a large lattice may permit good point estimation even when η is larger than 1. How much larger remains an open question, but we offer the following data-based heuristic for judging dependence strength.

Our simulations indicate that the bootstrap distribution corresponding to $\tilde{\eta}$, i.e., the distribution of $\tilde{\eta}^{(j)}, j = 1, \dots, B$, is approximately normal provided the dependence is neither too weak nor too strong. As the strength of dependence increases, however, the distribution departs from normality, and this departure tends to signal the beginning of poor point estimation for β .

In any case, $\tilde{\eta} < 1$ for the thrush data, and so we conclude that regression inference is valid.

Fitting the uncentered autologistic and standard logistic models to the thrush data yielded the estimates shown in the final two panels of Table 2. The differences among the three estimates of β imply considerable discrepancies among the corresponding large-scale structures. Since the traditional autologistic model tends to misrepresent large-scale structure and the standard logistic model leads to bias when dependence is appreciable, we conclude that the centered autologistic model provides the best point estimate for these data. Moreover, the standard logistic procedure finds conifer/mixed cover to be a highly significant predictor of Hermit Thrush presence, a conclusion that hardly seems tenable given the standard model's propensity to underestimate variability when dependence is consequential.

6. DISCUSSION

In this paper we have carefully examined the traditional and centered autologistic models along with Monte Carlo inference for these models. Because Besag's traditional autologistic model lacks interpretable parameters, we recommend the newly proposed centered autologistic model of Caragea and Kaiser. Their reparameterization of the autocovariate yields a model with easily interpreted parameters, which

results in estimates of the regression parameters that accord with the large-scale structure of the data.

We have explained in detail three approaches to inference: maximum pseudolikelihood followed by parametric bootstrap, Monte Carlo maximum likelihood, and MCMC for rigorous Bayesian inference. Our implementations of these algorithms employ perfect sampling, in parallel except for the Bayesian procedure. We compared the performance of the three approaches in a thorough simulation study that used four lattice sizes and up to five degrees of spatial dependence.

Our analysis found autologistic regression unreliable for lattices with fewer than 900 points. Unfortunately, binary spatial datasets often contain far fewer than 900 points, especially in ecology. But we remain optimistic about the centered autologistic model because our simulation study showed that pseudolikelihood inference, which is far easier to understand and to implement than the MCML and Bayesian approaches, is both statistically and computationally efficient for datasets that are large enough to permit valid inference. This happy confluence of model interpretability; reliable inference; statistical and computational efficiency; and ease of application recommends the centered autologistic model and corresponding pseudolikelihood inference to those researchers with large binary areal datasets like the Hermit Thrush dataset analyzed in this paper.

ACKNOWLEDGEMENTS

We thank Dr. Andrew Wilson for providing the Hermit Thrush dataset and for helpful discussions.

REFERENCES

- [1] Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33(2):339–347.
- [2] Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall Ltd.
- [3] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B: Methodological*, 36:192–236.
- [4] Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician: Journal of the Institute of Statisticians*, 24:179–196.
- [5] Besag, J. E. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, Series B: Methodological*, 34:75–83.

- [6] Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51:481–483.
- [7] Caragea, P. and Kaiser, M. (2009). Autologistic models with interpretable parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 14(3):281–300.
- [8] Cressie, N. A. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York, 2nd. edition.
- [9] De Oliveira, V. (2000). Bayesian prediction of clipped Gaussian random fields. *Computational Statistics & Data Analysis*, 34(3):299–314.
- [10] Dormann, C. (2007). Assessing the validity of autologistic regression. *Ecological Modelling*, 207(2-4):234–242.
- [11] Geyer, C. (1994). On the convergence of monte carlo maximum likelihood calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 261–274.
- [12] Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:131–156.
- [13] He, F., Zhou, J., and Zhu, H. (2003). Autologistic regression model for the distribution of vegetation. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(2):205–222.
- [14] Hoeting, J. A., Leecaster, M., and Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(1):102–114.
- [15] Huffer, F. and Wu, H. (1998). Markov chain monte carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, 54(2):509–524.
- [16] Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5:299–314.
- [17] Jones, P. W. and Donovan, T. M. (1996). Hermit thrush (*Catharus guttatus*). In Poole, A., editor, *The Birds of North America Online*. Cornell Lab of Ornithology, Ithaca.
- [18] Kaiser, M. and Cressie, N. (2000). The construction of multivariate distributions from markov random fields. *Journal of Multivariate Analysis*, 73(2):199–220.
- [19] Koutsias, N. (2003). An autologistic regression model for increasing the accuracy of burned surface mapping using Landsat Thematic Mapper data. *International Journal of Remote Sensing*, 24:2199–2204.
- [20] Møller, J. (1999). Perfect simulation of conditionally specified models. *Journal of the Royal Statistical Society, Series B, Methodological*,

