Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 783494, 8 pages
http://dx.doi.org/10.1155/2014/783494

Research Article

Bayesian Analysis for Dynamic Generalized Linear Latent Model with Application to Tree Survival Rate

Yu-sheng Cheng,[1] Mei-wen Ding,[1] Ye-mao Xia,[2] and Wen-fa Zhan[1]

[1] *School of Mathematics & Computation Sciences, Anqing Normal University, Anqing 246011, China*
[2] *Department of Applied Mathematics, Nanjing Forestry University, Nanjing 210037, China*

Correspondence should be addressed to Yu-sheng Cheng; chengyshaq@163.com

Received 8 January 2014; Revised 10 June 2014; Accepted 18 June 2014; Published 7 July 2014

Academic Editor: Jen-Tzung Chien

Logistic regression model is the most popular regression technique, available for modeling categorical data especially for dichotomous variables. Classic logistic regression model is typically used to interpret relationship between response variables and explanatory variables. However, in real applications, most data sets are collected in follow-up, which leads to the temporal correlation among the data. In order to characterize the different variables correlations, a new method about the latent variables is introduced in this study. At the same time, the latent variables about AR (1) model are used to depict time dependence. In the framework of Bayesian analysis, parameters estimates and statistical inferences are carried out via Gibbs sampler with Metropolis-Hastings (MH) algorithm. Model comparison, based on the Bayes factor, and forecasting/smoothing of the survival rate of the tree are established. A simulation study is conducted to assess the performance of the proposed method and a pika data set is analyzed to illustrate the real application. Since Bayes factor approaches vary significantly, efficiency tests have been performed in order to decide which solution provides a better tool for the analysis of real relational data sets.

1. Introduction

Logistic regression model, a widely appreciated model for analyzing categorical response data especially for binary/dichotomous data in social science research, marketing, biomedical studies, and ecology, has been received a substantive attention [1, 2]. The primary concern of logistic regression analysis is to build the interrelationships between explanatory variables and responses and explain the variability of the response probability in terms of the variability in the observed covariates. The common analysis for the logistic regression model is based on the independent subjects, and statistical inference is carried out via the maximum likelihood approach (see [3–5]).

However, in practice, many studies need to account for correlations among the items within time as well as correlations among the same items between times. Independent assumptions cannot capture such correlations. In order to interpret the correlations among the responses, one popular choice is to introduce the latent variables. Even more, taking into account the temporal correlations, a dynamic system is established which leads to the latent variable models [6–8]. However, this dynamic system is associated with dynamic factor. So, a Bayesian approach is proposed in this paper. With an alternative approach about the ML, the development is based on the generalized logistic model in the well-known linear dynamic factor analysis model. To cope with this complicated model, Gibbs sampler [9] with MH algorithm [10, 11] is implemented to generate a sequence of observations from the appropriate joint posterior distribution. The joint Bayes estimates of the parameters and latent variables scores associated with standard error estimates can be obtained directly based on the simulated observations. Beyond the estimation problem, another main objective of this paper is to introduce the Bayes factor (see [12]) to test various hypotheses about the posited model. In general, the computation of the Bayes factor involving multiple integrals is rather intractable and, more often, are evaluated via various computational methods (see [13–19], and among others). Among easy-to-construct, we adopted the well-known path sampling technique [20] for computing Bayes factor. The nice feature of path sampling lies in its simplicity in implementation and effectiveness over such an important sampling (e.g., [21]) and bridge sampling [22].

A basic intent of modeling dynamic model is to represent series of observations generated in time and to predict the future evolution of the series. The well-known Kalman recursion [23] is commonly employed for forecasting and provides the optimal forecasts in the case where the state transitions are linear and dynamic system is Gaussian. However, if Gaussian assumptions, in particular, of the distributions of the measurement errors are violated, Kalman recursion only yields the best linear predictor for linear dynamic system which is substantially different from the optimal forecasts. Many authors have suggested various alternatives to the Kalman filter; see, among others, Kitagawa [24], Meinhold and Singpurwalla [25], Carlin et al. [26], and Smith and West [27]. Motivated by the key idea in Carlin et al. [26], in this paper, the missing data is treated as the future observations and augmented them with parameters of latent variables as observed data. Therefore, it can use Bayesian analysis approach to deal them. Simulated observations generated by Gibbs sampler from the joint posterior distribution can be directly applied to forecasts and/or smoothing.

This paper is organized as follows. In Section 2, we describe the problem under consideration. Section 3 gives the estimation procedure and explores the Bayesian approach in detail. For model selection, Bayes factor and Bayesian forecasting are introduced in Section 4. Simulation study and a real example are given in Section 5. Some concluding remarks are presented in Section 6.

2. Materials and Methods about Model

In this section, the problem and model structure are presented as follows, which will be considered throughout the paper. For $i = 1, \ldots, T$, $j = 1, \ldots, K$, let y_{ij} denote the number of living tree within the jth plot in the ith month. We assume that y_{ij} follows the binomial distribution with the unknown survival rate p_{ij}

$$\left[y_{ij} \mid p_{ij}\right] \sim Bin\left(n_{ij}, p_{ij}\right), \tag{1}$$

where n_{ij} is the total number of tree being investigated at time i in the jth plot. Further, the probability p_{ij} is in relation to the covariates via the following link function:

$$\log it\left(p_{ij}\right) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \lambda_j \xi_i + \varepsilon_{ij}, \tag{2}$$

