

**MARKOV CHAIN MONTE CARLO FOR
AUTOLOGISTIC REGRESSION MODELS
WITH APPLICATION TO THE
DISTRIBUTION OF PLANT SPECIES**

Fred W. Huffer

and

Hulin Wu

Department of Statistics

Department of Mathematical Sciences

The Florida State University

The University of Memphis

Tallahassee, FL 32306-3033

Memphis, TN 38152

Tel. (904) 644-6696

Tel. (901) 678-4147

Email: huffer@stat.fsu.edu

Email: wuh@hermes.msci.memst.edu

MARKOV CHAIN MONTE CARLO FOR AUTOLOGISTIC REGRESSION MODELS WITH APPLICATION TO THE DISTRIBUTION OF PLANT SPECIES

SUMMARY

In this paper, we explore using autologistic regression models for spatial binary data with covariates. Autologistic regression models can handle binary responses exhibiting both spatial correlation and dependence on covariates. We use Markov chain Monte Carlo (MCMC) to estimate the parameters in these models. The distributional behavior of the MCMC maximum likelihood estimates (MCMC MLEs) is studied via simulation. We find that the MCMC MLEs are approximately normally distributed and that the MCMC estimates of Fisher information may be used to estimate the variance of the MCMC MLEs and to construct confidence intervals. Finally, we illustrate by example how our studies may be applied to model the distribution of plant species. In our simulations and data analysis, we have used the maximum pseudolikelihood estimates (MPLEs) as the “starting values” to generate the initial Monte Carlo sample. This seems to work well except when the degree of spatial correlation is very large.

1 Introduction

This project began as an attempt to model the distribution of plant species in terms of climate variables like temperature and rainfall. In particular, we were given data on the distribution of about 180 plant species in the state of Florida, and data on the values of 9 climate variables which were expected to be important factors in determining the distribution of the plant species. This data was assembled by Dr. D.W. Crumpacker of the University of Colorado at Boulder. The plant distribution data was obtained by digitizing maps from Little (1978). The climate information was compiled from a number of sources.

Key words: Spatial binary data; Markov chain Monte Carlo; Environmental statistics.

Figure 1 presents four of the distribution maps, species No. 1 (*Chamaecyparis Thyoides*), No. 4 (*Pinus Clausa*), No. 38 (*Castanea Pumila*) and No. 158 (*Zanthoxylum Clava-herculis*).

Place of Figure 1

The climate variables we are using are the following:

TMM = mean minimum temperature in degrees centigrade of coldest month
(usually January).

TM = mean temperature in degrees centigrade of coldest month (usually January).

TAV = mean annual temperature in degrees centigrade.

LT = lowest temperature recorded from 1931 to 1990 in degrees centigrade.

FZF = median freeze-free period in days.

PRCP = mean total annual precipitation in millimeters.

MI = moisture index = $(PRCP)/(TAV \times 58.93)$.

where $TAV \times 58.93$ = estimate of mean annual potential evapotranspiration
by the Holdridge method.

PMIN = mean total precipitation of driest month in millimeters.

ELV = elevation in feet.

Inspection of maps like those in Figure 1 suggests that there are strong climate effects on species distribution. In particular, the species boundaries are often observed to roughly follow contour lines for one of the temperature variables TMM, TM, TAV, LT or FZF. Inspection also reveals a strong degree of spatial correlation in the data, that is, whether a species is present or absent at a given site is strongly related to its presence or absence at neighboring sites. Some correlation is to be expected simply because the binary responses (presence or absence) depend on the climate covariates which

are themselves spatially correlated and display definite spatial patterns. However, for many species the degree of spatial correlation is much larger than can be explained in terms of the covariates alone.

Ecologists have analyzed similar data (sometimes using pollen abundance as a proxy for the actual species distribution) by a variety of means. Box et al. (1993) built a climatic-envelope model for many woody Florida plant species. This paper is particularly relevant to the current work as they use very similar climate variables and many of the same plant species that we do. Bartlein et al. (1986) and Huntley et al. (1989) constructed response surfaces for some plant species data by a locally-weighted averaging technique. This gave them graphical displays of the relationship between species distribution and climate. Austin et al. (1990) used logistic regression to model the dependence on the climate variables, but they did not consider spatial interaction/correlation in their models. In this paper, we try to model the binary data (species present or absent) in ways which properly account for both the spatial correlation and the dependence on the climate covariates.

This kind of spatial binary data with covariates occurs frequently in many fields such as geology, meteorology, medical science, and environmental science, so it is of great interest to develop and study models for dealing with this data. The model we shall use in this paper is the autologistic regression model which is explained in the next section (Section 2). The autologistic regression model is a straightforward generalization of the autologistic model introduced by Besag (1974, 1975) and studied by many authors (see Zhao and Prentice (1990), Geyer and Thompson (1992), Huffer and Wu (1993)). The parameters in this model will be estimated by a Markov Chain Monte Carlo (MCMC) approximation to the maximum likelihood estimates as in Geyer (1991a). This method of parameter estimation is reviewed in Section 3. Simulation studies on the distributional properties of MCMC estimates and confidence intervals are presented in Section 4. We find that the MCMC estimates are roughly normally distributed and that confidence intervals based on the estimated Fisher information perform very well so long as the amount of spatial correlation is not too large. In Section 5, we apply the methods proposed in this paper to one of the plant species

(No. 38).

2 The Autologistic Regression Model

We shall suppose that our data is recorded at m locations (sites) forming a subset \mathcal{S} of a rectangular lattice. Each site in \mathcal{S} is described by giving coordinates (k, ℓ) specifying the row and column of the lattice at which it is located. The sites in \mathcal{S} are numbered from 1 to m in some arbitrary fashion. At each site i we observe a binary response y_i and a $p \times 1$ vector of covariates \mathbf{x}_i . Taken altogether, the m binary responses constitute a map $Y = (y_i)$. The autologistic regression model specifies the conditional probability p_i that $y_i = 1$ given all the other values y_j ($j \neq i$) as follows:

$$\begin{aligned} p_i &= P(y_i = 1 | \text{all other values}) = P(y_i = 1 | \text{nearest neighbors}) \\ &= \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad \text{where} \quad \eta_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta}_1 + \gamma y_i^*. \end{aligned} \quad (2.1)$$

Here y_i^* denotes the neighborhood sum for site i , that is,

$$y_i^* = \sum_{j=1}^m y_j I(i \sim j) = \sum_{j: i \sim j} y_j$$

where $i \sim j$ indicates that sites i and j are “neighbors”. For a site i with coordinates (k, ℓ) , the neighbors j are any members of \mathcal{S} occupying the four nearest locations $(k-1, \ell)$, $(k+1, \ell)$, $(k, \ell-1)$, $(k, \ell+1)$. A site in the interior of \mathcal{S} will have 4 neighbors. Sites on the boundary of \mathcal{S} will have fewer neighbors.