- 61:251–264.
- [21] Møller, J., Pettitt, A., Berthelsen, K., and Reeves, R. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika*, 93(2):451–458.
 - [22] Moon, S. and Russell, G. J. (2008). Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science*, 54(1):71–82.
 - [23] Murray, I., Ghahramani, Z., and MacKay, D. (2006). Mcmc for doubly-intractable distributions. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 359–366.
 - [24] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9(1-2):223–252.
 - [25] Sanderson, R. A., Eyre, M. D., and Rushton, S. P. (2005). Distribution of selected macroinvertebrates in a mosaic of temporary and permanent freshwater ponds as explained by autologistic models. *Ecography*, 28(3):355–362.
 - [26] Tierney, L., Rossini, A., and Li, N. (2009). Snow : A parallel computing framework for the r system. *International Journal of Parallel Programming*, 37(1):78–90.
 - [27] Zheng, Y. and Zhu, J. (2008). Markov Chain Monte Carlo for a Spatial-Temporal Autologistic Regression Model. *Journal of Computational and Graphical Statistics*, 17(1):123–137.

APPENDIX

Derivation of the Negpotential Function. Here, following Kaiser and Cressie (2000), we derive the negpotential function for the centered autologistic model in a regression setting. We assume pairwise-only dependencies, i.e., that the underlying graph has clique number 2, but the extension to graphs with larger cliques is straightforward.

We start with the conditional density for Z_i , which is given by

$$f_i(Z_i | \mathbf{Z}_{-i}, \boldsymbol{\theta}) = p_i^{Z_i} (1 - p_i)^{1-Z_i},$$

where

$$\begin{aligned} p_i &= \frac{\exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \sum_{j \sim i} (Z_j - \mu_j))}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \sum_{j \sim i} (Z_j - \mu_j))} \\ &= \frac{\exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \mathbf{A}_i (\mathbf{Z} - \boldsymbol{\mu}))}{1 + \exp(\mathbf{X}_i \boldsymbol{\beta} + \eta \mathbf{A}_i (\mathbf{Z} - \boldsymbol{\mu}))}. \end{aligned}$$

Thus the log-density, which will be important in the sequel, is

$$\log f_i(Z_i | \mathbf{Z}_{-i}, \boldsymbol{\theta}) = Z_i(\mathbf{X}_i\boldsymbol{\beta} + \eta\mathbf{A}_i(\mathbf{Z} - \boldsymbol{\mu})) - \log(1 + \exp(\mathbf{X}_i\boldsymbol{\beta} + \eta\mathbf{A}_i(\mathbf{Z} - \boldsymbol{\mu}))). \quad (9)$$

Since we assume pairwise-only dependencies, only the first- and second-order H -functions are required to construct the negpotential function. More precisely,

$$Q(\mathbf{Z}) = \sum_{1 \leq i \leq n} H_i(Z_i) + \sum_{1 \leq i < j \leq n} H_{i,j}(Z_i, Z_j),$$

where, for some suitable $\mathbf{Z}^* \in \Omega$,

$$H_i(Z_i) = \log \frac{f_i(Z_i | \mathbf{Z}_{-i}^*)}{f_i(Z_i^* | \mathbf{Z}_{-i}^*)} \quad (10)$$

and

$$H_{i,j}(Z_i, Z_j) = \mathbf{1}_{\{i \sim j\}} \log \left(\frac{f_i(Z_i | Z_j, \mathbf{Z}_{-i,-j}^*)}{f_i(Z_i^* | Z_j, \mathbf{Z}_{-i,-j}^*)} \frac{f_j(Z_j^* | \mathbf{Z}_{-j}^*)}{f_j(Z_j | \mathbf{Z}_{-j}^*)} \right). \quad (11)$$

Choosing $\mathbf{Z}^* = \mathbf{0}$ and using (9) with (10) and (11) gives

$$\begin{aligned} H_i(Z_i) &= Z_i(\mathbf{X}_i\boldsymbol{\beta} - \eta\mathbf{A}_i\boldsymbol{\mu}) \\ H_{i,j}(Z_i, Z_j) &= \eta\mathbf{1}_{\{i \sim j\}} Z_i Z_j. \end{aligned}$$