in which "log *it*" denotes the log *it* function, $\mathbf{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijr})^T$ is a $(r+1)$ vector of fixed covariates, $\boldsymbol{\beta}$ is a $(r+1)$ regression parametric vector, ξ_i is the common factor, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iK})'$ is the unique error with distribution $N_K(\mathbf{0}, \boldsymbol{\Psi}_\varepsilon)$ where $\boldsymbol{\Psi}_\varepsilon = \operatorname{diag}\{\psi_{\varepsilon 1}, \ldots, \psi_{\varepsilon p}\}$ is the diagonal matrix. The factor loading vector is $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_K)'$, which measures the effect of ξ_i on the functions of the response probability. Moreover, it assumes that ξ_i satisfies the following one-order autoregressive model

$$\xi_i = \phi \xi_{i-1} + \delta_i, \tag{3}$$

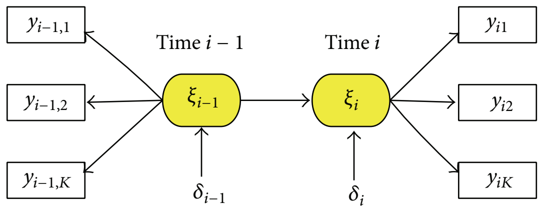

Figure 1: Path diagram about interaction with the latent variables and manifest variables.

where $|\phi| < 1$ and δ_i is the iid random errors distributed according to the normal distribution $N(0, \sigma^2)$ $(\sigma^2 > 0)$. For the initial variable ξ_0, we assume that $\xi_0 \sim N(\mu_0, \sigma_0^2)$ with mean μ_0 and variance σ_0^2. δ_i and $\boldsymbol{\varepsilon}_{ij}$ are mutually independent and both are independent of ξ_i. In this paper, we treat μ_0 and σ_0^2 to be known and fix them at some preassigned values.

Let $\vartheta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \lambda_j \xi_i + \varepsilon_{ij}$, $b(\vartheta_{ij}) = \log(1 + \exp(\vartheta_{ij}))$. Then, (1), (2), and (3) can be reformulated as follows:

$$\begin{gathered} p\left(y_{ij} \mid \vartheta_{ij}\right) = C_{n_{ij}}^{y_{ij}} \exp\left\{y_{ij}\vartheta_{ij} - n_{ij} b\left(\vartheta_{ij}\right)\right\}, \\ \vartheta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \lambda_j \xi_i + \varepsilon_{ij}, \quad \boldsymbol{\varepsilon}_i \sim N\left(\mathbf{0}, \boldsymbol{\Psi}_\varepsilon\right), \\ \xi_i = \phi \xi_{i-1} + \delta_i, \\ \xi_0 \sim N\left(\mu_0, \sigma_0^2\right), \qquad \delta_i \sim N\left(0, \sigma^2\right), \\ \delta_i \perp \varepsilon_i, \quad \xi_0 \perp \varepsilon_i, \quad \delta_i \perp \xi_0. \end{gathered} \tag{4}$$

The main aim of introducing the latent variables is to characterize the temporal correlation between the consecutive observations and explain the interrelationships among the response probability. To give more details, note that for $i > i'$, and $j, l = 1, \ldots, K$, the correlation between $\vartheta_{i'j}$ and ϑ_{il} is given by

$$\operatorname{Corr}\left(\vartheta_{i'j}, \vartheta_{il}\right) = \frac{\lambda_j \lambda_l \sigma_\xi^2(i) \phi^{i-i'}}{\sqrt{\lambda_j \sigma_\xi^2 + \psi_{\varepsilon j}} \sqrt{\lambda_l^2 \sigma_\xi^2 + \psi_{\varepsilon l}}}, \tag{5}$$

in which

$$\sigma_\xi^2(i) = \sigma_0^2 \phi^{2i} + \sigma^2 \frac{1 - \phi^{2i}}{1 - \phi^2}. \tag{6}$$

A path diagram of latent variables and manifest variables at time $i-1$ and i is given in Figure 1. Following the conventions of path diagrams, the ellipses represent the latent variables, the rectangles denote the observed measurements, and the arrows represent the direct effect.

For ease of exposition, let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iK})'$, $\boldsymbol{\vartheta}_i = (\vartheta_{i1}, \ldots, \vartheta_{iK})'$, and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iK})'$, which denote the vector of random observations, the vector of latent variables, and the vector of covariates cross K items at time i, respectively. Let $\mathbf{Y} = \{\mathbf{y}_i : i = 1, \ldots, T\}$ be the collection of the observed data, let $\boldsymbol{\Omega} = \{\xi_i : i = 0, \ldots, T\}$ be the collection of the factor variables, let $\boldsymbol{\Xi} = \{\boldsymbol{\vartheta}_i : i = 1, \ldots, T\}$ be the collection of the latent variables, and let $\boldsymbol{\theta}$ be the vector of unknown parameters in $\boldsymbol{\beta}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Psi}_\varepsilon$, ϕ, and σ^2 involved in the model and then the joint distribution of $(\mathbf{Y}, \Xi, \boldsymbol{\Omega})$ conditioning on $\boldsymbol{\theta}$ is given by

$$p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}\boldsymbol{\theta}) = p(\mathbf{Y} \mid \Xi, \boldsymbol{\theta})\, p(\Xi \mid \boldsymbol{\Omega}, \boldsymbol{\theta})\, p(\boldsymbol{\Omega} \mid \boldsymbol{\theta}) = \left(\prod_{i=1}^{T}\prod_{j=1}^{K} p(y_{ij} \mid \vartheta_{ij})\, p(\vartheta_{ij} \mid \xi_i, \boldsymbol{\theta})\right) \times p(\xi_0) \prod_{i=1}^{T} p(\xi_i \mid \boldsymbol{\theta}). \tag{7}$$

