NISS

Estimating Hunting Success Rates via Bayesian Generalized Linear Models

Roger Woodard, Dongchu Sun,
Zhuoqiong He, and Steven L. Sheriff

Estimating Hunting Success Rates via Bayesian Generalized Linear Models

Roger Woodard and Dongchu Sun

University of Missouri-Columbia

Columbia,Missouri 65211

Zhuoqiong He and Steven L. Sheriff

Missouri Department of Conservation

Columbia,Missouri 65201 [1]

Post-season harvest surveys provide data used in the management of Missouri wildlife. These surveys provide information on the number of animals harvested, hunting pressure and hunter success rate. These estimates provide unbiased results at the statewide level due to the large sample size. However, if this survey information is used to make county estimates, poor results often occur due to small sample sizes. To estimate hunter success at the county level for the 1996 Missouri Turkey Hunting Survey, we developed a hierarchical Bayesian model. Specifically, we evaluate a generalized linear model that incorporates linear covariate terms in addition to a conditional auto-regressive structure for spatial correlation. Calculation of the posterior distribution is achieved through Gibbs sampling and adaptive rejection sampling. The inclusion of covariate terms is then evaluated using Bayes factors.

Key Words: Bayes Factor; Conditional auto-regressive model; Inverse gamma distribution; Log-linear mixed model; Spatial correlation.

1 INTRODUCTION

The Missouri Turkey Hunting Survey (MTHS) is a post-season mail survey conducted by the Missouri Department of Conservation to monitor and aid in the regulation of the turkey hunting season. Questionnaires are distributed after the hunting season to a simple random sample of persons who purchased permits to hunt wild turkey during the spring season. For the 1996 turkey hunting season 95,801 persons purchased hunting permits. From these individuals a simple random sample of 6,999 hunters were selected for the survey and 5,005 of these responded. The MTHS

---

provides information concerning the number of turkeys harvested by hunters, the total number of hunter-days used to pursue turkeys, and the hunter success rates. This information is sufficient in supplying usable estimates at the state level. At levels less than statewide in scope, however, accuracy and precision of the estimates become questionable.

For our efforts here, we wish to estimate the hunter success rate in each county of Missouri. Turkey population management focuses on the county level as the smallest unit due to the hunters' ability to indicate where they have hunted in reporting their success. We define hunter success rate as the number of birds harvested per hunter-day(day) of hunting. At the statewide level, the resulting sampled number of days is in excess of 10,000, and this number produces acceptable estimates. After spreading the reported sample days across 114 counties and within two separate weeks of the turkey hunting season, however, sample sizes are extremely low for some counties.

During the sample selection process we cannot determine in which county a particular person hunted. This eliminates the possibility of using stratification to collect an adequate sample size for each county. Using the simple random sample approach, the resulting sample may have some counties with few, if any, respondents. For the 1996 survey four counties had fewer than 10 days of turkey hunting reported in the sample for either of the two weeks of the season. These small sample sizes produce large standard errors at the county level.

To achieve our goal of estimating hunter success rates at the county level we used hierarchical Bayesian methodologies. Hierarchical Bayesian and empirical Bayesian methods are especially applicable to small area estimation problems. Ghosh and Rao (1994) point out that these types of models allow borrowing of strength between the regions. The empirical Bayes methodology has been applied by several authors (Stasny 1991, Tsutakawa 1985, and Tsutakawa 1988). Empirical Bayes methods do not account for the uncertainty in the estimated prior distribution of the parameters (Dempster, Rubin, and Tsutakawa 1981). However, a fully Bayesian approach can account for this uncertainty. Hierarchical and empirical Bayesian methods both suffer from the need for powerful computing resources. Recent advances in techniques, such as Gibbs sampling (Gelfand and Smith 1990), have provided a remedy for this computational problem and allowed widespread use of hierarchical Bayesian methods.

We examine the contribution of a county level covariate when added to a hierarchical Bayesian model. Specifically, we evaluate the capability of the covariate to improve a model which includes random effects for individual counties. Our evaluation is carried out through the use of Markov Chain Monte Carlo (MCMC) methods (Gilks, Richardson, and Spiegelhalter 1996).

Section 2 provides the basic model for response and possible forms for the covariate terms. The MCMC methods utilized to fit these models are described in section 3. Section 4 provides the results of this research, and Section 5 provides some conclusions and insights into the covariate

relationship within the MTHS.

2 Hierarchical Model

The baseline model to which we will add covariate terms has a form similar to the classical mixed model. A mixed model is generally written as

$$\mathbf{v} = \mathbf{X}_1\boldsymbol{\theta} + \mathbf{X}_2\mathbf{z} + \mathbf{e}, \tag{1}$$

where $\boldsymbol{\theta}$ is an unknown vector of *fixed* parameters, $\mathbf{z}$ is an unknown vector of *random* variables, $\mathbf{X}_1$ and $\mathbf{X}_2$ are known design matrices, $\mathbf{v}$ is our vector of response variables, and $\mathbf{e}$ is a vector of random errors. The response vector need not be of simple form, but can be transformed variables. For the MTHS data, $\mathbf{v}$ will become the logit-transformed hunter success rates.