Thus

$$\begin{aligned} Q(\mathbf{Z} | \boldsymbol{\theta}) &= \sum_i Z_i(\mathbf{X}_i\boldsymbol{\beta} - \eta\mathbf{A}_i\boldsymbol{\mu}) + \frac{\eta}{2} \sum_{i,j} \mathbf{1}_{\{i \sim j\}} Z_i Z_j \\ &= \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} - \eta\mathbf{Z}'\mathbf{A}\boldsymbol{\mu} + \frac{\eta}{2} \mathbf{Z}'\mathbf{A}\mathbf{Z}, \end{aligned}$$

which implies the joint distribution

$$\pi(\mathbf{Z} | \boldsymbol{\theta}) = \frac{\exp(Q(\mathbf{Z} | \boldsymbol{\theta}))}{\sum_{\mathbf{Y} \in \Omega} \exp(Q(\mathbf{Y} | \boldsymbol{\theta}))}.$$

For the traditional autologistic model, the second-order H -functions are the same while the first-order functions are given by $H_i(Z_i) = Z_i\mathbf{X}_i\boldsymbol{\beta}$. Consequently, the negpotential function for the traditional model is

$$\begin{aligned} Q(\mathbf{Z} | \boldsymbol{\theta}) &= \sum_i Z_i\mathbf{X}_i\boldsymbol{\beta} + \frac{\eta}{2} \sum_{i,j} \mathbf{1}_{\{i \sim j\}} Z_i Z_j \\ &= \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \frac{\eta}{2} \mathbf{Z}'\mathbf{A}\mathbf{Z}. \end{aligned}$$

Tabulated Results of Our Simulation Study. *E-mail address:* JPH264@PSU.EDU

E-mail address: MHARAN@STAT.PSU.EDU

E-mail address: CARAGEA@STAT.IASTATE.EDU

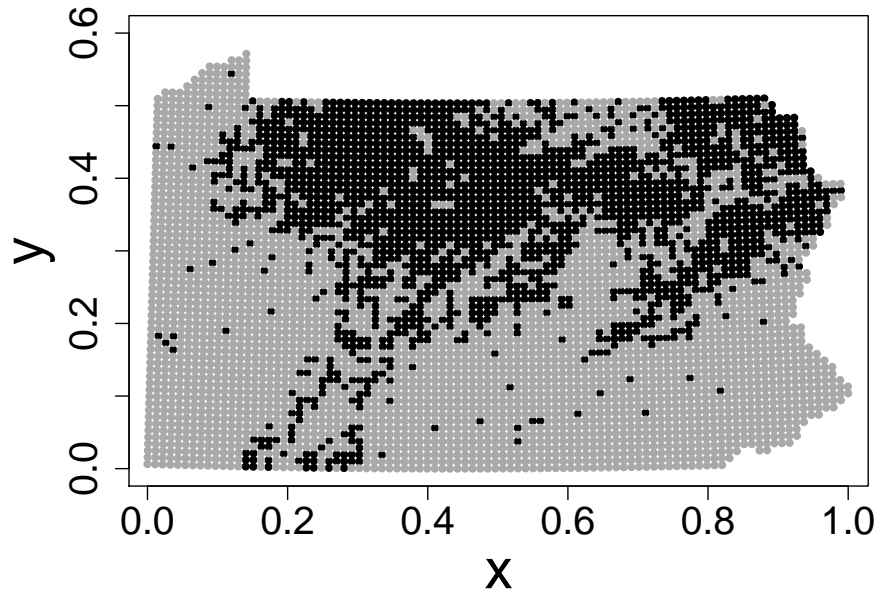


FIGURE 1. Hermit Thrush presence in Pennsylvania, 2004-2009. Black dots indicate presence.

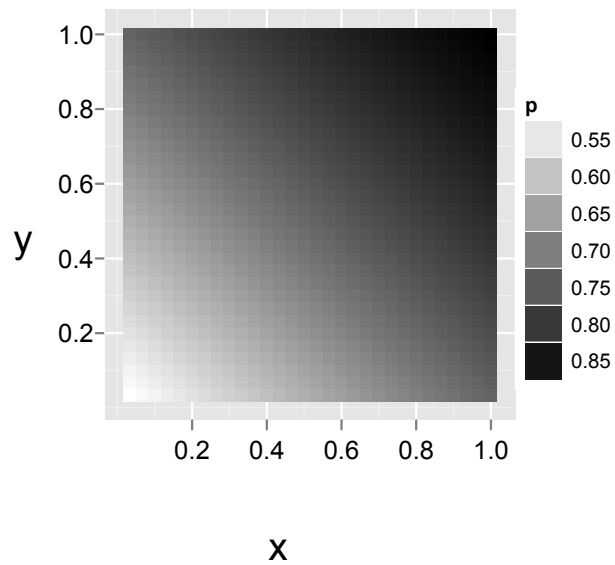


FIGURE 2. The probability gradient imposed by the regression component of our study.