The parameters in this model are the intercept β_0 , a $p \times 1$ vector $\boldsymbol{\beta}_1$ which specifies the covariate effects, and a parameter γ which determines the degree of spatial interaction or correlation in the data. When $\gamma = 0$, the model (2.1) reduces to the ordinary logistic regression model which is appropriate for independent binary responses. When $\boldsymbol{\beta}_1 = 0$, the model becomes the autologistic model of Besag (1974, 1975).

3 Markov Chain Monte Carlo Method

The spatial correlation term γy_i^* in (2.1) causes difficulty for parameter estimation. Carrying out exact maximum likelihood estimation is difficult because the likelihood

function is not tractable except when the number of sites m is quite small. Two estimation methods, the coding method and maximum pseudolikelihood (MPL) method have been proposed by Besag (1974, 1975). But these two methods are not efficient. Huffer and Wu (1993) and other authors concluded that the MCMC method is superior to the other two methods. In particular, Huffer and Wu (1993) have carried out simulation studies comparing the coding method, MPL method, and the MCMC method on data generated from the model (2.1). They found that when the amount of spatial interaction (as determined by γ) is sufficiently large, the MCMC MLE's are a substantial improvement on the other two methods. For the species data we are modeling, the spatial interaction is quite large. For this reason we shall restrict our attention to the MCMC MLE's in this paper.

3.1 Description of the MCMC Method

Markov chain Monte Carlo methods can be used to approximate the MLE for any family of distributions having probability densities known up to a constant of proportionality (Moyeed and Baddeley, 1991; Geyer, 1991a,b, 1992a,b; Geyer and Thompson, 1992; Gelman and Rubin, 1992). We shall first describe the MCMC method in general, following the notation of Geyer (1991b), and then specialize our discussion to the model (2.1).

Suppose the probability measure P_θ of our data y has a density or mass function f_θ (with respect to a measure μ) which can be written as:

$$f_\theta(y) = \frac{1}{z(\theta)} h_\theta(y)$$

where the function h_θ is known, but the normalizing constant

$$z(\theta) = \int h_\theta(y) d\mu(y)$$

is intractable. If we can generate observations Y_1, Y_2, \dots, Y_n from P_ψ , then we can estimate the ratio $z(\theta)/z(\psi)$ using the relation

$$\frac{z(\theta)}{z(\psi)} = \int \frac{h_\theta(y)}{h_\psi(y)} f_\psi(y) d\mu(y) = E_\psi \frac{h_\theta(Y)}{h_\psi(Y)} \approx \frac{1}{n} \sum_{i=1}^n \frac{h_\theta(Y_i)}{h_\psi(Y_i)}.$$

Here we are taking θ as a variable and ψ as fixed. This gives an estimate $l_n(\theta)$ of the log likelihood:

$$l_n(\theta) = \log \frac{h_\theta(Y_{\text{obs}})}{h_\psi(Y_{\text{obs}})} - \log \left(\frac{1}{n} \sum_{i=1}^n \frac{h_\theta(Y_i)}{h_\psi(Y_i)} \right). \quad (3.1)$$

Here we are using Y_{obs} to denote the observed data, and Y_1, Y_2, \dots, Y_n to denote the Monte Carlo sample generated from P_ψ . The maximizer of $l_n(\theta)$ is the Monte Carlo approximant of the MLE.

When the probability measures P_θ form an exponential family with $\theta \in R^p$, we can write

$$h_\theta(y) = e^{\theta' t}$$

where $t = t(y)$ is the vector of sufficient statistics. In this case, for given observed data Y_{obs} , the maximum likelihood estimate is obtained by finding θ which satisfies $E_\theta t(Y) = t(Y_{\text{obs}})$. The Monte Carlo analog of this result can be obtained by differentiating the log-likelihood function (3.1) and setting it equal to 0. i.e.

$$\frac{\sum_{j=1}^n T_j e^{(\theta-\psi)' T_j}}{\sum_{j=1}^n e^{(\theta-\psi)' T_j}} = T_{\text{obs}} \quad (3.2)$$

where we define $T_{\text{obs}} = t(Y_{\text{obs}})$ and $T_j = t(Y_j)$ for $j = 1, 2, \dots, n$. The Newton-Raphson method can be used to solve this equation. The solutions of Equation (3.2) for θ are the Monte Carlo MLEs. Note that this solution depends on the Monte Carlo sample Y_1, Y_2, \dots, Y_n only through the sufficient statistics T_1, T_2, \dots, T_n which form a set of points in R^p .

In our situation the data y consists of a binary map $y = (y_i)_{i=1}^m$ and the joint probability function of model (2.1) forms an exponential family with

$$h_\theta(y) = \exp \left(\beta_0 \sum_i y_i + \beta_1' \left(\sum_i \mathbf{x}_i y_i \right) + (\gamma/2) \sum_i y_i y_i^* \right)$$

so that

$$t(y) = \left(\sum_i y_i, \sum_i \mathbf{x}_i y_i, \frac{1}{2} \sum_i y_i y_i^* \right) \text{ and} \\ \theta = (\beta_0, \beta_1, \gamma).$$

Note that

$$\frac{1}{2} \sum_i y_i y_i^* = \sum_{i < j} y_i y_j I(i \sim j).$$

See Cressie (1991) and Besag (1974) for more detail. Thus the MCMC method outlined in the previous paragraphs applies in our situation.

The closer ψ is to the exact MLE, the smaller the Monte Carlo error is for a fixed Monte Carlo sample size n . The maximum pseudolikelihood estimate (MPLE) is a good candidate for ψ ; it is easily computed using standard software for logistic regression.

3.2 MCMC Estimation of the Fisher Information

The first derivative of the log likelihood approximant is:

$$\frac{\partial l_n(\theta)}{\partial \theta} = T_{\text{obs}} - \frac{\sum_{j=1}^n T_j e^{(\theta-\psi)'T_j}}{\sum_{j=1}^n e^{(\theta-\psi)'T_j}} = -\frac{\sum_{j=1}^n u_j e^{(\theta-\psi)'u_j}}{\sum_{j=1}^n e^{(\theta-\psi)'u_j}}$$

where $u_j = T_j - T_{\text{obs}}$. The second derivative evaluated at the MCMC maximum likelihood estimate $\hat{\theta}_n$ is

$$\left. \frac{\partial^2 l_n(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}_n} = -\frac{\sum_{j=1}^n u_j u_j' e^{(\hat{\theta}_n-\psi)'u_j}}{\sum_{j=1}^n e^{(\hat{\theta}_n-\psi)'u_j}}.$$

In obtaining the above expression, we have used the fact that

$$\left. \frac{\partial l_n(\theta)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0.$$

Thus we can estimate the Fisher information evaluated at $\hat{\theta}_n$ by

$$\widehat{I}(\hat{\theta}_n) = -\left. \frac{\partial^2 l_n(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}_n} = \frac{\sum_{j=1}^n u_j u_j' e^{(\hat{\theta}_n-\psi)'u_j}}{\sum_{j=1}^n e^{(\hat{\theta}_n-\psi)'u_j}}. \quad (3.3)$$

Under some assumptions, Geyer (1992b) claimed that $\widehat{I}(\hat{\theta}_n)$ converges to $I(\hat{\theta})$ in probability, as the MCMC sample size, n , tends to ∞ , where

$$I(\hat{\theta}) = -\left. \frac{\partial^2 l(\theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}},$$

$\hat{\theta}$ is the exact maximum likelihood estimate, and $l(\theta)$ is the exact log likelihood function.