In the framework of Bayesian analysis, the following priors for unknown parameter vector $\boldsymbol{\theta}$ are used to complete Bayesian specifications: $p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}) p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\varepsilon) p(\phi) p(\sigma^2)$ and

$$\begin{gathered} \boldsymbol{\beta} \sim N_{r+1}(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0), \qquad \phi \sim \text{Uniform}(-1, 1), \\ \sigma^{-2} \sim \text{Gamma}(\alpha_0, \tau_0), \\ (\boldsymbol{\Lambda} \mid \boldsymbol{\Psi}_\varepsilon) \sim \prod_{k=1}^{K} N(\boldsymbol{\Lambda}_{0k}, \psi_{\varepsilon k}\mathbf{H}_{0k}), \\ \boldsymbol{\Psi}_\varepsilon \sim \prod_{k=1}^{K} \text{Gamma}(\alpha_{\varepsilon 0k}, \beta_{\varepsilon 0k}), \end{gathered} \tag{8}$$

in which $\boldsymbol{\beta}_0$, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Lambda}_{0k}$, $\mathbf{H}_{0k}$, α_0, τ_0, $\alpha_{\varepsilon 0k}$, and $\beta_{\varepsilon 0k}$ are the known superparameters.

3. Parameters Estimation via Gibbs Sampler with MH Algorithm

For the Bayesian analysis, we are required to evaluate the complicated posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{Y})$ which involves the high-dimensional integrals. Data augmentation [28] technique is used to cope with the posterior analysis in relation to the complicated $p(\boldsymbol{\theta} \mid \mathbf{Y})$. Specifically, the observed data $\mathbf{Y}$ are augmented with the latent quantities $\{\Xi, \boldsymbol{\Omega}\}$ in the posterior analysis. A sequence of random observations will be generated by the Gibbs sampler [9], coupled with the Metropolis-Hastings algorithm [10, 11] from the joint posterior distribution $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y})$, specifically, at the mth iteration with current values $(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)}) = (\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Lambda}^{(m)}, \boldsymbol{\Psi}_\varepsilon^{(m)}, \phi^{(m)}, \sigma^{2(m)})$. We do the following:

(i) draw $\Xi^{(m+1)}$ from $p(\Xi \mid \boldsymbol{\Omega}^{(m)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Lambda}^{(m)}, \boldsymbol{\Psi}_\varepsilon^{(m)}, \phi^{(m)}, \sigma^{2(m)}, \mathbf{Y})$;

(ii) draw $\boldsymbol{\Omega}^{(m+1)}$ from $p(\boldsymbol{\Omega} \mid \Xi^{(m+1)}, \boldsymbol{\beta}^{(m)}, \boldsymbol{\Lambda}^{(m)}, \boldsymbol{\Psi}_\varepsilon^{(m)}, \phi^{(m)}, \sigma^{2(m)}, \mathbf{Y})$;

(iii) draw $\boldsymbol{\beta}^{(m+1)}$ from $p(\boldsymbol{\beta} \mid \Xi^{(m+1)}, \boldsymbol{\Omega}^{(m+1)}, \boldsymbol{\Lambda}^{(m)}, \boldsymbol{\Psi}_\varepsilon^{(m)}, \phi^{(m)}, \sigma^{2(m)}, \mathbf{Y})$;

(iv) draw $(\boldsymbol{\Lambda}^{(m+1)}, \boldsymbol{\Psi}_\varepsilon^{(m+1)})$ from $p(\boldsymbol{\Lambda}, \boldsymbol{\Psi}_\varepsilon \mid \Xi^{(m+1)}, \boldsymbol{\Omega}^{(m+1)}, \boldsymbol{\beta}^{(m+1)}, \phi^{(m)}, \sigma^{2(m)}, \mathbf{Y})$;

(v) draw $\phi^{(m+1)}$ from $p(\phi \mid \Xi^{(m+1)}, \boldsymbol{\Omega}^{(m+1)}, \boldsymbol{\beta}^{(m+1)}, \boldsymbol{\Lambda}^{(m+1)}, \boldsymbol{\Psi}_\varepsilon^{(m+1)}, \sigma^{2(m)}, \mathbf{Y})$;

(vi) draw $\sigma^{2(m+1)}$ from $p(\sigma^2 \mid \Xi^{(m+1)}, \boldsymbol{\Omega}^{(m+1)}, \boldsymbol{\beta}^{(m+1)}, \boldsymbol{\Lambda}^{(m+1)}, \boldsymbol{\Psi}_\varepsilon^{(m+1)}, \phi^{(m+1)}, \mathbf{Y})$.

It has been shown [9, 29] that under mild conditions and, for sufficiently large m, such as J, the joint distribution of $(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)})$ converges at an exponential rate to the desired posterior distribution $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y})$. Hence, $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y})$ can be approximated by the empirical distribution of $\{(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)}) : m = J+1, \ldots, J+M\}$, where M is chosen to give sufficient precision to the empirical distribution. The convergence of the hybrid Gibbs sampler with MH algorithm can be monitored by the "estimated potential scale reduction (EPSR)" values as suggested by Gelman and Rubin [30]. To implement Gibbs sampler with MH algorithm, it requires the full conditional distributions for each parameters and latent quantities. For saving space, these technical details are omitted.