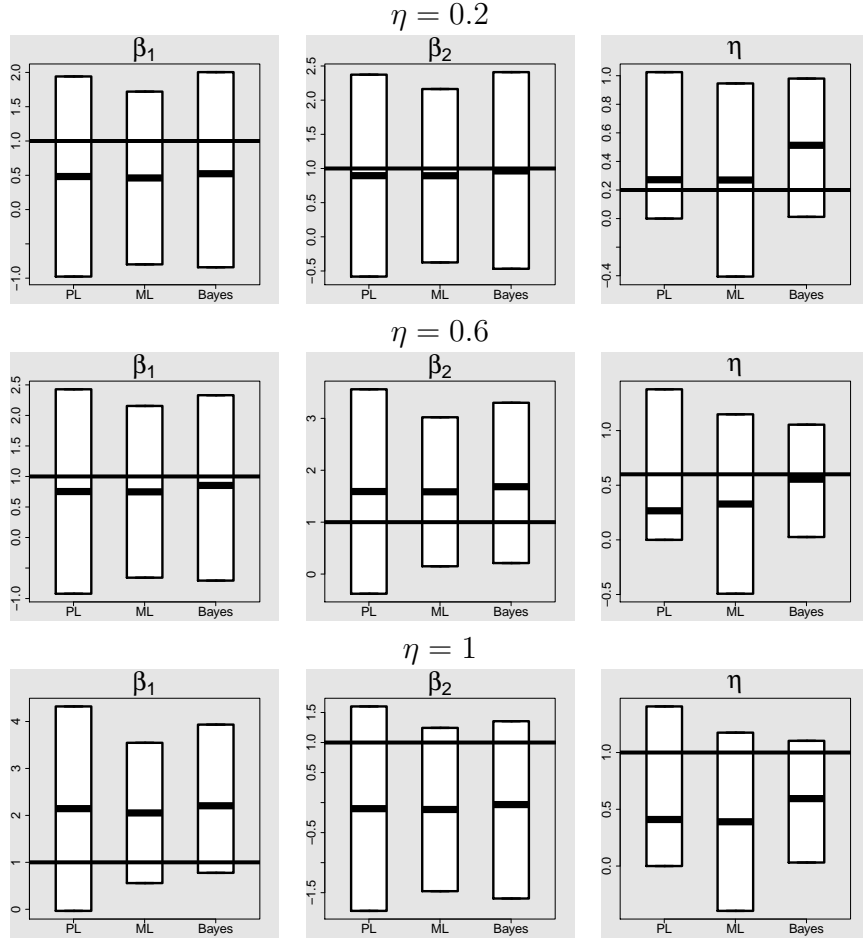


FIGURE 3. (10×10 lattice) The three approaches—MPLE, MCMLE, Bayes—perform about equally poorly for a lattice of this size. MCMLE yields the narrowest confidence intervals for β_1 and β_2 , but the intervals for all three procedures generally cover 0. The Bayesian approach gives the narrowest intervals for η , and this procedure alone was able to find statistically significant dependence among the data, but the significance is only marginal. A solid black line marks the true parameter value in each panel.