We will let n_{ij} be the number of days hunted in county i during week j. The number of turkeys killed will be denoted y_{ij}. During the 1996 season, each hunter could take only one turkey during each of the two weeks of hunting. We, therefore, included an effect for weeks, (θ_j). Our model also contained a random effect for each county (z_i). This allowed us to model the general heterogeneity among counties and incorporates a spatial correlation.

The number of turkeys taken in a particular county may be considered to have a binomial distribution with parameters n_{ij} and probability of success, p_{ij}, that is

$$y_{ij}|(n_{ij}, p_{ij}) \quad \sim \quad Binomial\,(n_{ij}, p_{ij}). \tag{2}$$

It is this probability of success, p_{ij}, in which we are interested. To model these success rates we used the logit transformation to approximate linearity. Thus, the following model is considered:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_j + z_i + e_{ij}, \quad i = 1, \cdots, I, \, j = 1, 2, \tag{3}$$

where $I = 114$, the number of counties in Missouri, θ_j is the week effect, z_i is the county effect, and e_{ij} terms are random errors for each week-county combination with mean 0 and common variance component δ_0. We use this error term to account for those components which are not explicitly listed in the model. We assumed a normal distribution for these error terms.

2.1 Priors for Model Components

Priors for a Bayesian model should be chosen with knowledge of the inherent nature of the data. When inherent knowledge is not available we attempt to assign a reasonable prior which is relatively flexible. For instance, preliminary examination of the data indicated that the hunter success rates during the two weeks may differ. The statewide success rates for week 1 and week 2 were 10.25% and 6.8%, respectively. To model this difference we assign each week effect a normal distribution with unique mean and variance. Thus, we assume $\theta_j \sim Normal\,(\mu_j, \delta_j), \, j = 1, 2$.

2.1.1 Random County Effects

Turkey habitat and behavior are not restricted by political boundaries. Similarly, turkey hunting methods are not unique to a particular county. We expected hunter success rates in neighboring counties in Missouri to be quite similar. This similarity may give rise to spatial correlation between neighboring counties. In fact, He and Sun (1998) found that the MTHS had a significant spatial correlation between counties.

The normal distribution is often used to model the random effect in a mixed model. We apply the normal distribution but introduce a spatial correlation structure. Several possibilities are available to model this spatial correlation. Whittle (1954) proposed a simultaneous auto-regressive (AR) model for dealing with spatial correlation. It was also introduced by Ord (1975). He and Sun (1998) used this simultaneous AR model for the MTHS data. This model assumes that $\mathbf{z} = (z_1, \ldots, z_I)'$ has a multivariate normal distribution with correlations represented through the equations:

$$z_i = \rho \sum_{m=1}^{I} C_{im} z_m + \epsilon_i, \quad i = 1, \cdots, I, \tag{4}$$

where C_{im} is an indicator variable of the two counties sharing a common boundary. That is, the elements of $\mathbf{C}$ take on the value of 1 if regions i and m share a common boundary and 0, otherwise. This approach uses the intuitive idea that adjoining areas should have similar responses. This approach considered the counties with common boundaries to be spatially correlated.

Alternatively, we apply the conditional auto-regressive (CAR) model of Clayton and Kaldor (1987), which is a special case of the model presented by Besag (1974).

In general, a *CAR Model* is specified by the full conditional densities,

$$f(z_i|z_m, m \neq i) = \left(\frac{a_i}{2\pi\delta_z}\right)^{\frac{1}{2}} \exp\left\{-\frac{a_i}{2\delta_z}\left(z_i - \sum_{m\neq i}^{I} \beta_{im} z_m\right)^2\right\}, \tag{5}$$

for $i = 1, \cdots, I$ and where δ_z is a common variance component for spatial regions. Let $\mathbf{B}$ be the $I \times I$ matrix with diagonal elements a_i and im^{th} off-diagonal elements $-a_i\beta_{im}$. Besag (1974) proved that if $\mathbf{B}$ is symmetric and positive definite, then the joint probability density of $\mathbf{z}$ exists and is given by

$$f(\mathbf{z}) = (2\pi\delta_z)^{-I/2} |\mathbf{B}|^{1/2} \exp\left\{-\frac{1}{2\delta_z}\mathbf{z}'\mathbf{B}\mathbf{z}\right\}. \tag{6}$$

We let

$$\mathbf{B} = \mathbf{I} - \rho\mathbf{C}, \tag{7}$$

where ρ is a measure of spatial correlation, $\mathbf{I}$ is the identity matrix, and $\mathbf{C}$ is an adjacency matrix as was previously defined. We restrict the common spatial correlation to $\lambda_1^{-1} < \rho < \lambda_I^{-1}$, where

4

λ_1 is the minimum eigenvalue of $\mathbf{C}$ and λ_I is the largest eigenvalue of $\mathbf{C}$. As noted by Cressie (1993, p. 471), if this condition holds then $\mathbf{B}$ is positive definite. Sun, Tsutakawa and Speckman (in press) show the conditional density of z_i is of the form (5), or equivalently,

$$(z_i | z_m, m \neq i) \sim N\left(\rho \sum_{m \in \Lambda_i} z_m, \delta_z\right),\tag{8}$$

where Λ_i is the set of counties adjacent to county i.