Simulated observations obtained from the posterior values can be used for statistical inference via straightforward analysis procedure. For brevity, let $\{(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)}) : m = 1, \ldots, M\}$ be the random observations of $(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta})$ generated by the Gibbs sampler from $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y})$. The joint Bayesian estimate of $\boldsymbol{\theta}$ and $\boldsymbol{\Omega}$ can be obtained easily via the corresponding sample means of the generated observations as follows:

$$\widehat{\boldsymbol{\theta}} = \sum_{m=1}^{M} \frac{\boldsymbol{\theta}^{(m)}}{M}, \qquad \widehat{\boldsymbol{\Omega}} = \sum_{m=1}^{M} \frac{\boldsymbol{\Omega}^{(m)}}{M}. \tag{9}$$

The consistent estimates of covariance matrix of $(\boldsymbol{\Omega}, \boldsymbol{\theta})$ can be obtained as follows:

$$\begin{aligned} \text{Cov}(\widehat{\boldsymbol{\theta}}) &\approx \sum_{m=1}^{M} \frac{(\boldsymbol{\theta}^{(m)} - \widehat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(m)} - \widehat{\boldsymbol{\theta}})'}{(M-1)}, \\ \text{Cov}(\widehat{\boldsymbol{\Omega}}) &\approx \sum_{m=1}^{M} \frac{(\boldsymbol{\Omega}^{(m)} - \widehat{\boldsymbol{\Omega}})(\boldsymbol{\Omega}^{(m)} - \widehat{\boldsymbol{\Omega}})'}{(M-1)}. \end{aligned} \tag{10}$$

4. Model Selection and Bayesian Forecasting

Model selection is an important issue for Bayesian generalized logistic analysis since it is of interest to determine whether the latent variables are involved or whether the significant effects exist among the competing covariates. Consider the problem of comparing competing models M_1 and M_0 which are nested or nonnested. Let $p(\mathbf{Y} \mid M_1)$ and $p(\mathbf{Y} \mid M_0)$ denote the marginal density of the data $\mathbf{Y}$ under M_1 and M_0, respectively. A popular choice for selecting models (e.g., [16, 18]) is achieved via the following Bayes factor:

$$\text{BF}_{10} = \frac{p(\mathbf{Y} \mid M_1)}{p(\mathbf{Y} \mid M_0)} = \frac{\int p(\mathbf{Y}, \Xi, \boldsymbol{\Omega} \mid \boldsymbol{\theta}, M_1)\, p(\boldsymbol{\theta} \mid M_1)\, d\boldsymbol{\theta}\, d\Xi\, d\boldsymbol{\Omega}}{\int p(\mathbf{Y}, \Xi, \boldsymbol{\Omega} \mid \boldsymbol{\theta}, M_0)\, p(\boldsymbol{\theta} \mid M_0)\, d\boldsymbol{\theta}\, d\Xi\, d\boldsymbol{\Omega}}, \tag{11}$$

Table 1: The evidence about Bayes factor against M_0 [18].

$\log\mathrm{BF}_{10}$	BF_{10}	Evidence against M_0
0 to 1	1 to 3	Not worth more than a bare mention
1 to 3	3 to 20	Positive
3 to 5	20 to 150	Strong
>5	>150	Very strong

where $p(\boldsymbol{\theta} \mid M_k)$ and $p(\mathbf{Y}, \Xi, \boldsymbol{\Omega} \mid \boldsymbol{\theta}, M_k)$ are the prior density of $\boldsymbol{\theta}$ and the joint probability density of $(\mathbf{Y}, \Xi, \boldsymbol{\Omega})$ given the values of $\boldsymbol{\theta}$ under M_k, respectively. Kass and Rafter [18] provide the following categories to furnish appropriate guidelines in Table 1.

Computing BF_{10} or $\log \mathrm{BF}_{10}$ involves the high-dimensional integrations of likelihood with respect to $(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta})$ which is rather difficult. Various techniques are explored to address this problem (see [7]). Among easy-to-construct, we consider the path sampling method [20]. The core of path sampling is to construct a series of linked models which link up the competing models.

Specifically, consider a class of densities indexed by a continuous parameter t in $[0, 1]$ such that $t = 0$ corresponding to M_0 and $t = 1$ to M_1. Note that

$$p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y}, t) = \frac{1}{m(t)} p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid t), \tag{12}$$

where $m(t)$ is the normalizing constant of $p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid t)$ given by

$$m(t) = m(\mathbf{Y} \mid t) = \int p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid t)\, d\Xi\, d\boldsymbol{\Omega}\, d\boldsymbol{\theta}. \tag{13}$$

A path using the parameter t in $[0, 1]$ is constructed so that $\mathrm{BF}_{10} = m(1)/m(0)$. Taking logarithm and differentiating (13) with respect to t give

$$\begin{aligned}\frac{d \log m(t)}{dt} &= \int \frac{d}{dt} \log p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid t)\, p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y}, t)\, d\Xi\, d\boldsymbol{\Omega}\, d\boldsymbol{\theta} \\ &= E_{\Xi,\Omega,\theta} \mathbf{U}(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{Y}, t),\end{aligned} \tag{14}$$

in which $E_{\Xi,\Omega,\theta}[\mathbf{U}(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{Y}, t)]$ denotes the expectation with respect to the distribution $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y}, t)$ and

$$\mathbf{U}(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{Y}, t) = \frac{d \log p(\mathbf{Y}, \Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid t)}{dt}. \tag{15}$$