We obtain the Fisher information estimate (3.3) as a by-product of our parameter estimation program. It does not require any extra computational effort.

The diagonal elements of the inverse Fisher information matrix can be used as rough estimates of the variances of the MCMC parameter estimates. These variance estimates will be asymptotically valid. We study this issue via simulation in the next section.

4 Distributional Properties of MCMC Estimates

Theoretical study of the behavior of MCMC estimates is difficult, so we shall approach this problem by conducting simulation studies. In our simulations we shall use the autologistic regression model (2.1) on a complete 40×40 lattice; the number m of sites is 1600. When computing the neighborhood sums y_i^* , the corner sites have two neighbors, and the sites along the edges have three neighbors. We shall restrict consideration to a single covariate ($p = 1$) so that η_i in (2.1) can be written simply as $\eta_i = \beta_0 + \beta_1 x_i + \gamma y_i^*$. This covariate has the form of a diagonal sine wave; for a site i with coordinates (k, ℓ) , the covariate takes on the value

$$x_i = 2.5 \times \sin(0.1 \times (k + \ell)).$$

For each set of true parameter values $(\beta_0, \beta_1, \gamma)$, 500 sets of the pseudo-observed data were generated by a Gibbs sampler (Geman and Geman (1984), Gelman and Rubin (1992), Geyer (1992)) using the model (2.1), i.e., the 500 data sets are obtained from 500 independent Markov chains. For each pseudo-observed data set, MCMC estimates of the parameters are calculated. So we have 500 independent MCMC estimates for each set of true parameter values, which may be examined for their bias, variance, and other distributional properties. We shall use this same simulated data throughout this section.

The “one long run” scheme is used to implement the MCMC. We start from the “zero-state”, i.e., the initial value is $y_i = 0$ for all sites i . The burn-in or warm-up period of 100, spacing of 1, and MCMC sample size of 1000 are chosen on the basis of our experience and the suggestions of Geyer and Thompson (1992) and Geyer (1991a, 1992a). In other words, with Z_i denoting the Monte Carlo sample after the i^{th} sweep of the entire lattice by the Gibbs sampler, we are collecting samples $Y_j = Z_{100+2j}, j = 1, 2, \dots, 1000$. Our computational work was done using an S-PLUS interface to FORTRAN on SUN SPARC Stations.

For generating our Monte Carlo sample, we take ψ to be the maximum pseudolikelihood estimate (MPLE) obtained by fitting a logistic regression model for the response y_i on the covariates x_i and y_i^* . Occasionally, the MPLE turned out to be a bad value

for ψ and led to a Monte Carlo sample $T_1, T_2, \dots, T_{1000}$ which did not contain the pseudo-observed value T_{obs} within its convex hull. For these samples the MCMC MLE does not exist, and attempting to solve (3.2) by Newton-Raphson iteration leads to a sequence of estimates $\theta_1, \theta_2, \theta_3, \dots$ which drift off to infinity. When this happened, we took the first step θ_1 of the Newton-Raphson iteration as a new value for ψ and generated another Monte Carlo sample using this new value. This somewhat ad hoc adjustment worked (at least for the range of parameter values reported in our tables) and produced samples which led to well-defined MCMC MLE's.

Our choice of parameter values was motivated partly by the species data we wish to model. In most of the species we examined, the covariate effects and the spatial interaction were both quite large. So in these simulations we have taken $\beta_1 = 2$ which produces strong covariate effects. (We have also run simulations using other values of β_1 and the results were qualitatively similar.) We use a wide range of γ values. In applications γ is generally positive, but out of curiosity we have also examined negative values of γ and have given results for one of these ($\gamma = -1.5$) in our tables.

We shall examine only univariate aspects of the sampling distribution of the MCMC estimates, that is, we shall examine each parameter estimate separately. Studying the multivariate behavior of the MCMC estimates is a more complicated problem we shall not undertake here. In this subsection we shall mainly be concerned with showing that the estimates are approximately normally distributed with the correct mean. The next subsection examines the behavior of the estimated Fisher information and the approximate confidence intervals which may be formed using it.

First we shall examine the following quantities for each sample of 500 estimates. *Bias*, the difference between the estimator sample mean and the known true value of the parameter; *skewness*, the skewness coefficient $g_1 = m_3/m_2^{3/2}$; *kurtosis*, the excess kurtosis coefficient $g_2 = m_4/m_2^2 - 3$. Here m_2, m_3 , and m_4 are the estimator sample moments computed about the mean. To test whether the estimator is normally distributed, the coefficients of skewness and excess kurtosis may be compared with their null distributions under sampling from a normal distribution: When the simulation sample size, N , is big enough, g_1 and g_2 are approximately normally distributed with

mean zero and standard deviation $(6/N)^{1/2}$ and $(24/N)^{1/2}$ respectively (Ratkowsky, 1983, p. 25–26). In our simulation study, $N = 500$, and thus the standard deviations of g_1 and g_2 are 0.11 and 0.22, respectively.

Place of Figure 2

Place of Tables 1, 2 and 3

The sampling distributions of $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}$ are summarized in Tables 1, 2, and 3. These tables show that the estimation error (both bias and variance) and the skewness increase as the spatial interaction increases. There is no such apparent trend for the kurtosis coefficients. We display Q-Q normal plots of the three parameter estimates in two cases, one with $\gamma = 0.2$ and the other with $\gamma = 0.8$ representing a small and moderately large value of γ respectively. The parameters (β_0, β_1) were equal to $(1.0, 2.0)$ in both cases. These plots show that our MCMC estimators are approximately normally distributed, although some skewness and kurtosis exist. In summary, in our simulations the normal approximation to the distribution of the MCMC estimates appears reasonable so long as the spatial interaction γ is not too large.

A number of things start to go wrong when the value of γ becomes large. First, as γ increases, the correlation between successive iterations of the Gibbs sampler increases. Eventually, this correlation becomes large enough so that a sample size of 1000 with the “warm-up” and “spacing” values chosen in our simulations is no longer appropriate. (We need to either enlarge our sample size or increase the warm-up and spacing

values.) Second, when γ exceeds some critical value a phase transition occurs and the realizations of the Gibbs sampler tend toward states in which nearly all of the values y_i are the same. Third, when γ is large, the maximum pseudolikelihood estimate performs poorly and is no longer suitable for us to use as the value ψ for generating our Monte Carlo samples. Because of these reasons, in our simulations the performance of the MCMC MLE deteriorates as γ increases. We have tried values of γ larger than those we report in our tables and the deterioration in performance continues and accelerates. The first and third problems mentioned above can be dealt with by fine-tuning our MCMC procedure and improving our method for choosing ψ . Thus, it may be possible to improve the performance for large γ to some degree.