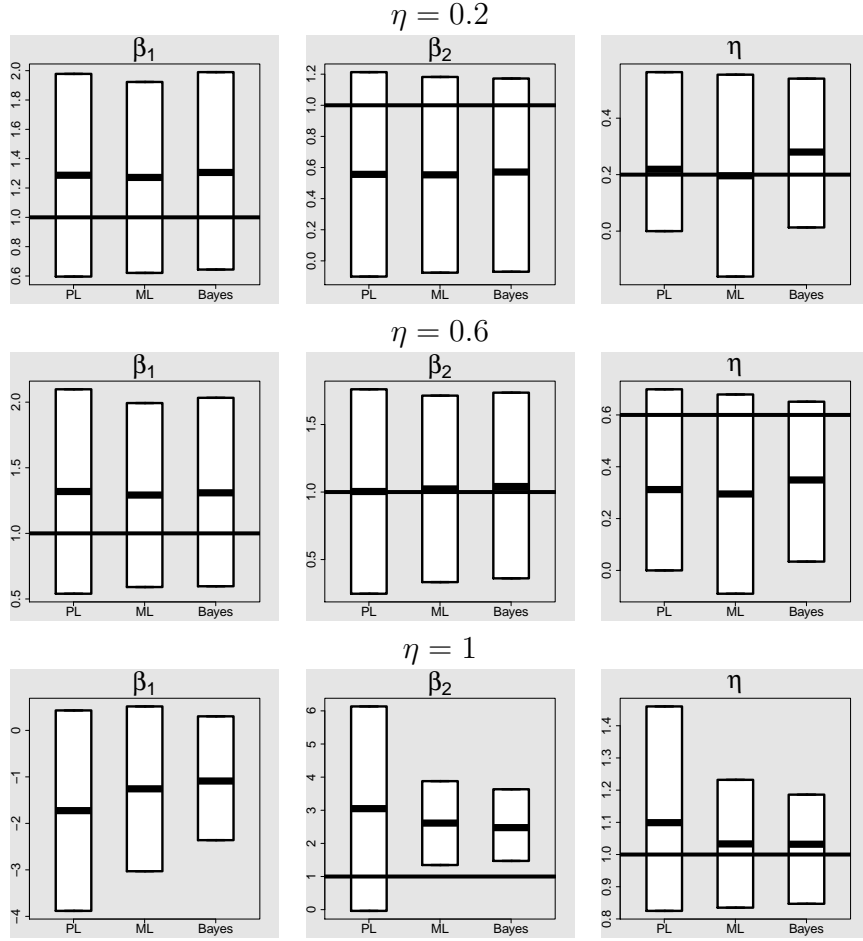


FIGURE 4. (20×20 lattice) The three procedures again performed comparably (and poorly) with respect to β , but the Bayesian procedure once again gave the narrowest intervals for η and was the only approach capable of finding significant dependence when the dependence was small to moderate. All three approaches provided reliable inference for η when $\eta = 1$, but a high degree of dependence is likely to prevent recovery of the regression parameters.

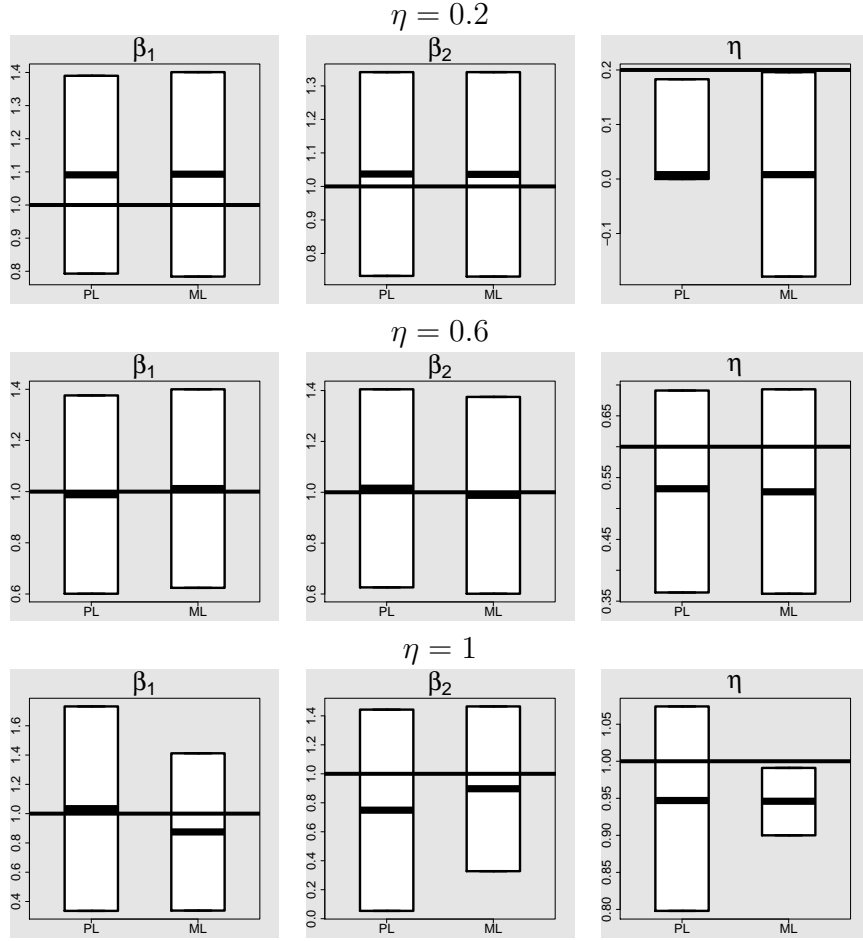


FIGURE 5. (40×40 lattice) The Bayesian procedure is impractical for a lattice this large. MPLE and MCMLE performed comparably. We note that $\eta = 1$ posed little problem for this larger lattice.

