Bayesian first order autoregressive latent variable models for multiple binary sequences

Federica Giardina, Alessandra Guglielmi[1],

Fernando A. Quintana[2] and Fabrizio Ruggeri[3]

[1]Dipartimento di Matematica, Politecnico di Milano, Milano, ITALY,
e-mail: alessandra.guglielmi@polimi.it

[2]Departamento de Estadística, Pontificia Universidad Católica de Chile,
Santiago, CHILE, e-mail: quintana@mat.puc.cl

[3] Istituto di Matematica Applicata e Tecnologie Informatiche,
Consiglio Nazionale delle Ricerche, Milano, ITALY,
e-mail: fabrizio@mi.imati.cnr.it.

Abstract

Longitudinal clinical trials often collect long sequences of binary data monitoring a disease process over time. Our application is a medical study conducted by VACURG to assess the effectiveness of a chemioterapic treatment (thiotepa) in preventing recurrence on subjects affected by bladder cancer. We propose a generalized linear model with latent autoregressive structure for longitudinal binary data following a Bayesian approach. We describe a suitable posterior simulation scheme and discuss inference and sensitivity issues for the bladder cancer data.

Keywords: binary longitudinal data; first order autoregressive model; hierarchical Bayesian modelling; latent variables.

AMS 2000 Mathematics Subject Classification: 62F15, 62J12, 62P10.

1 Introduction

Many prediction studies in medical research lead to discrete longitudinal data with continuous, ordinal or categorical outcomes. For instance, longitudinal data arise in clinical trials when a time sequence of measurements is taken from each of a number of experimental units allocated to one of several treatments. The main peculiarity of longitudinal data is the dependency of multiple responses from the same individual.

We consider here a generalized linear model for binary longitudinal variables. Specifically, we develop a Bayesian first order autoregressive model with the introduction of latent variables for a set of binary repeated measurements. The dataset is part of a bladder cancer study conducted in the USA by the Veterans Administration Cooperative Urological Research Group (VACURG) about comparing the effectiveness of three treatments (placebo, pyridoxine, and topical thiotepa) in preventing recurrence of Stage I bladder cancer (Byar et al., 1977). Davis and Wei (1988) analyzed data coming from placebo and thiotepa treatment groups. They considered a class of univariate one-sided global asymptotically distribution-free tests for the equality of the two treatments. Their testing and estimation procedures take advantage of the fact that repeated measurements of the same characteristic scheduled to be taken over a common set of time points for each study subject are non-decreasing. Davis (1996) extended the non-parametric methods to the case of comparison of multiple treatment groups with a control group, using data from placebo, pyridoxine and topical thiotepa treatment groups. Quintana and Müller (2004) considered the problem of optimal sampling design approaching it as a Bayesian decision problem. They transformed the original non-decreasing data from placebo and thiotepa treatment group of VACURG study into binary repeated measurements, assuming a nonparametric Bayesian model, together with partial exchangeability, as meant in Quintana and Newton (1998).

We revisit the binary data discussed in Quintana and Müller (2004), assuming a first order autoregressive structure for the latent variables, as suggested by them. We construct a generalized linear model for binary longitudinal variables, with the introduction of normal latent variables or, which is the same, by means of the probit link function. The dependence between consecutive binary variables is obtained through an

autoregressive model for the latent variables. A consequence of assuming serial dependence at the level of latent variables is that the induced marginal correlation structure for observations is richer than first order homogeneous Markovian. We specifically introduce terms that allow for trends in the conditional probability and for treatment effect. As will be explained later, our analysis does reveal a significant treatment effect. Furthermore, we conduct sensitivity analysis on all the parameters of the model showing a certain robustness across different specification of prior distributions. A Gibbs sampler algorithm was implemented in Matlab to compute all the Bayesian estimates of the parameters as well as the predictive probabilities of recurrence for individuals in the study and new patients. As a double-check, we compared the results with those obtained using WinBUGS (Lunn et al., 2000).

The article is organized as follows. In Section 2 we illustrate the medical problem while Section 3 describes the main features of the proposed model, emphasizing the role of the latent variables and calculating predictive distributions. In Section 4 we present posterior simulation results and discuss inference on the quantities of interest. We also present the results of a sensitivity analysis conducted to assess changes in the posterior distribution under different prior scenarios. We conclude with a discussion and possible extensions of our analysis in Section 5. The full conditionals used in the Gibbs sampler are presented in the Appendix.