Hence,

$$\begin{aligned}\log \mathrm{BF}_{10} &= \log \frac{m(1)}{m(0)} = \int_0^1 E_{\Xi,\Omega,\theta} \mathbf{U}(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta}, \mathbf{Y}, t)\, dt \\ &\cong \frac{1}{2} \sum_{s=0}^{S} \left(\overline{\mathbf{U}}_{s+1} + \overline{\mathbf{U}}_s\right) \Delta t_s,\end{aligned} \tag{16}$$

where $\{t_s : s = 0, \ldots, S+1\}$ are the fixed grids such that $t_0 = 0 < t_1 < \cdots < t_{s+1} = 1.0$, $\Delta t_s = t_{s+1} - t_s$, and

$$\overline{\mathbf{U}}_s = \frac{1}{M} \sum_{m=1}^{M} \mathbf{U}\left(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)}, \mathbf{Y}, t_s\right), \tag{17}$$

in which $\{(\Xi^{(m)}, \boldsymbol{\Omega}^{(m)}, \boldsymbol{\theta}^{(m)}) : m = 1, \ldots, M\}$ are observations sampled from $p(\Xi, \boldsymbol{\Omega}, \boldsymbol{\theta} \mid \mathbf{Y}, t)$. The MCMC method proposed for estimation produced can be directly applied to simulate the above observations for computing BF_{10} or $2\log \mathrm{BF}_{10}$.

With all the complete conditionals available for sampling, a predictive or estimated values for a future $\mathbf{y}_{T+1}$ can be obtained, provided that $\mathbf{X}_{T+1}$ was available, which offers a solution to the so-called forecasting problem. More generally, the predictions of the future observations $\mathbf{Y}^* = \{\mathbf{y}_{T+1}, \ldots, \mathbf{y}_{T+t0}\}$ can be obtained in terms of the observed data $\mathbf{Y}$ and the covariates $\{\mathbf{X}_1, \ldots, \mathbf{X}_{T+t0}\}$. As mentioned in introduction, we treat $\mathbf{Y}^*$ as missing data and augment them with $\boldsymbol{\Omega}^* = \{\xi_0, \ldots, \xi_{T+t0}\}$, and $\Xi^* = \{\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\theta}_{T+t0}\}$ and data $\mathbf{Y}$ in the posterior analysis. Gibbs sampler now is implemented to sample observations from the posterior $p(\mathbf{Y}^*, \Xi^*, \boldsymbol{\Omega}^*, \boldsymbol{\theta} \mid \mathbf{Y})$. Predictive values for $\mathbf{Y}^*$ are given by $\mathbf{Y}^* = E[\mathbf{Y}^* \mid \mathbf{Y}]$ which can be solved by

$$\widehat{\mathbf{Y}}^* = E\left[\mathbf{Y}^* \mid \mathbf{Y}\right] = \sum_{m=1}^{M} \frac{\mathbf{Y}^{*(m)}}{M}, \tag{18}$$

where $\{\mathbf{Y}^{*(m)} : m = 1, \ldots, M\}$ are the random observations generated by the Gibbs sampler with MH algorithm from $p(\mathbf{Y}^*, \Xi^*, \boldsymbol{\Omega}^*, \boldsymbol{\theta} \mid \mathbf{Y})$. Consistent estimates of covariance matrix of (18) can be obtained via sample covariance matrix.

5. Results of Experiment

5.1. Simulation Study. A simulation study is presented to assess the performance of our proposed procedure. The data set is simulated from the model defined by (1), (2), and (3) with $T = 100$, $K = 4$, and $r = 3$. The true population values of the unknown parameters are taken as $\boldsymbol{\beta} = (-0.6, -0.7, 0.6)'$, $\boldsymbol{\Lambda} = (1.0, 1.2, 1.0, 1.8)'$, $\phi = 0.5$, $\sigma^2 = 1.0$, $\mu_0^* = 0.0$, and $\sigma_0^{2*} = 1.0$ in which elements with asterisk are treated as fixed known parameters for model identification (see [6]). For $i = 1, \ldots, 100$, $j = 1, \ldots, 4$, we first sampled independently the observations from the normal distribution $N(\mathbf{0}, \mathbf{I}_3)$ to form the covariates $\mathbf{x}_{ij}$, where $\mathbf{I}_k$ is a $k \times k$ identity matrix, and then generate ξ_i from model (3) with $\xi_0 = 1.0$. Based on these settings, we generate the observations $\mathbf{Y} = \{y_{ij} : i = 1, \ldots, 100, j = 1, \ldots, 4\}$ from the binomial distribution Bin $(50, p_{ij})$ with $n_{ij} = 50$. Priors inputs for the hyperparameters in the prior distributions are taken as (I) $\boldsymbol{\beta}_0 = \mathbf{0}$, $\boldsymbol{\Sigma}_0 = 1000\mathbf{I}_3$, $\boldsymbol{\Lambda}_{0k} = \mathbf{0}$, $H_{0k} = 1000$, $\alpha_0 = 2.0$, $\tau_0 = 2.0$, $\alpha_{\varepsilon 0k} = 4.0$, and $\beta_{\varepsilon 0k} = 4.0$; (II) $\boldsymbol{\beta}_0$ and $\boldsymbol{\Lambda}_{0k}$ are set to be equal to the true values, respectively; $\boldsymbol{\Sigma}_0 = \mathbf{I}_3$, $H_{0k} = 1.0$, $\alpha_0 = 4.0$, $\tau_0 = 4.0$, $\alpha_{\varepsilon 0k} = 9.0$, and $\beta_{\varepsilon 0k} = 8.0$. Note that prior (I) approaches the noninformative prior and prior (II) gives the informative prior.