The range of γ values for which the MCMC MLE performs well increases with the value of β_1 . One reason for this may be that the critical value of γ at which a phase transition occurs increases with the value of β_1 . As one indication of this trend we give Table 4. We ran a number of simulations (similar to those described above) holding β_0 fixed at 1.0 and varying the values of β_1 and γ . When γ was sufficiently large, the MCMC MLE would fail to exist for many of the Monte Carlo samples, that is to say the procedure described earlier, using the MPLE or the first step θ_1 of the Newton-Raphson iteration as ψ , did not produce a Monte Carlo sample for which the MCMC MLE was well-defined. For each value of β_1 that we used, Table 4 records the largest value γ for which the MCMC MLE existed for all samples. This largest value increases steadily with β_1 .

Place of Table 4

4.1 Asymptotic Variance and Confidence Intervals

For a maximum likelihood estimate, the inverse of the Fisher information matrix may be used as a rough estimate of the covariance matrix of the parameter estimates. In many

situations this estimate can be shown to be asymptotically valid, that is, it becomes progressively more accurate as the “sample size” or “amount of data” increases. For the autologistic regression model, the MCMC estimate of the Fisher information matrix is given in equation (3.3). We shall use the diagonal elements of the inverse of this matrix to estimate the variances of our parameter estimates. We shall call this type of variance estimate the *asymptotic variance*. For each set of parameter values used in our simulation study, we generated 500 data sets and computed MCMC estimates for each of them. For each parameter, the sample variance of the 500 estimates supplies us with a value we shall refer to as the *empirical variance*. The empirical variance will be a fairly accurate estimate of the true variance of our parameter estimates. In this subsection we shall compare the asymptotic and empirical variances.

Table 5 summarizes the asymptotic variance and empirical variance for the different parameter cases. It shows that the mean of the asymptotic variance estimates is usually very close to the empirical variance. The standard deviation of the asymptotic variance estimates is relatively small when γ is small, but increases steadily with the value of γ . Thus the asymptotic variance seems to be a roughly unbiased estimator of the actual variance, but one that is highly variable for larger values of γ . In spite of this difficulty at large values of γ , the confidence intervals (described below) based on the asymptotic variance estimates work very well even for larger values of γ . On the whole, the asymptotic variance estimate is a good substitute for the empirical variance, and also it is easier to apply to practical problems.

Place of Table 5

Let $\hat{\theta}_i$ be the estimate of some parameter θ , and $\hat{\sigma}_i^2$ be the asymptotic variance estimate for that parameter obtained from the i^{th} data set in one of our simulation cases. Let $r_i = (\hat{\theta}_i - \theta)/\hat{\sigma}_i$. We call r_i the standardized estimate. We have examined

the normality of r_i and found that the distribution of r_i is closer to normal (in fact, $N(0, 1)$) than that of the original estimate $\hat{\theta}_i$. The sampling distribution of the standardized estimates in the two cases $(\beta_0, \beta_1, \gamma) = (1, 2, 0.2)$ and $(\beta_0, \beta_1, \gamma) = (1, 2, 0.8)$ is summarized in Table 6. The corresponding Q-Q normal plots are shown in Figure 3.

Place of Figure 3

Place of Table 6

We may form confidence intervals for each parameter based on the approximate $N(0, 1)$ distribution of the values $r_i = (\hat{\theta}_i - \theta)/\hat{\sigma}_i$, i.e., we take the $(1 - \alpha)$ confidence interval of θ to be $\hat{\theta}_i \pm z_{\alpha/2}\hat{\sigma}_i$, where $\hat{\theta}_i$ is the MCMC estimate, $z_{\alpha/2}$ is the normal quantile corresponding to the $(1 - \alpha)$ confidence level, and $\hat{\sigma}_i$ is the asymptotic standard deviation estimate.

For each parameter case, we have generated 500 data sets and obtained the 500 corresponding sets of parameter estimates and confidence intervals. Table 7 provides the actual coverage probabilities of the 99%, 95%, 90%, 70%, 50%, 30% and 10% confidence intervals. The smaller confidence levels are included only to test the normal approximation. In order to test that the coverage probabilities are equal to the nominal confidence level, we have used the exact Binomial two-sided test. For the level of significance $\alpha = 0.05$, the critical values for the various confidence levels are: 99%, (0.980, 0.998); 95%, (0.930, 0.969); 90%, (0.872, 0.926); 70%, (0.659, 0.740); 50%, (0.456, 0.544); 30%, (0.260, 0.341); 10%, (0.074, 0.128).

Place of Table 7

Table 7 shows that the actual coverage probabilities of our confidence intervals are close to the nominal confidence levels. In all cases, the difference between the actual coverage probability and the nominal confidence level is insignificant at the level $\alpha = 0.05$. Since these excellent results hold for a broad range of values for γ , we are optimistic that these confidence intervals will be reliable in most applications of the autologistic regression model.

Some interesting research which in many ways parallels our own was carried out by Graham (1994). His work deals with hypothesis testing for a variety of autologistic models: he considers both bi-directional (asymmetric) and second-order autologistic models in addition to the simple first-order autologistic models used in the present paper. However, Graham (1994) does *not* include covariates in his models. Also, Graham studies hypothesis testing whereas our paper concentrates on the properties of point estimates and confidence intervals. Much of the emphasis in both our paper and Graham (1994) is placed on showing that standard likelihood-based procedures have their usual properties when applied to the autologistic model with a Monte Carlo approximation of the likelihood function. In particular, Graham carries out simulations to verify that both the Monte Carlo likelihood ratio test statistic and the Monte Carlo Wald test statistic have their usual asymptotic chi-squared distributions. Other similarities include the following: both papers use Monte Carlo estimates of the Fisher information matrix, and both papers use the MPLE as the parameter ψ for generating the Monte Carlo samples used to approximate the likelihood function.

5 Application to the Distribution of Plant Species

To illustrate our methods we shall use the distribution data for plant species No. 38 (*Castanea pumila*) introduced in Section 1. The other species can be modeled in a

similar way. For this analysis, we have “digitized” the map of Florida so that it is represented as a collection of 1845 sites forming a subset of a rectangular 68×80 lattice. These 1845 sites form the set \mathcal{S} of Section 2. At each site i we record $y_i = 1$ or 0 according to whether the species was present or absent at this site.

After fitting a variety of models, we found that the distribution of species No. 38 could be explained fairly well in terms of the single climate variable FZF. The general topic of covariate selection will be pursued in future research. For now, we shall simply assume that a model using FZF is appropriate.