2.2   Other Priors

At this stage in the modeling process we also specified a prior distribution for the variance components. The common prior for the variance of a normal distribution is the inverse gamma distribution (Gelman, Carlin, Stern, and Rubin 1995). For e_{ij} in equation (3), we specified $\delta_0 \sim IG(a_0, b_0)$. We also had a prior for the common variance component(δ_z) of the random county effects. This was considered a third stage to the hierarchical model. Again, we used the inverse gamma distribution and set $\delta_z \sim IG(a_z, b_z)$. The densities of δ_0 and δ_z were of the form

$$\pi(\delta_k) \propto \frac{1}{\delta_k^{a_k+1}} \exp(-b_k/\delta_k), \quad \delta_k > 0, \quad k = 0 \text{ or } z.\tag{9}$$

The remaining hyper-parameters, such as (a_0, b_0) and (a_z, b_z), were considered fixed. Finally, we assume that ρ is uniformly distributed on the interval $(\lambda_1^{-1}, \lambda_I^{-1})$.

2.3   Covariate Terms

Our model can be applied to the hunter success rates obtained by the MTHS. It is, however, possible that hunter success rates may be more accurately modeled using covariates. These covariates may be obtained from a known biological relationship with the dependent variable. Covariates may be incorporated to reduce the variance of the random effects. These covariates take the form of additional linear terms in the model. As an example of a covariate, we use the proportion of a given county covered by forest (Giessman, Barney, Haithcoat, Myers, and Massengale 1986). We refer to this covariate as forest coverage (s_i).

2.3.1   Exploratory Data Analysis

Prior to fitting the actual models, including the covariates, we used exploratory techniques to examine the relationship between the covariate and the hunter success rate produced by the Bayesian CAR model without covariates. We begin by examining a scatter plot of the covariate with the estimated success rates for each county (Fig. 1). It would seem there is little or no linear trend. If a linear relationship were present, it would appear to be negative. The scatter plots in

Figure 1 also may indicate some curvature in the relationship, which might be modeled using a quadratic formulation of the covariate.

2.3.2   Covariate as a linear term

We include the forest coverage covariate and use η to indicate the common parameter. The model presented in (3) can be generalized to include several covariate terms and takes on the form

$$v_{ij} \equiv \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_j + \mathbf{w}_i'\boldsymbol{\eta} + z_i + e_{ij}, \quad i = 1, \cdots, I, j = 1, 2, \tag{10}$$

where $\mathbf{w}_i$ is a vector of forest coverage covariates and $\boldsymbol{\eta}$ is the corresponding coefficients. For forest coverage, this allows use of either a simple linear model for the covariate term or the inclusion of higher order relationships, such as a quadratic term. Covariates and week effects can easily be combined into a single design matrix. Thus, in our model the week and forest coverage effects may be considered fixed effects in the sense of a mixed model. This model would be considered our first stage of the hierarchical model. To define the second stage, prior distributions for each component of our linear model must be addressed. We retain the priors for other model components defined above and place a multivariate normal distribution on the vector $\boldsymbol{\eta}$ with mean vector $\boldsymbol{\xi}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

In addition to the specification of the priors we make assumptions on the nature of the model. Taking the week effects as a vector $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ then, given $(\boldsymbol{\theta}, \boldsymbol{\eta}, \mathbf{z}, \delta_0)$, the parameter p_{ij} is independent of δ_z and ρ. Also, given (δ_z, ρ), $\mathbf{z}$ is independent of $(\boldsymbol{\theta}, \boldsymbol{\eta}, \delta_0)$. Finally, we assume the parameters $(\boldsymbol{\theta}, \boldsymbol{\eta}, \delta_0, \delta_z, \rho)$ are mutually independent. We have little information on appropriate prior distributions for the fixed effects $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. To specify these distributions we first explore the posterior of these distributions using a constant prior. We then use the posterior mean and variance of the resulting posterior to specify the prior of $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. This procedure may fail if the posterior is not a proper distribution.

2.4   Existence of the Posterior

We often do not have subjective information concerning the nature of the data. In these instances we can use a non-informative prior. Non-informative priors, however, do not produce proper Bayes factors. Thus, if we wish to compare several models, we will need Bayes factors. Instead, we apply non-informative priors to obtain a posterior mean and variance. This posterior mean and variance can then be used to construct a proper prior.