At first, four parallel methods are conducted for different starting values as a pilot study to obtain some idea about the

Table 2: Statistical results of Gibbs under prior (I) and prior (II) with relevant parameters.

Para.	Priors					
	(I)			(II)		
	BIAS	RMS	SD	BIAS	RMS	SD
β_0	0.0127	0.0456	0.0508	0.0228	0.0510	0.0467
β_1	0.0267	0.0526	0.0469	0.0231	0.0533	0.0431
β_2	−0.0202	0.0472	0.0490	−0.0246	0.0455	0.0451
λ_2	0.0510	0.0943	0.1114	0.0160	0.0839	0.0964
λ_3	0.0294	0.0971	0.1087	0.0154	0.0880	0.0961
λ_4	0.0178	0.0739	0.0822	−0.0063	0.0545	0.0714
$\psi_{\varepsilon 1}$	0.1617	0.1701	0.1017	0.0573	0.0923	0.0953
$\psi_{\varepsilon 2}$	0.1845	0.1928	0.1110	0.0591	0.0821	0.1015
$\psi_{\varepsilon 3}$	0.1791	0.1869	0.1096	0.0595	0.0851	0.1020
$\psi_{\varepsilon 4}$	0.1313	0.1412	0.0909	0.0255	0.0569	0.0829
ϕ	−0.0115	0.0694	0.0664	−0.0110	0.0692	0.0662
σ^2	−0.1013	0.1944	0.1905	−0.0345	0.1464	0.1725

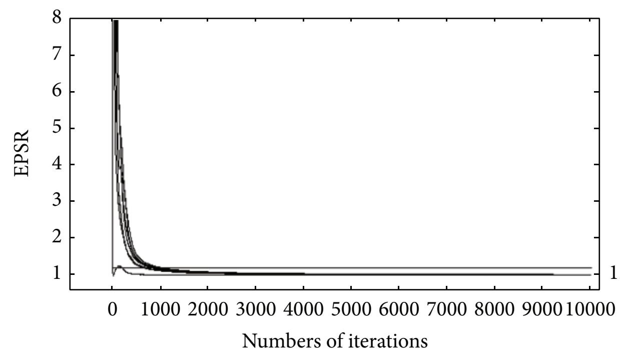

Figure 2: EPSR values change with numbers of iterations under prior (I).

number of the Gibbs sampler iterations in getting convergence. For MH algorithm, the acceptance rate is also adjusted in posterior sampling to give about 40% [31]. Hence, in the following analysis, experiment always takes such values for the tuned parameters. Figure 2 gives the estimated potential scaled reduction (EPSR) values [30] of the parameters against the number of the iterations in Gibbs sampler under prior (I).

In Figure 2, it can be seen clearly that all these EPSR values are less than 1.2 in about 2000 iterations. To be conservative, we collect 5000 observations after 5000 "burn-in" in computing the absolute bias (BIAS) and the root mean squares (RMS) of the estimates and the true values in 100 replications. Results obtained from the Gibbs sampler are summarized in Table 2, where SD denotes the standard deviation of the estimates. Based on these results, it can be seen that our proposal is rather accurate and effective. It also shows that there are no significant differences of estimates between prior (I) and prior (II). Therefore, the proposal method is more robust against the choices of values of hyperparameters. An interesting phenomenon is that the estimates of variance parameters in unique errors under prior (I) are more accurate than those under prior (II). This reflects that it provides more information in estimation.

For model comparison, we consider the following competing models to illustrate the performance of the Bayes factor.

M_1 is the aforementioned model given by (4). Consider

$$\begin{aligned} M_0\colon\ & p\left(y_{ij} \mid \vartheta_{ij}\right) = C_{n_{ij}}^{y_{ij}} \exp\left\{y_{ij}\vartheta_{ij} - n_{ij} b\left(\vartheta_{ij}\right)\right\}, \\ & \vartheta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \lambda_j \xi_i, \xi_i, \quad \text{iid}\, N\left(0, \sigma^2\right). \end{aligned} \tag{19}$$

Note that M_0 is nested to M_1 in the sense that all the parameters in model M_0 are included in model M_1. M_0 indicates that no temporal correlation exists among the random effects. The linked model between M_1 and M_0 is taken as

$$\begin{aligned} M_t\colon\ & p\left(y_{ij} \mid \vartheta_{ij}\right) = C_{n_{ij}}^{y_{ij}} \exp\left\{y_{ij}\vartheta_{ij} - n_{ij} b\left(\vartheta_{ij}\right)\right\}, \\ & \vartheta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \lambda_j \xi_i, \quad \xi_i = t\phi\xi_{i-1} + \delta_i, \end{aligned} \tag{20}$$

in which $t = 1$ and $t = 0$ correspond to M_1 and M_0, respectively. For computation, we choose $S = 20$ in $[0, 1]$ and draw 5000 observations after 5000 burn-ins deleted from the posterior distribution for each t_s to calculate the $\log \mathrm{BF}_{10}$. Figure 3 gives the histogram of the values of $\log \mathrm{BF}_{10}$ across 100 replications. The estimate of $\log \mathrm{BF}_{10}$ is 11.2222 with standard deviation 6.6242. Based on the guidelines given by Kass and Raftery [18], M_1 is selected with positive evidence.