5.1 Markov Chain Monte Carlo Estimates

In obtaining our maximum likelihood estimates, we implemented the MCMC as follows: We start the Gibbs sampler from the “zero-state” and take the warm-up period to be 100 and the spacing to be 1. We used two different Monte Carlo sample sizes, 2000 and 10000. For this data, the MPLE is not a good choice for ψ . We obtained our final choice for ψ by a process of successive modification beginning with the MPLE, that is, we generated a sequence of Monte Carlo samples and used each sample to suggest a new value of ψ to be used for the next sample. The new value for ψ was typically the first step θ_1 in the Newton-Raphson iterations for solving (3.2), the same procedure we used in our simulation studies. This approach to choosing ψ is fairly ad hoc; a more systematic approach is the iterative approach with constrained step sizes suggested by Geyer and Thompson (1992). Our parameter estimates are reported in Table 8. From Table 8, we can see that there is little difference between the parameter estimates for the sample sizes 2000 and 10000. In the work below, we use the parameter estimates obtained from the sample of size 2000.

After estimating the parameters in the autologistic model, Monte Carlo samples from this model can be generated using the Gibbs sampler. We can then compare these generated maps with the actual species map. An overall comparison can be obtained by using the model to compute fitted probabilities for each site and comparing these fitted values with the actual data values (0 or 1) at each site. The fitted probabilities are estimated using one long run of our Markov chain, that is, we use the Gibbs sampler

to generate N samples (maps) and take the proportion of “1’s” among the N values for a given site to be the fitted probability for that site. We have used our model to compute fitted probabilities for species No. 38. A gray scale plot of these values is presented in Figure 4. We can see that this closely resembles the observed distribution map in Section 1.

Huffer and Wu (1993) compared the MCMC MLEs in Table 8 with estimates obtained using two other methods: the coding method and maximum pseudo-likelihood method. The fitted probabilities produced by the MCMC estimates were closer to the actual data values than those produced by the other two methods. More precisely, the MCMC estimates are best in the sense of minimizing the sum of the absolute fitting errors.

Place of Table 8

Place of Figure 4

5.2 Bootstrap Study Using Gibbs Sampler

In order to study the distributional properties of the MCMC estimates for Species No. 38, we conducted a parametric bootstrap study using the Gibbs sampler. The estimated parameter vector, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}) = (10.467, -0.0516, 2.5620)$ was used to generate 500 independent samples (maps) of observations using the Gibbs sampler. Using the procedures described earlier, we succeeded in obtaining MCMC estimates for 497 of these samples. These 497 estimates were normalized as in section (4.1). The distribu-

tion of these normalized estimates is summarized in Table 9.

Place of Table 9

Table 9 shows that the normalized MCMC estimates have approximately a standard normal distribution. This suggests that interval estimates based on the estimated Fisher information will be fairly reliable in this situation.

References

- Austin, M. P., Nicholls, A. O., and Margules, C. R. (1990), "Measurement of the Realized Qualitative Niche: Environmental Niches of Five Eucalyptus Species," *Ecological Monographs*, 60(2), 161-177.
- Bartlein, P. J., Prentice, I. C., and Webb, T. (1986), "Climatic Response Surfaces from Pollen Data for Some Eastern North American Taxa," *Journal of Biogeography*, 13, 35-57.
- Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems," *Journal of the Royal Statistical Society (with Discussion)*, Series B, 36, 192-236.
- (1975), "Statistical Analysis of Non-lattice Data," *The Statistician*, 24, 179-195.
- Box, E. O., Crumpacker, D. W., and Hardin E. D. (1993), "A Climatic Model for Location of Plant Species in Florida, U.S.A.," *Journal of Biogeography*, 20, 629-644.
- Cressie, N. (1991), *Statistics for Spatial Data*, New York: John Wiley.
- Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using

- Multiple Sequences (with discussion)", *Statistical Science*, Vol.7, 457–472.
- Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Geyer, C. J. (1991a), "Markov Chain Monte Carlo Maximum Likelihood," *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*(E.M. Keramides, ed.), 156–163.
- (1991b), "Reweighting Monte Carlo Mixtures", *Technical Report No.568, School of Statistics, University of Minnesota*.
- (1992a), "Practical Markov Chain Monte Carlo (with discussion)", *Statistical Science*, Vol.7, No.4, 473–511.
- (1992b), "On the Convergence of Monte Carlo Maximum Likelihood Calculations", *Technical Report No.571, School of Statistics, University of Minnesota*.
- Geyer, C. J. and Thompson, E. A. (1992), "Constrained Monte Carlo Maximum Likelihood for Dependent Data (with discussion)," *Journal of the Royal Statistical Society*, Ser.B 54 657–699.
- Graham, J. (1994), "Monte Carlo Markov Chain Likelihood Ratio Test and Wald Test for Binary Spatial Lattice Data", *Technical Report, Dept. of Statistics, North Carolina State University, Raleigh, NC*.
- Huffer, F. W., and Wu, H. (1993), "Modeling and Regressing Spatial Binary Data with Application to the Distribution of Species," Contributed and presented at the 1993 Joint Statistical Meetings in San Francisco.
- Huntley, B., Bartlein, P. J., and Prentice, I. C. (1989), "Climatic Control of the Distribution and Abundance of Beech (*Fagus L.*) in Europe and North America," *Journal of Biogeography*, 16, 551-560.
- Little, Jr., E. L. (1978), *Atlas of United States Trees*, Volume 5. Florida. Misc.

Publ. No. 1361, USDA Forest Service. Washington, D. C.: U.S. Government Printing Office. 256 maps, with indices of common and scientific names.

Ratkowsky, D. A. (1983), *Nonlinear Regression Modeling*, Marcel Dekker, Inc.

Zhao, L. P., and Prentice, R. L. (1990), "Correlated Binary Regression Using a Quadratic Exponential Model," *Biometrika*, 77, 642–648.