When a non-informative prior is used, the posterior distribution may be improper (Hobert and Casella 1996). Under such a situation, an MCMC algorithm would not converge. To examine some properties of the priors used and the resulting posterior distributions, suppose the data

$\mathbf{y} = (y_{11}, \ldots, y_{I1}, y_{12}, \ldots, y_{I2})'$ follow the binomial distribution as in (2), whose first stage prior is given by the linear mixed model (10). Equation (10) can be rewritten as $\mathbf{v} = \mathbf{X}_1 \boldsymbol{\theta} + \mathbf{W} \boldsymbol{\eta} + \mathbf{X}_2 \mathbf{z} + \mathbf{e}$, where

$$
\mathbf{v} = \begin{pmatrix} v_{11} \\ \vdots \\ v_{I1} \\ v_{12} \\ \vdots \\ v_{I2} \end{pmatrix}, \quad
\mathbf{X}_1 = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}, \quad
\mathbf{W} = \begin{pmatrix} \mathbf{w}_1' \\ \vdots \\ \mathbf{w}_I' \\ \mathbf{w}_1' \\ \vdots \\ \mathbf{w}_I' \end{pmatrix}, \quad
\mathbf{X}_2 = \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \end{pmatrix}, \quad
\mathbf{e} = \begin{pmatrix} e_{11} \\ \vdots \\ e_{I1} \\ e_{12} \\ \vdots \\ e_{I2} \end{pmatrix}.
$$

The following fact is a corollary of a theorem from Sun, Tsutakawa and He (1998).

Fact 1 *Suppose that the design matrix $(\mathbf{X}_1, \mathbf{W})$ is of full column rank.*

Assume the following prior distributions: $p(\boldsymbol{\theta}, \boldsymbol{\eta}) \propto 1$, and $(\mathbf{z} \mid \rho, \delta_z)$ has the CAR prior given in (6), where $\mathbf{B}$ is given in (7) and ρ has a uniform prior on $(\lambda_1^{-1}, \lambda_I^{-1})$. Then the joint posterior distribution of $(\mathbf{v}, \boldsymbol{\theta}, \mathbf{z}, \delta_0, \delta_z, \rho)$ given $\mathbf{y}$ exists if the prior of (δ_0, δ_z) (not necessary of the form (9)) is proper.

This theorem is readily applied to our model for the MTHS. Here the design matrix $(\mathbf{X}_1, \mathbf{W}, \mathbf{X}_2)$ does not have full rank, but $(\mathbf{X}_1, \mathbf{W})$ has full rank. The problem is that the vector of covariates for forest coverage $\mathbf{W}$ can be considered a linear combination of the columns of $\mathbf{X}_2$.

3   Computations

Computing the appropriate posterior in this case is not feasible due to the high dimensionality of the integration. Instead, we use MCMC methods such as Gibbs sampling. The use of Gibbs sampling requires sampling from full conditional distributions of each parameter. In this section, we briefly summarize these conditional distributions based on our model.

The following fact gives the full conditional distributions when the prior includes the effect $\boldsymbol{\eta}$ with the corresponding design matrix $\mathbf{W}$. The full conditional distributions without covariates can be derived similarly.

Fact 2 *The full conditional distributions are given as follows:*
(a) Given $(\mathbf{y}, \boldsymbol{\eta}, \boldsymbol{\theta}, \mathbf{z}, \delta_0, \delta_z, \rho)$, $v_{ij} = \log[p_{ij}/(1 - p_{ij})], (i = 1, \ldots, I; j = 1, 2)$, are independent, and have conditional density proportional to

$$
\exp\left\{ v_{ij} y_{ij} - n_{ij} \log(1 + e^{v_{ij}}) - \frac{(v_{ij} - \theta_j - \mathbf{w}_i' \boldsymbol{\eta} - z_i)^2}{2\delta_0} \right\}.
$$

(b) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\eta}, z, \delta_0, \delta_z, \rho)$, θ_1 and θ_2 are independent and the distribution of θ_j is

$$N\left(\frac{\overline{d}_j \delta_j + \delta_0 \mu_j}{2\delta_j + \delta_0}, \ \frac{\delta_j \delta_0}{2\delta_j + \delta_0}\right),$$

where $d_{ij} = v_{ij} - z_i - \boldsymbol{w}_i'\boldsymbol{\eta}$ and $\overline{d}_j = I^{-1}\sum_i^I d_{ij}$.