TABLE 1. These results are for 1,000 simulated 30×30 datasets with $\eta = 0.6$. The second column shows the coverage rates for the 95% bootstrap confidence intervals, the third column gives standard errors, and the final column gives a Type II error rate for each of β_1 and β_2 , i.e., the proportion of the intervals that covered 0.

Parameter	Coverage Rate	SE	Type II Error Rate
β_1	95.3%	0.007	3.8%
β_2	96.1%	0.006	5.1%
η	95.1%	0.007	NA

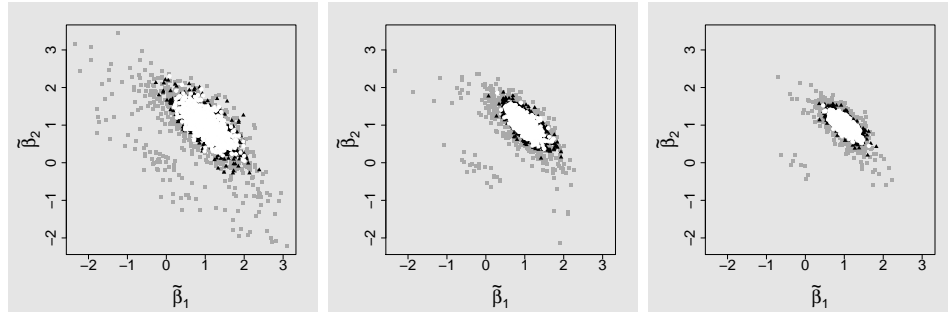


FIGURE 6. MPLEs of β for 20×20 , 30×30 , and 40×40 lattices, respectively. Each plot shows 1,000 estimates for each of three values of η : 0.2 (white), 0.6 (black), and 1 (gray).

TABLE 2. Results from fitting the centered autologistic, traditional autologistic, and standard logistic models to the Hermit Thrush data shown in Figure 1.

Model	Parameter	Estimate	CI
Centered	β_0	-2.18	(-2.345, -2.105)
	β_1	0.000555	(-0.00955, 0.0107)
	β_2	0.00736	(0.00606, 0.00866)
	η	0.668	(0.644, 0.692)
Traditional	β_0	-2.86	(-3.01, -2.71)
	β_1	0.00686	(-0.00202, 0.0157)
	β_2	0.00343	(0.00256, 0.00430)
	η	0.656	(0.610, 0.702)
Logistic	β_0	-0.736	(-0.811, -0.661)
	β_1	0.0288	(0.0218, 0.0358)
	β_2	0.00951	(0.00889, 0.0101)

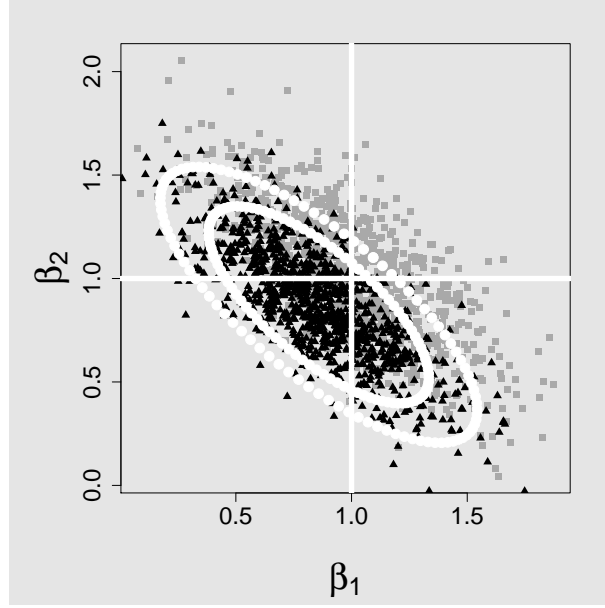


FIGURE 7. Comparison of standard logistic regression and autologistic regression for a 30×30 lattice and $\eta = 0.7$. The autologistic estimates are shown in gray, the standard logistic estimates in black. The inner white ellipse shows the mean asymptotic 95% confidence region for β for the standard logistic model. The outer ellipse shows the sample 95% confidence region. Note that, for this scenario, standard logistic regression biased $\hat{\beta}_1$ and $\hat{\beta}_2$ by 14.2% and 12.7%, respectively, while the PL estimators were biased by 0.2% and 2.2%.