To investigate the smoothing and forecasting of our proposal, we regenerate a data set with sample size 100 and treat the last 5×4 observations as unobserved. The Gibbs sampler with MH algorithm is used to compute the first ninety-five smoothing values and the last five forecasting values of the latent variables ξs. Figure 4 gives the true values and smoothing/forecasting values based on 5000 simulated observations after deleting first 5000 observations. It can be seen that it fits well between the estimated values and true values.

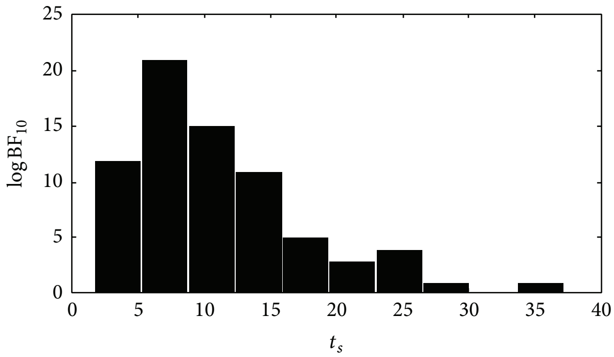

Figure 3: Histogram of $\log \mathrm{BF}_{10}$ of posterior distribution for each t_s.

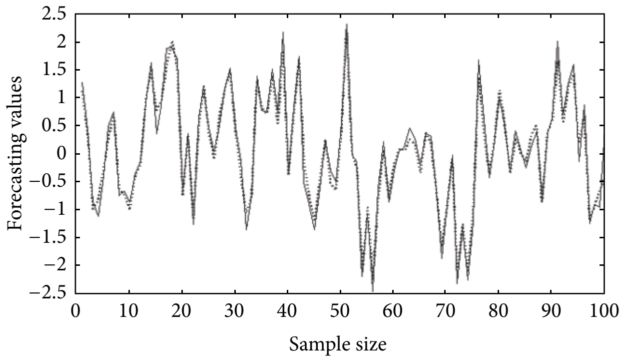

Figure 4: Comparison with true values (dashed line) and forecasting values (pointed line).

5.2. Pika's Data. An application is considered in relation to a study of Guo et al., 2009 [32], to illustrate the developed methodology. The data set is collected from Qinshui Forest farm of Jincheng Coal Industry Group sited on the northwest of Qinshui County of Jincheng City in Shanxi Province, 112°19′40″–112°19′52 east longitude and 35°48′43″–35°48′57 north latitude. Details about experiment design can be found in Guo et al., 2009 [32]. The original data set is constituted of the number of the living tree and pika within four plots in five months. The primary concern is to assess the relationship between the survival rate of young plantation and the pika population and model the optimal pika population per hectare. The data set is reanalyzed by Xia et al. [33] to investigate the correlations among the observed responses. In these studies the latent variables model is established through AR(1) model (3) coupled with model (1) and (2), which interprets the correlations among the outcomes resulting from the time dependence.

Firstly, the following hypothesis is considered: H_0: $\phi = 0$. If H_0 is true, the model becomes the common generalized logistic model considered by Xia et al. [33]. Path sampling is used to compute $\log \mathrm{BF}_{10}$. The linked model is defined similarly as (20). $S = 20$ grids were chosen in $[0, 1]$, and, for each t_s $(s = 1, \ldots, 20)$, 10000 random observations were drawn from the each posterior distributions and the first 5000 observations were deleted in view of the burn-in phase. The logarithm of Bayes factor is 3.0143. It shows that there

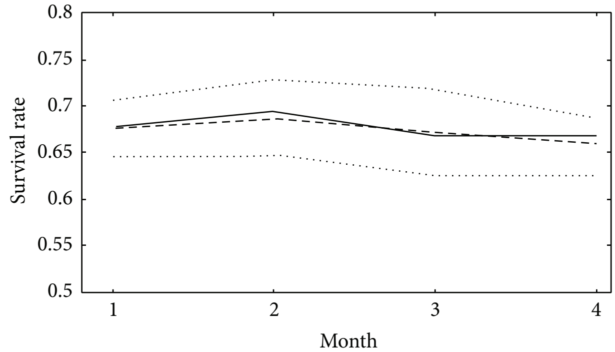

Figure 5: Comparison with the predictions for survival rate (solid line) and the actual values (dashed line).

exists significant evidence against H_0; hence, our proposal is appropriate.

To assess the effect of forecasts of the proposal methodology, the original data set is divided into two parts: one is formed by the previous sixteen observations in four months within four plots and the other is made up of four observations in the following fifth month. The proposal algorithm computes the predictive values of the proportion of the living tree in the fifth month based on the previous 20 observations. Figure 5 shows the forecasts and actual values of the proportions of the living tree within four plots. It can be seen that fitted values match the actual values closely.

6. Concluding Remarks

Logistic regression model is the most popular model used to interpret the relationship between the responses and covariates and to assess the effect of covariates on the responsible probability. Bayesian analysis of logistic model has received a lot of attention recently; see Johnson and Albert [34] and Congdon [35]. For example, Choi et al. [36] discussed a Bayesian statistical inference about missing information on the basis of the logistic regression model. They also used Gibbs sampler to get the estimates and the posterior analysis of the posited model. Since their model is not an AR(1) model, the underlying development is less complicated. Recently, McCormic et al. [37] proposed an online binary classification procedure based on the dynamic logistic regression and dynamic model averaging. However, their developments are restricted within a single response y_t, which do not need to explore the relationships among the observed variables. Though Xia et al. [33] establish the Bayesian analysis for generalized logistic model, the essential difference lies in that their latent variables are identically and independently distributed according to normal distribution and, hence, can not capture the temporal correlation.