(c) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}, z, \delta_0, \delta_z, \rho)$, the distribution of $\boldsymbol{\eta}$ is multivariate normal with mean $\boldsymbol{A}_1^{-1}\boldsymbol{G}_1$ and covariance matrix $\boldsymbol{A}_1^{-1}$, where $\boldsymbol{A}_1 = \frac{2}{\delta_0}\boldsymbol{W}'\boldsymbol{W} + \boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{G}_1 = \frac{2}{\delta_0}\boldsymbol{W}'\boldsymbol{t} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\xi}$ with

$$\boldsymbol{t} = \left(\frac{1}{2}\sum_{j=1}^2 (v_{1j} - \theta_j) - z_1, \ \ldots, \ \frac{1}{2}\sum_{j=1}^2 (v_{Ij} - \theta_j) - z_I\right)'. \tag{11}$$

(d) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\eta}, \delta_0, \delta_z, \rho)$, z is multivariate normal with mean $\delta_0^{-1}\boldsymbol{A}_2^{-1}\boldsymbol{G}_2$ and covariance matrix $\boldsymbol{A}_2^{-1}$, where $\boldsymbol{A}_2 = \frac{2}{\delta_0}\boldsymbol{I} + \frac{1}{\delta_z}\boldsymbol{B}$ and

$$\boldsymbol{G}_2 = \left(\sum_{j=1}^2 (v_{1j} - \theta_j - \boldsymbol{w}_1'\boldsymbol{\eta}), \ldots, \sum_{j=1}^2 (v_{Ij} - \theta_j - \boldsymbol{w}_I'\boldsymbol{\eta})\right)'. \tag{12}$$

(e) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\eta}, z, \delta_z, \rho)$, δ_0 has an inverse gamma distribution with parameters

$$a_0 + I \quad and \quad b_0 + \frac{1}{2}\sum_{i,j}(v_{ij} - \theta_j - \boldsymbol{w}_i'\boldsymbol{\eta} - z_i)^2.$$

(f) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\eta}, z, \delta_0, \rho)$, δ_z has an inverse gamma distribution with parameters

$$a_z + 1 \quad and \quad b_z + \frac{1}{2}\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z}.$$

(g) Given $(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{\eta}, z, \delta_0, \delta_z)$, ρ has a conditional density proportional to

$$|\boldsymbol{B}|^{1/2}\exp\left\{-\frac{1}{2\delta_z}\boldsymbol{z}'\boldsymbol{B}\boldsymbol{z}\right\}.$$

These distributions allow the use of Gibbs sampling for producing our Bayesian estimates of the parameters. The distributions of the v_{ij} and ρ are not a standard distribution from which we can easily obtain samples. These distributions can be sampled using the adaptive rejection sampling procedures of Gilks and Wild (1992). This methodology is very straight forward for distributions which are log-concave as shown by Berger and Sun (1993). It can be shown that the conditional densities of v_{ij} and ρ are indeed log-concave. For v_{ij}, the second derivative of the log density is given by:

$$-n_{ij}e^{v_{ij}}(1 + e^{v_{ij}}) - \frac{1}{\delta_0}.$$

This value is negative for values of v_{ij} and thus the distribution is log-concave. For ρ, we note that the second derivative of the log density is given by:

$$-\frac{1}{2}\sum_{i=1}^I \frac{\lambda_i^2}{(1 - \rho\lambda_i)^2},$$

where $(\lambda_1, \cdots, \lambda_I)$ are the eigenvalues of the adjacency matrix $\boldsymbol{C}$. This derivative is obviously negative. We can use the log-concavity of the density to simplify sample generation.

4 Results

Our goal is to examine covariate effects in modeling the success rate of turkey hunters in counties of Missouri. The covariate of interest is the proportion of a county covered by woodland. Biologically, forest coverage may be linked with the number of turkeys in a given area. It is reasonable to assume that the hunter success rate might also be linked with forest coverage. It is also thought that the number of turkeys has a quadratic relationship with forest coverage. We thus considered both linear and quadratic relationships between hunter success rate and forest coverage.

4.1 Models

We considered three models for these data (Table 1). We examined a simple model with no covariate as a baseline to examine the covariate effect. This model included the fixed week and the random county effects but no covariates. For the second model we start with a simple linear relationship and include the covariate forest coverage (s_i). A third model that included both a linear term and a quadratic term of the forest coverage was also fit.

4.2 Hyper-parameters

To implement our Gibbs sampling routine we needed to specify the appropriate hyper-parameters. We had little practical knowledge of the situation so use of priors that were informative was not feasible. We first used non-informative priors to produce estimates of the posterior means and variances. The resulting posterior means and variances allowed us to establish more appropriate priors. We then used these proper priors for estimation and calculation of Bayes factors.

For example, in Model 2 we needed the prior for the covariate term. The posterior mean and variance for this term from the non-informative prior were found to be -0.1 and 0.2. We then set the mean of the covariate prior at -0.1 and the variance at 0.8, or four times the variance. For Model 3, we have both a linear and a quadratic term. As before, we used the posterior results from the non-informative prior to set the mean and variance of the linear term at 0.0075 and 0.008 respectively. Similarly for the quadratic term, we set the prior mean to -0.0107 and variance to 0.07 using the posterior results from the non-informative priors.