TABLE 3. The results of our simulation study for a 10×10 lattice.

η	Parameter	MPLE	PL CI	MCMLE	ML CI	Posterior Mean	Bayes CI
0.2	β_1	0.482	(-0.977, 1.941)	0.461	(-0.799, 1.720)	0.523	(-0.841, 2.004)
	β_2	0.895	(-0.582, 2.373)	0.894	(-0.375, 2.163)	0.964	(-0.468, 2.410)
	η	0.272	(0, 1.025)	0.270	(-0.406, 0.946)	0.513	(0.012, 0.980)
0.6	β_1	0.753	(-0.921, 2.428)	0.748	(-0.658, 2.155)	0.853	(-0.706, 2.331)
	β_2	1.590	(-0.379, 3.560)	1.585	(0.150, 3.021)	1.683	(0.212, 3.300)
	η	0.266	(0, 1.378)	0.328	(-0.493, 1.149)	0.555	(0.026, 1.054)
1	β_1	2.144	(-0.034, 4.321)	2.052	(0.558, 3.546)	2.205	(0.776, 3.934)
	β_2	-0.101	(-1.803, 1.601)	-0.115	(-1.475, 1.244)	-0.033	(-1.597, 1.355)
	η	0.410	(0, 1.406)	0.390	(-0.395, 1.175)	0.594	(0.031, 1.103)

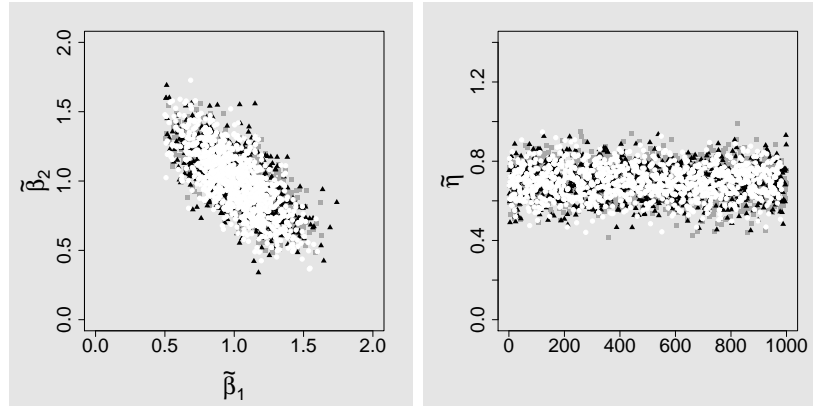


FIGURE 8. Edge percentage has little or no effect on inference. Estimates for 10% are shown in white, estimates for 20% are shown in black, and estimates for 40% are shown in gray.

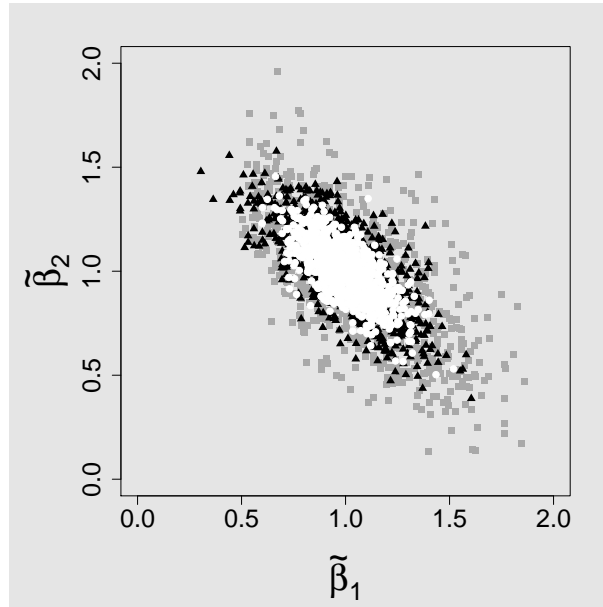


FIGURE 9. Replication has a dramatic effect on inference. Estimates for two replicates are shown in gray, estimates for five replicates are shown in black, and estimates for ten replicates are shown in white.