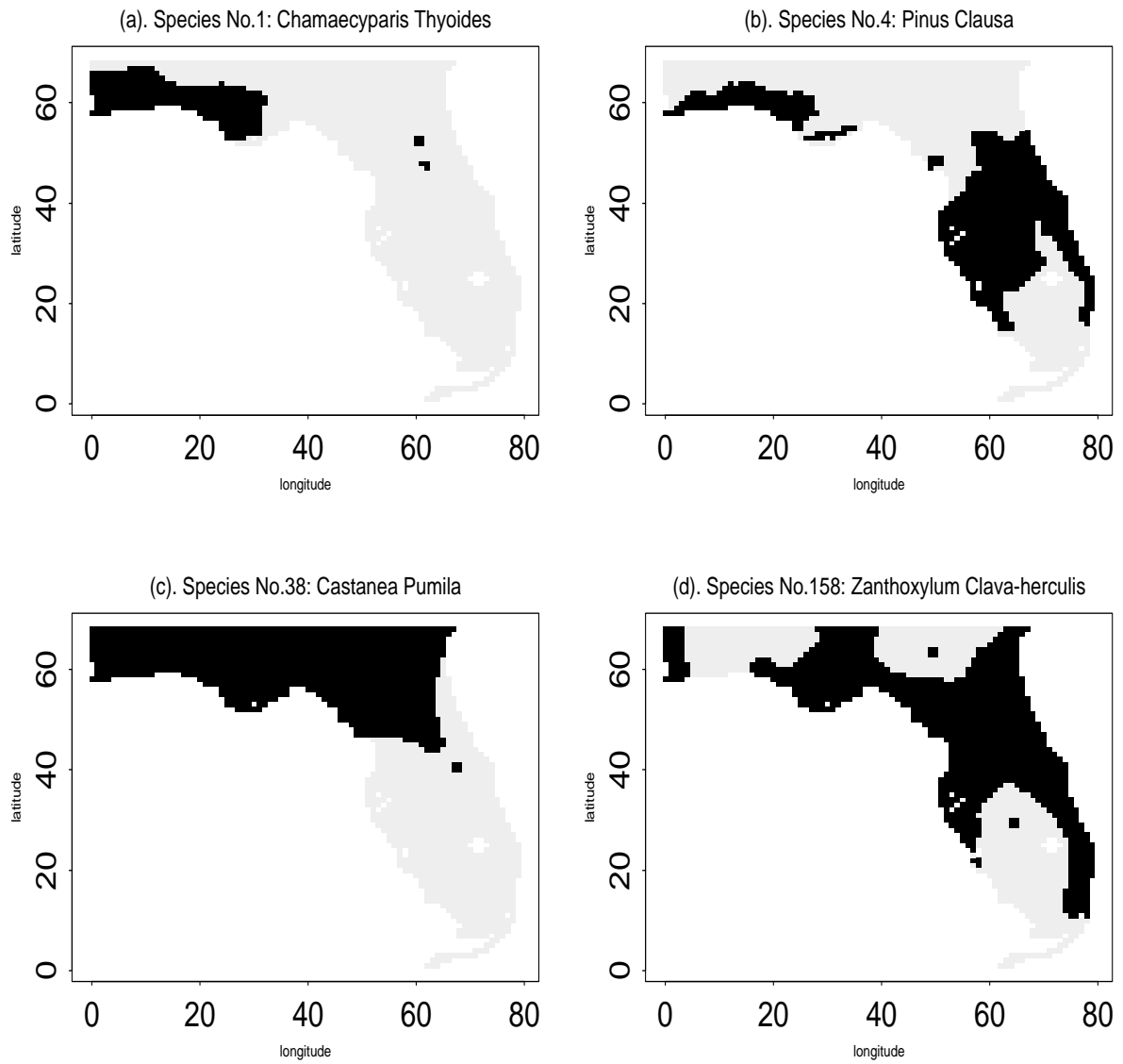


Figure 1: Distribution Maps of Plant Species

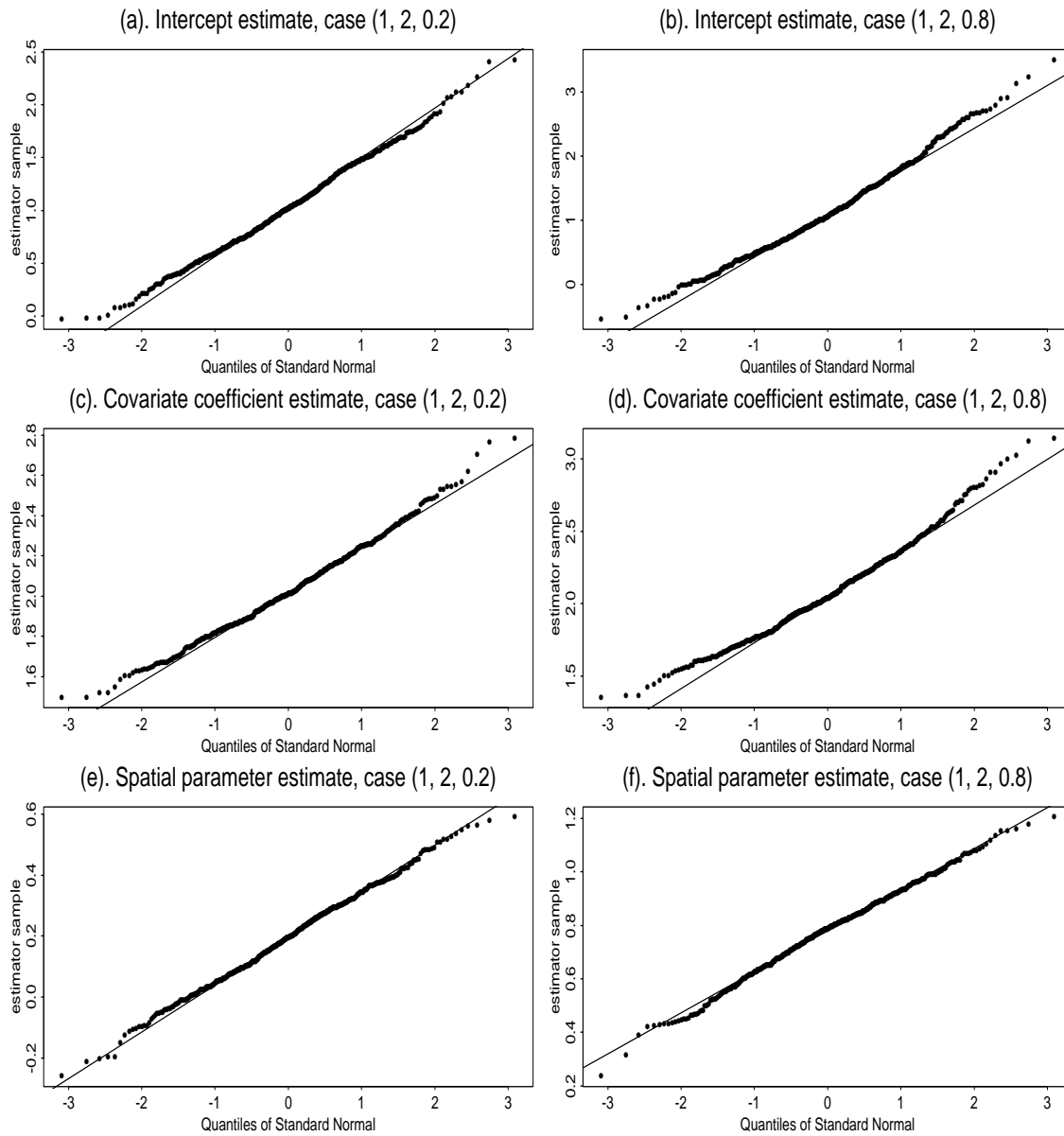


Figure 2: Q-Q Normal Plots of the MCMC Estimators

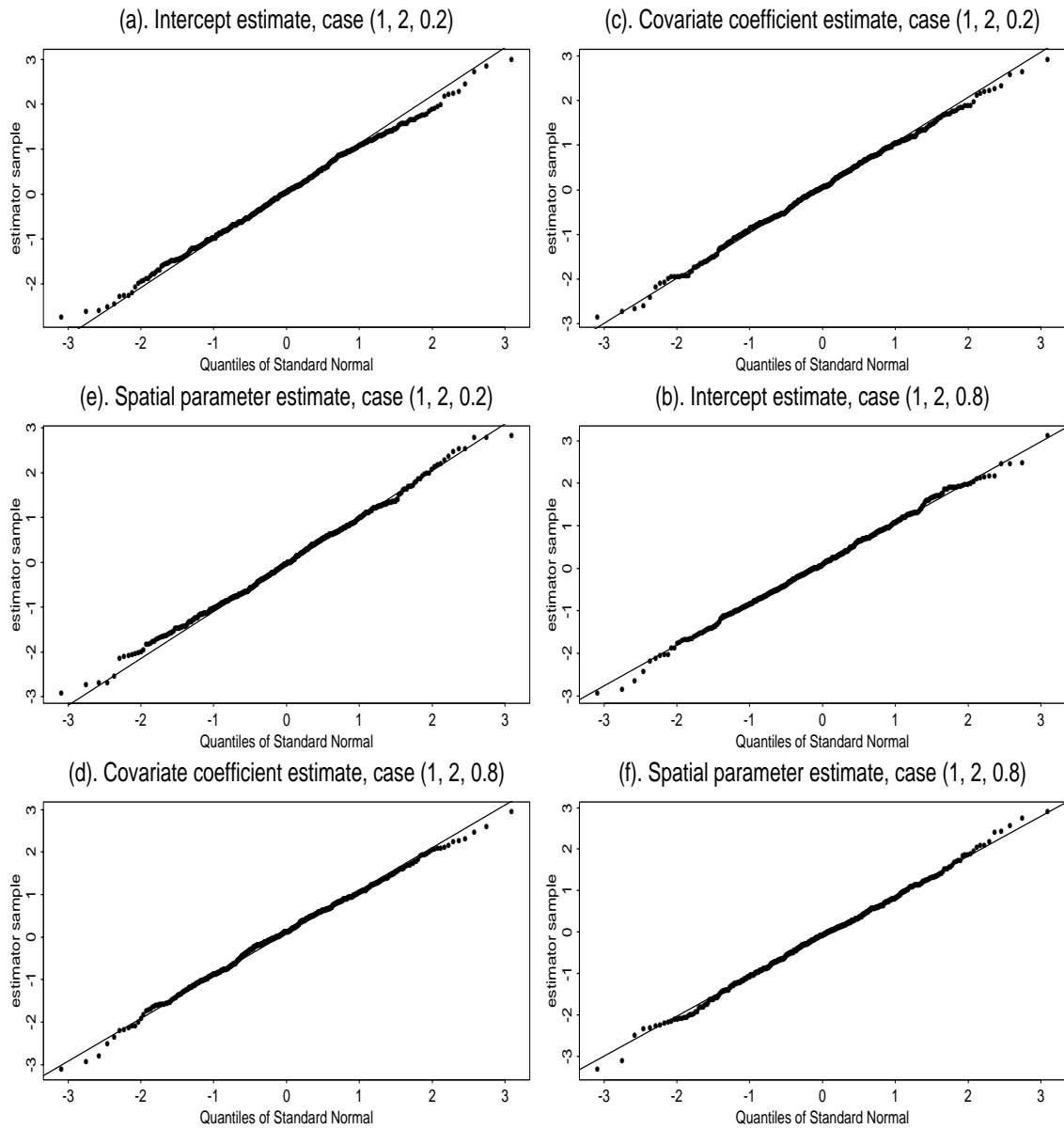


Figure 3: Q-Q Normal Plots of the Standardized Estimates

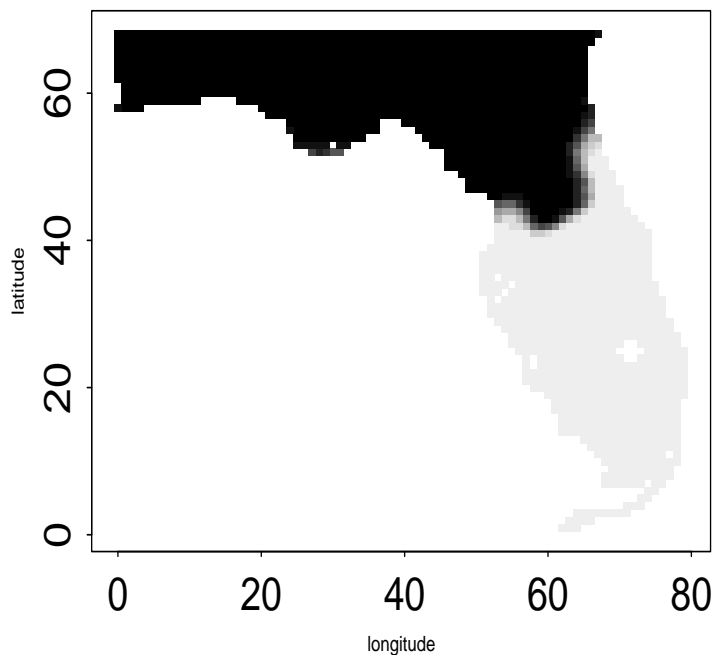


Figure 4: Gray-scale Plot of the Fitted Probabilities for Species No. 38

Table 1: Sampling Distribution of $\hat{\beta}_0$

True $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	1.0063	0.1551	0.0063	0.1318	-0.1748
(1, 2, 0.0)	1.0157	0.3679	0.0157	0.2370*	0.0400
(1, 2, 0.2)	1.0282	0.4346	0.0282	0.1827	-0.1132
(1, 2, 0.4)	1.0488	0.5064	0.0488	0.2824*	-0.0715
(1, 2, 0.6)	1.0463	0.5818	0.0463	0.3036*	-0.2154
(1, 2, 0.8)	1.1282	0.6757	0.1282	0.4348**	0.1161
(1, 2, 1.0) ¹	1.1821	0.8474	0.1821	0.6176**	1.3421**
(1, 2, 1.5)	1.2743	1.1532	0.2743	0.7468**	0.7793**