In the present paper, dynamic factor model is established to characterize the temporal correlation among responses. One contribution in this paper is the development of a feasible estimation procedure for obtaining the Bayesian estimates of the parameters and latent factor scores. The hybrid Gibbs sampler with MH algorithm was implemented to provide a convenient mechanism for implementing our method. Another contribution is the development of test statistics on the Bayes factor for testing hypothesis of the model. Computation of the Bayes factor in the current complex model is on the basis of path sampling. Further, Bayesian forecasting procedure which takes advantages of the full conditional distribution is presented. Results from empirical studies indicate that these procedures can be usefully applied to real studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The paper is supported by the Natural Science Foundation (no. 61306046) and the Anhui Provincial Educational Natural Science Foundation (no. KJ2013A177).

References

[1] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2000.

[2] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2002.

[3] J. Nelder and R. W. M. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society A*, vol. 135, pp. 370–384, 1972.

[4] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, UK, 2nd edition, 1989.

[5] B. C. Wei, *Exponential Family Nonlinear Models*, Springer, Singapore, 1998.

[6] K. A. Bollen, *Structural Equations with Latent Variables*, John Wiley & Sons, New York, NY, USA, 1989.

[7] S. Y. Lee, *Structural Equation Modeling: A Bayesian Perspective*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2007.

[8] A. Skrondal and S. Rabe-Hesketh, *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Taylor & Francis, London, UK, 2004.

[9] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721–741, 1984.

[10] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[11] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[12] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 1985.

[13] M. Aitkin, "Simpson's paradox and the Bayes factor," *Journal of the Royal Statistical Society B*, vol. 60, no. 1, pp. 269–270, 1998.

[14] A. E. Raftery, "Bayesian model selection in structural equation models," in *Testing Structural Equation Models*, K. A. Bollen and J. S. Long, Eds., pp. 163–180, Sage, Newbury Park, Calif, USA, 1993.

[15] A. E. Raftery, "Bayesian model selection in social research," in *Sociological Methodology*, A. E. Raftery, Ed., Blackwell, Oxford, UK, 1995.

[16] S. Chib, "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1313–1321, 1995.

[17] A. O'Hagan, "Fractional Bayes factors for model comparison," *Journal of the Royal Statistical Society B Methodological*, vol. 57, no. 1, pp. 99–138, 1995.

[18] R. E. Kass and A. E. Raftery, "Bayes factors," *Journals of the American Statistical Association*, vol. 90, pp. 773–795, 1995.

[19] T. J. DiCiccio, R. E. Kass, A. Raftery, and L. Wasserman, "Computing Bayes factors by combining simulation and asymptotic approximations," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 903–915, 1997.

[20] A. Gelman and X. Meng, "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling," *Statistical Science*, vol. 13, no. 2, pp. 163–185, 1998.

[21] A. E. Gelfand and D. K. Dey, "Bayesian model choice: asymptotics and exact calculations," *Journal of the Royal Statistical Society B. Methodological*, vol. 56, no. 3, pp. 501–514, 1994.

[22] X. Meng and W. H. Wong, "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration," *Statistica Sinica*, vol. 6, no. 4, pp. 831–860, 1996.

[23] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME D: Journal of Basic Engineering*, vol. 83, pp. 35–45, 1960.

[24] G. Kitagawa, "Non-Gaussian state-space modeling of nonstationary time series," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1032–1063, 1987.

[25] R. J. Meinhold and N. D. Singpurwalla, "Robustification of Kalman filter models," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 479–486, 1989.

[26] B. P. Carlin, N. G. Polson, and D. S. Stoffer, "A Monte Carlo approach to nonnormal and nonlinear state-space modeling," *Journal of the American Statistical Association*, vol. 87, pp. 493–500, 1992.

[27] A. F. Smith and M. West, "Monitoring renal transplants: an application of the multiprocess Kalman filter," *Biometrics*, vol. 39, no. 4, pp. 867–878, 1983.

[28] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–550, 1987.

[29] L. Tierney, "Markov chains for exploring posterior distributions," *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1728, 1994.

[30] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–472, 1992.

[31] M. K. Cowles, "Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models," *Statistics and Computing*, vol. 6, no. 2, pp. 101–111, 1996.

[32] A. H. Guo, Y. Chen, and L. Y. Fu, "Harm about the artifical growth forest with Pikas," *Journal of Shandong Forestry Science and Technology*, vol. 180, no. 1, pp. 54–55, 2009.

[33] Y. M. Xia, Y. A. Liu, and Z. Fang, "Modeling survival rate of forest through Bayesian generalized logistic regression model," *Journal of the Nanjing Forestry*, vol. 34, pp. 47–50, 2010.

[34] V. E. Johnson and J. H. Albert, *Ordinal Data Modeling*, John Wiley & Sons, New York, NY, USA, 1999.

[35] P. Congdon, *Bayesian Models for Categorical Data*, John Wiley & Sons, New York, NY, USA, 2005.

[36] T. Choi, M. J. Schervish, and K. A. a. Schmitt, "A Bayesian approach to a logistic regression model with incomplete information," *Biometrics. Journal of the International Biometric Society*, vol. 64, no. 2, pp. 424–430, 2008.

[37] T. H. McCormick, A. E. Raftery, D. Madigan, and R. S. Burd, "Dynamic logistic regression and dynamic model averaging for binary classification," *Biometrics*, vol. 68, no. 1, pp. 23–30, 2012.