TABLE 4. The results of our simulation study for a 20×20 lattice.

η	Parameter	MPLE	PL CI	MCMLE	ML CI	Posterior Mean	Bayes CI
0.2	β_1	1.287	(0.596, 1.978)	1.272	(0.621, 1.923)	1.306	(0.644, 1.990)
	β_2	0.556	(-0.101, 1.213)	0.553	(-0.076, 1.182)	0.571	(-0.070, 1.172)
	η	0.219	(0, 0.563)	0.196	(-0.161, 0.554)	0.280	(0.013, 0.540)
0.6	β_1	1.319	(0.540, 2.098)	1.292	(0.591, 1.993)	1.309	(0.597, 2.033)
	β_2	1.004	(0.247, 1.761)	1.023	(0.333, 1.714)	1.041	(0.361, 1.736)
	η	0.312	(0, 0.699)	0.295	(-0.090, 0.679)	0.349	(0.034, 0.651)
1	β_1	-1.726	(-3.882, 0.429)	-1.257	(-3.029, 0.515)	-1.089	(-2.361, 0.302)
	β_2	3.049	(-0.037, 6.134)	2.613	(1.347, 3.879)	2.476	(1.471, 3.634)
	η	1.099	(0.825, 1.460)	1.033	(0.835, 1.232)	1.032	(0.847, 1.186)

TABLE 5. The results of our simulation study for a 40×40 lattice.

η	Parameter	MPLE	PL CI	MCMLE	ML CI
0.2	β_1	1.091	(0.793, 1.390)	1.093	(0.784, 1.401)
	β_2	1.037	(0.733, 1.341)	1.036	(0.731, 1.341)
	η	0.008	(0, 0.183)	0.008	(-0.179, 0.196)
0.6	β_1	0.988	(0.601, 1.376)	1.012	(0.624, 1.400)
	β_2	1.016	(0.626, 1.405)	0.988	(0.601, 1.375)
	η	0.532	(0.364, 0.691)	0.527	(0.362, 0.693)
1	β_1	1.034	(0.336, 1.732)	0.875	(0.338, 1.412)
	β_2	0.749	(0.054, 1.443)	0.897	(0.328, 1.465)
	η	0.947	(0.798, 1.074)	0.946	(0.900, 0.991)

TABLE 6. (30×30 lattice) All three approaches perform reasonably well on a lattice of this size, provided the dependence is not too strong.

η	Parameter	MPLE	PL CI	MCMLE	ML CI	Posterior Mean	Bayes CI
0.2	β_1	0.987	(0.516, 1.458)	0.983	(0.550, 1.417)	0.995	(0.589, 1.448)
	β_2	1.291	(0.781, 1.801)	1.297	(0.858, 1.735)	1.297	(0.873, 1.713)
	η	0.155	(0, 0.434)	0.154	(0.039, 0.270)	0.206	(0.013, 0.400)
0.4	β_1	0.858	(0.348, 1.368)	0.824	(0.360, 1.288)	0.836	(0.370, 1.302)
	β_2	1.234	(0.699, 1.769)	1.267	(0.796, 1.738)	1.270	(0.812, 1.728)
	η	0.377	(0.122, 0.622)	0.343	(0.107, 0.578)	0.357	(0.124, 0.589)
0.6	β_1	1.108	(0.510, 1.706)	1.131	(0.631, 1.632)	1.134	(0.665, 1.621)
	β_2	1.025	(0.463, 1.587)	1.022	(0.515, 1.528)	1.028	(0.560, 1.515)
	η	0.495	(0.238, 0.741)	0.500	(0.403, 0.598)	0.520	(0.303, 0.736)
0.8	β_1	1.140	(0.203, 2.077)	1.266	(0.716, 1.816)	1.238	(0.684, 1.760)
	β_2	0.871	(0.016, 1.725)	0.753	(0.169, 1.338)	0.828	(0.219, 1.407)
	η	0.767	(0.554, 1.150)	0.772	(0.683, 0.860)	0.840	(0.671, 0.998)
1	β_1	0.604	(-0.260, 1.467)	0.644	(-0.048, 1.336)	0.638	(-0.059, 1.306)
	β_2	0.614	(-0.243, 1.470)	0.670	(-0.021, 1.360)	0.665	(-0.026, 1.338)
	η	0.811	(0.623, 1.008)	0.822	(0.682, 0.961)	0.832	(0.667, 0.983)