Table 2: Sampling Distribution of $\hat{\beta}_1$

true $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	2.0116	0.1079	0.0116	-0.0285	0.5295*
(1, 2, 0.0)	2.0183	0.1920	0.0183	0.3239*	0.5432*
(1, 2, 0.2)	2.0242	0.2170	0.0242	0.3307**	0.1626
(1, 2, 0.4)	2.0337	0.2452	0.0337	0.5288**	0.3732
(1, 2, 0.6)	2.0326	0.2736	0.0326	0.4070**	-0.0256
(1, 2, 0.8)	2.0679	0.3087	0.0679	0.5189**	0.3532
(1, 2, 1.0) ¹	2.0892	0.3746	0.0892	0.7865**	1.9230**
(1, 2, 1.5)	2.1225	0.4478	0.1225	0.7754**	0.7375**

Table 3: Sampling Distribution of $\hat{\gamma}$

true $(\beta_0, \beta_1, \gamma)$	Mean	S.D.	Bias	Skewness (g_1)	Kurtosis (g_2)
(1, 2, -1.5)	-1.5059	0.0997	-0.0059	0.2682*	0.7506**
(1, 2, 0.0)	-0.0040	0.1386	-0.0040	-0.1024	-0.1545
(1, 2, 0.2)	0.1939	0.1478	-0.0061	-0.0243	-0.1448
(1, 2, 0.4)	0.3894	0.1530	-0.0106	-0.1066	-0.1122
(1, 2, 0.6)	0.5919	0.1553	-0.0081	-0.1682	-0.3673
(1, 2, 0.8)	0.7773	0.1556	-0.0227	-0.1892	0.0393
(1, 2, 1.0) ¹	0.9722	0.1681	-0.0278	-0.1893	0.0553
(1, 2, 1.5)	1.4845	0.1723	-0.0155	-0.2612*	0.6410*

¹ In this case, there is an outlier. These are the results after deleting the outlier.

* the value exceeds two standard deviations

** the value exceeds three standard deviations

Table 4: Largest Value of γ for Which MCMC MLE Always Existed

Coefficient β_1	0.1	0.5	1.0	1.5	2.0
Critical Value for γ	1.1	1.2	1.8	1.9	2.5

Table 5: Asymptotic Variance and Empirical Variance

γ	-1.5	0.0	0.2	0.4	0.6	0.8	1.0	1.5
E.V.	0.024	0.135	0.189	0.256	0.338	0.457	0.718	1.330
$\hat{\beta}_0$ A.V.	0.027	0.145	0.191	0.256	0.345	0.479	0.658	1.310
S.D.	0.009	0.015	0.024	0.039	0.064	0.108	0.232	1.532
E.V.	0.012	0.037	0.047	0.060	0.075	0.095	0.140	0.200
$\hat{\beta}_1$ A.V.	0.012	0.039	0.049	0.061	0.077	0.099	0.126	0.204
S.D.	0.023	0.006	0.008	0.013	0.018	0.027	0.043	0.169
E.V.	0.010	0.019	0.022	0.023	0.024	0.024	0.028	0.030
$\hat{\gamma}$ A.V.	0.011	0.021	0.022	0.023	0.024	0.026	0.027	0.029
S.D.	0.027	0.001	0.002	0.002	0.003	0.004	0.014	0.061

Notes:

1. In all cases, the true values are $\beta_0 = 1, \beta_1 = 2$.
2. E.V. is the empirical variance.
3. A.V. is the average of the asymptotic variance estimates.
4. S.D. is the standard deviation of the asymptotic variance estimates.

Table 6: Sampling Distribution of Standardized Estimates

Parameter Case	Estimate	Mean	S.D.	Skewness	Kurtosis
(1, 2, 0.2)	$\hat{\beta}_0$	0.0267	0.9941	-0.0393	-0.1893
	$\hat{\beta}_1$	0.0413	0.9769	-0.0991	-0.0593
	$\hat{\gamma}$	-0.0263	1.0073	0.0707	-0.0995
(1, 2, 0.8)	$\hat{\beta}_0$	0.1035	0.9706	-0.0409	-0.0152
	$\hat{\beta}_1$	0.1008	0.9749	-0.1816	0.0647
	$\hat{\gamma}$	-0.1147	0.9794	-0.0331	0.1495

Table 7: Actual Coverage Probability of Confidence Intervals

True	γ	-1.5	0.0	0.2	0.4	0.6	0.8	1.0	1.5
	99%	0.992	0.992	0.988	0.994	0.994	0.992	0.992	0.994
	95%	0.960	0.952	0.960	0.954	0.948	0.956	0.932	0.954
C.I.	90%	0.912	0.902	0.918	0.900	0.904	0.894	0.884	0.914
of	70%	0.696	0.732	0.694	0.680	0.694	0.724	0.677	0.700
β_0	50%	0.532	0.514	0.476	0.480	0.488	0.502	0.481	0.450
	30%	0.354	0.278	0.298	0.298	0.278	0.304	0.293	0.270
	10%	0.126	0.078	0.106	0.096	0.084	0.108	0.084	0.088
	99%	0.986	0.984	0.988	0.992	0.996	0.990	0.992	0.996
	95%	0.946	0.960	0.960	0.958	0.952	0.952	0.944	0.964
C.I.	90%	0.898	0.920	0.900	0.918	0.910	0.916	0.888	0.908
of	70%	0.706	0.738	0.716	0.690	0.706	0.716	0.671	0.690
β_1	50%	0.498	0.514	0.496	0.494	0.490	0.498	0.491	0.496
	30%	0.302	0.322	0.290	0.292	0.296	0.304	0.313	0.290
	10%	0.096	0.118	0.122	0.100	0.094	0.122	0.108	0.086
	99%	0.992	0.992	0.986	0.990	0.996	0.992	0.988	0.984
	95%	0.944	0.956	0.946	0.950	0.950	0.942	0.944	0.950
C.I.	90%	0.898	0.904	0.906	0.892	0.898	0.904	0.876	0.888
of	70%	0.708	0.710	0.696	0.686	0.688	0.708	0.685	0.686
γ	50%	0.504	0.500	0.478	0.484	0.502	0.518	0.497	0.468
	30%	0.318	0.310	0.266	0.296	0.300	0.330	0.269	0.296
	10%	0.114	0.100	0.088	0.104	0.114	0.118	0.080	0.100

Notes:

1. In all cases, the true values are $\beta_0 = 1, \beta_1 = 2$.
2. For case $\gamma = 1.0$, the sample size is 499. One observation is an outlier.

Table 8: Estimation Results of Species No. 38

		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$
Sample Size	Parameter Estimate	10.467	-0.0516	2.5620
2000	Asymptotic S.D	3.1644	0.0100	0.2795
Sample Size	Parameter Estimate	10.444	-0.0515	2.5589
10000	Asymptotic S.D	3.1136	0.0098	0.2745

Table 9: Sampling Distribution of Normalized Estimates for Species No. 38

Estimate	Mean	S.D.	Skewness	Kurtosis
$\hat{\beta}_0$	0.1204	1.0294	-0.2226	0.1942
$\hat{\beta}_1$	-0.1214	1.0327	0.2691	0.1559
$\hat{\gamma}$	-0.0724	0.9545	-0.3976	0.2225