For other hyper-parameters, we followed similar procedures. For the week effects, θ_j we assigned $\mu_1 = -2.2$, $\mu_2 = -2.6$ and $\delta_1 = \delta_2 = 1.0$. For the variance components, we set the priors on the inverse gamma distributions with shape parameters a_0 and a_z, to 2.5 and the scale parameters b_0 and b_z to 0.1 for both the component from the random effects and the overall error. Experimentation with a variety of hyper-parameters found very similar results regardless of the hyper-parameters.

9

4.3 Model Comparison

To compare the various models we use Bayes factors. The Bayes factor gives the ratio of the posterior odds to the prior odds in comparing two models. To use this, we had to specify prior odds for particular model comparisons. In this instance, we took a neutral point of view and choose equal probability for both models. The Bayes factor for Model h versus Model k is $B_{hk} = P(\mathbf{y}|M_h)/P(\mathbf{y}|M_k)$. The difficulty with the Bayes factor is calculating $P(\mathbf{y}|M_k)$. Several methods are available for calculating this quantity.

The simplest method is the harmonic mean estimator for the marginal likelihood based on Gibbs outputs presented by Newton and Raftery (1994). Unfortunately, as it was noted by Kass and Raftery (1995) and others, the estimated marginal likelihood is often dominated by a few values with small likelihood. Alternatively, we use the bridge sampling method of Meng and Wong (1996). This method makes use of a random sample $\boldsymbol{\gamma}(k) = (\boldsymbol{\gamma}_{k1}, \cdots, \boldsymbol{\gamma}_{kn_k})$ from the posterior produced by Gibbs sampling for Model k. In our case $\boldsymbol{\gamma}_{km} = (\mathbf{v}^{(m)}, \boldsymbol{\theta}^{(m)}, \mathbf{z}^{(m)}, \delta_0^{(m)}, \delta_z^{(m)}, \rho^{(m)})$, $m = 1, \ldots, n_k$, from the output of Gibbs sampling for Model k. Meng and Wong (1996) presented an iterative algorithm for obtaining an optimal estimate of the Bayes factor. Specifically, at the $(g+1)st$ iteration the estimate of the Bayes factor of Model h to Model k is given by

$$\hat{B}_{hk}^{(g+1)} = \frac{\frac{1}{n_k}\sum_{m=1}^{n_k}\frac{l_{km}}{d_h l_{km}+d_k \hat{B}_{hk}^{(g)}}}{\frac{1}{n_h}\sum_{m=1}^{n_h}\frac{1}{d_h l_{hm}+d_k \hat{B}_{hk}^{(g)}}}, \tag{13}$$

where $d_h = 1 - d_k = n_h/(n_h + n_k)$, $l_{km} = q_h(\boldsymbol{\gamma}_{km})/q_k(\boldsymbol{\gamma}_{km})$, $l_{hm} = q_h(\boldsymbol{\gamma}_{hm})/q_k(\boldsymbol{\gamma}_{hm})$, and q_k is the product of the likelihood and the marginal prior density of $\boldsymbol{\gamma} = (\mathbf{v}, \boldsymbol{\theta}, \mathbf{z}, \delta_0, \delta_z, \rho)$ under Model k. We often choose $n_h = n_k$ so that $d_h = d_k = 1/2$.

To see if a linear or quadratic term for forest coverage is necessary, we let $\mathbf{W}_2 = (s_1, \ldots, s_I)'$ and $\mathbf{W}_3$ be the $I \times 2$ matrix, whose ith row is of the form (s_i, s_i^2). We also let $\boldsymbol{\xi}_k$ and $\boldsymbol{\Sigma}_k$ be the prior mean and variance of the corresponding parameters $\boldsymbol{\eta}_k$ under Model k, for $k = 2, 3$. In this instance, we have,

$$\frac{q_1(\boldsymbol{\gamma})}{q_k(\boldsymbol{\gamma})} = (2\pi)^{(1-p_k)/2}|\boldsymbol{\Sigma}_k|^{1/2}|\mathbf{A}_{3k}|^{1/2}\exp\Big(\frac{1}{2}\Big[\boldsymbol{\xi}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\xi}_k - \mathbf{G}_{3k}'\mathbf{A}_{3k}^{-1}\mathbf{G}_{3k}\Big]\Big), \quad k = 2, 3,$$

where p_k is the dimension of $\boldsymbol{\eta}_k$ (i.e., $q_2 = 1$ and $q_3 = 2$), $\mathbf{G}_{3k} = \frac{2}{\delta_0}\mathbf{W}_k'\mathbf{t} + \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\xi}_k$ and $\mathbf{A}_{3k} = \frac{2}{\delta_0}\mathbf{W}_k'\mathbf{W}_k + \boldsymbol{\Sigma}_k^{-1}$. Here $\mathbf{t}$ is defined as in equation (11) and is evaluated using the random sample $\boldsymbol{\gamma}(k)$.

4.3.1 Bayes Factors

The models presented in 4.1 were fit and the Meng and Wong (1996) procedure was used to calculate appropriate Bayes factors. The resulting Bayes factors for comparing Model 2 (the CAR

model with a linear covariate) with Model 1 (CAR model without covariate) was $B_{21} = 2.753$. The value for comparing Model 3 (CAR with quadratic covariates) and Model 1 was $B_{31} = 3.009$. The Bayes factors give the posterior odds in favor of first model. The interpretations outlined by Kass and Raftery (1995) find that Bayes factors between 1 and 3 are "not worth more than a bare mention" and values from 3 to 20 are considered to be "positive" evidence. Here we see that while the Model 2 is probably not an improvement over Model 1, Model 3 may be more effective.

The borderline value of the Bayes factor in comparison of Model 3 with Model 1 necessitates further examination. To support the conclusions of the Bayes factors, we examined the posterior distributions of the covariate coefficients. If the covariates add little to the model, we would expect the coefficients to be near zero. The posterior marginal distributions are displayed in Figure 2. Both covariate coefficients (η_1, η_2) for the quadratic model (Model 3) appear to be centered near zero. Similarly, for the linear covariate model (Model 2) we see that the distribution of η is centered near zero. In light of this information it is unlikely that the covariates improve the model in either form for these data.

4.3.2 Comparison of Success Rates

Another point of comparison among the candidate models concerns the hunter success rates for each county. Figure 3 displays plots of the comparative success rates generated by each model. From these plots it appears that success rates of the covariate models were very similar to the simple CAR model for the 1996 turkey data. Figure 4 also gives maps of the success rates for under the Bayesian CAR model as well as the simple frequency estimators.

4.4 Bayesian Estimators

The Gibbs sampling routine produced estimates for the posterior distributions of the week and county effects. The posterior means and standard deviations for Model 1 are given in Table 2. Other posterior parameter estimates of interest include the variance components and the spatial correlation. Model comparisons given above imply that our spatial correlation should be non-zero. This conclusion is supported by the posterior mean and standard deviation for the correlation coefficient ρ. As expected the random county effects had a significant variance δ_z. The estimated posterior densities are displayed in Figure 5.

5   Comments

The forest coverage covariate did not seem to provide substantive improvement over the simpler CAR models for the 1996 MTHS data. This result is somewhat unexpected. Examination of scatter

plots seemed to indicate a general relationship between the covariate and success rates.

A possible explanation of this general relationship may be attributed to the county effects z_i. In fitting models with the forest coverage covariate but without the county effects, we found very different posterior estimates for the covariate parameters. Table 3 provides a comparison of the posterior estimates and standard deviations of the covariate effects with and without the random county effects.

For the quadratic model without the inclusion of county effects, a 95% confidence interval would not contain zero. In contrast the model which contained random county effects had a 95% confidence interval that contained zero. This would indicate that the forest coverage covariate may be useful if one chooses not to fit random county effects.

In general, Bayesian estimates provided a smoother fit than standard frequency estimators used in standard survey methods. We did find that even in counties where no data were available or where no successes were reported in the sample, we had a positive estimate. These Bayesian estimates seem more reasonable in such situations.

REFERENCES

Berger, J. O., and Sun, D. (1993), "Bayesian Analysis for the Poly-Weibull Distribution," *Journal of the American Statistical Association* 88, 1412-1418.

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with Discussion)," *Journal of the Royal Statistical Society*, Ser. B, 36, 192-236.

Clayton, D., and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671-681.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981), "Estimation in Covariance Components Models," *Journal of the American Statistical Association,* 76, 341-353.

Gelfand, A., and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 298-409.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995), *Bayesian Data Analysis*, London, U.K.: Chapman and Hall.

Ghosh, M., and Rao, J. N. K. (1994), "Small Area Estimations: an Appraisal," *Statistical Science*, 9, 65-93.

Giessman, N. F., Barney, T. W., Haithcoat, T. L., Myers, J. W., and Massengale, R. (1986), "Distribution of Forestland in Missouri," *Transactions, Missouri Academy of Science*, 20, 5-14.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice*, London,U.K.: Chapman and Hall.

Gilks, W. R., and Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics*, 41, 337-348.

He, Z., and Sun, D. (1998), "Hierarchical Bayes Estimation of Hunting Success Rates with Spatial Correlations," submitted.

Hobert, J. P., and Casella, G. (1996), "The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models," *Journal of the American Statistical Association*, 91, 1461-1473.

Kass, R., and Raftery, A. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773-794.

Meng, X., and Wong, W. H. (1996), "Simulating Ratios of Normalizing Constants Via A Simple Identity: A Theoretical Exploration," *Statistica Sinica* 6, 831-860.

Newton, M. A., and Raftery, A. E. (1994), "Approximate Bayesian Inference by Weighted Likelihood Bootstrap (with Discussion)," *Journal of the Royal Statistical Society*, Ser. B, 56, 3-48.

Ord, K. (1975), "Estimation Methods for Models with Spatial Interaction," *Journal of the American Statistical Association*, 70, 120-126.

Stasny, E. A. (1991), "Hierarchical Models for the Probabilities of a Survey Classification and Nonresponse: an Example from the National Crime Survey," *Journal of the American Statistical Association*, 86, 296-303.

Sun, D., Tsutakawa, R. K., and He, Z. (1998), "Propriety of Posteriors with Improper Priors in Hierarchical Linear Mixed Models," submitted.

Sun, D., Tsutakawa, R. K., and Speckman, P. L. (in press), "Posterior Distribution of Hierarchical Models Using CAR (1) Distributions," *Biometrika*.

Tsutakawa, R. K. (1985), "Estimation of Cancer Mortality Rates: a Bayesian Analysis of Small Frequencies." *Biometrics*, 41, 69-79.

Tsutakawa, R. K. (1988), "Mixed Model for Analyzing Geographic Variability in Mortality Rates," *Journal of the American Statistical Association*, 83, 37-42.

Whittle, P. (1954), "On Stationary Process in the Plane," *Biometrika*, 41, 434-449.

Table 1: Models

Model	Description	Form
1	CAR model without covariate	$\theta_j + z_i + e_{ij}$
2	CAR model with linear covariate	$\theta_j + \eta s_i + z_i + e_{ij}$
3	CAR model with quadratic covariate	$\theta_j + \eta_1 s_i + \eta_2 s_i^2 + z_i + e_{ij}$

Table 2: Posterior Estimates of Parameters for Model 1 (CAR model without covariates included.)

Parameter	Posterior Mean	Posterior Std. Dev.
θ_1	-2.2040	.0791
θ_2	-2.6705	.0816
δ_0	0.0601	.0151
δ_z	0.0222	.0079
ρ	0.1663	.0039

Table 3: Covariate Estimates (Std. Dev) With and Without County Effects

Model	Linear Term	Quadratic Term
$\theta_j + \eta s_i + e_{ij}$	$-130.66(75.45)$	$-$
$\theta_j + \eta s_i + z_i + e_{ij}$	$0.0204(0.0413)$	$-$
$\theta_j + \eta_1 s_i + \eta_2 s_i^2 + e_{ij}$	$2090.56(270.96)$	$-3855.05(441.00)$
$\theta_j + \eta_1 s_i + \eta_2 s_i^2 + z_i + e_{ij}$	$0.0132(0.0414)$	$-0.0257(0.1222)$

Figure 1: Scatter plots of logit success rates obtained from the CAR model $logit(p_{ij}^B)$ and the forest coverage s_i for week 1 (a) and week 2 (b).

Figure 2: Relative frequency histograms of posterior distributions of the covariate terms from (a) the linear term of Model 2, (b) the linear term of Model 3 and (c) the quadratic term of Model 3.

Figure 3: Comparisons of the predicted success rates $(\tilde{p}_{ij}^B)$ for (a) Model 2, and (b) Model 3 with those of Model 1 (p_{ij}^B).

Figure 4: Maps of naive frequency estimators of $p_{ij}^F = y_{ij}/n_{ij}$ and Bayesian estimators p_{ij}^B: (a): frequency estimators of p_{i1}^F for Week 1; (b): frequency estimators of p_{i2}^F for Week 2; (c): Bayesian estimators of p_{i1}^B for Week 1; (d): Bayesian estimators of p_{i2}^B for Week 2. Here the counties in white are those where $n_{ij} = 0$ so that p_{ij}^F do not exist.

Figure 5: Relative frequency histograms from posterior distributions produced by Model 1 for the week effects (a) θ_1, and (b) θ_2, the variance components (c) δ_0, and (d) δ_z, and the common spatial correlation (e) ρ.

(a)

(b)

(a)

(b)

(c)

η

η_1

η_2

17

(a)

(b)

(a)   (b)

(c)   (d)

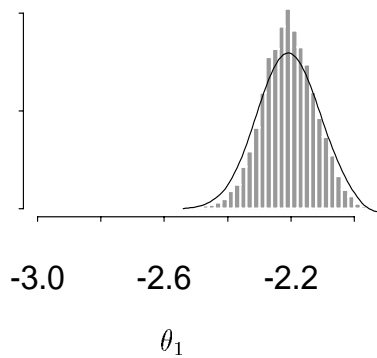|		
0 - 0.04	0.08 - 0.12	0.16 - 0.2
0.04 - 0.08	0.12 - 0.16	> 0.2

(a)

(b)

(d)

(c)

(e)