2 Data

Because of its high recurrence rate and the need for lifelong surveillance, bladder cancer is the most expensive cancer to treat on a per-patient basis. If a bladder cancer only affects the inner lining of the bladder, it is known as a superficial cancer or Stage I bladder cancer. Stage I bladder tumors can usually be completely removed by transurethral resection, but many patients have multiple recurrences. The subsequent tumors sometimes show a higher degree of malignancy and may even progress to invasive carcinoma.

At the beginning of the randomized clinical trial conducted by the VACURG all

patients had superficial bladder tumors. In order to determine if recurrences of Stage I bladder cancer can be prevented, these tumors are removed transurethrally and patients are assigned to one of three treatments: placebo, thiotepa, or pyridoxine (Vitamin B-6). At subsequent follow-up visits, all recurrent tumors are removed and treatment is continued. Although patients are scheduled to be reexamined every three months for tumor recurrences, there are many missing observations. Observations from thiotepa treatment patients are generally obtained more frequently than from those in placebo group and pyridoxine group, since the thiotepa is scheduled to be administered on a regular basis. Thiotepa is a chemotherapy drug used to reduce the size of a cancerous tumor and prevent the growth of new cancer cells.

The study conducted by VACURG consists of $m = 82$ patients with up to a maximum of $n_i = 12$ observations taken every three months for each patient. We consider only patients grouped into treatment (thiotepa) and placebo: group T (36 subjects) and group P (46 subjects). The original data set (see Davis and Wei, 1988) consists of non-decreasing repeated measures of the same characteristics, specifically the cumulative counts of recurrent tumors at each visit j for patients from groups T and P.

Since the aim here is evaluating the treatment effect, we decide to look at differences between every response variable of subject i at time j and the value taken at the previous visit. Each response variable is indicated with Y_{ij} where $i = 1, \ldots, m$ denotes individuals and $j = 1, \ldots, n_i$ denotes the measurements (or the measurement time for each individual i). Of course, each observation records an indicator of recurrence of bladder cancer tumors, *i.e.* $y_{ij} = 1$ if an increased number of tumors was detected at time j for patient i, and $y_{ij} = 0$ otherwise. Each missing value was substituted by a linear interpolation and rounding before the reduction to binary data.

3 The model

For any $i = 1, \ldots, m$, and $j = 1, \ldots, n_i$, Y_{ij} is a Bernoulli r.v. with mean p_{ij}, which represents the probability of recurrence in bladder cancer tumors. Patients are grouped

into P (placebo) group, $x_i = 0$, and T (treatment) group, $x_i = 1$, for $i = 1, \ldots, m$. This variable will represent the only covariate considered in our regression analysis. Of course, the proposed model can be straightforwardly adapted to datasets with more covariates than the VACURG data. Latent variable models have gradually become a standard tool and an active area of study in a wide range of problems in medical research. Such models provide a useful and intuitive way to motivate the distribution of a discrete outcome assuming that the binary event occurs only if an unobservable continuous variable (latent variable) exceeds a certain level.

Here we introduce N latent variables Z_{ij}, where $N = \sum_{i=1}^{m} n_i$ is the total number of observations. The binary r.v.'s Y_{ij} are modelled as

$$Y_{ij} = \begin{cases} 1 & \text{if } Z_{ij} \geq c \\ 0 & \text{if } Z_{ij} < c, \end{cases} \tag{3.1}$$

c being a fixed threshold that we assume equal to zero for reference. Moreover,

$$\mathbf{Y}_1, \ldots, \mathbf{Y}_m | \mathbf{Z}_1, \ldots, \mathbf{Z}_m \overset{ind}{\sim} \text{ as in (3.1)}, \tag{3.2}$$

where $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{in_i})'$. The latent vectors $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ are independent but a Markovian dependence within each $\mathbf{Z}_i$ is introduced via

$$Z_{ij} | Z_{ij-1} \sim \mathcal{N}\left(\beta_0 + \beta_1 x_i + \alpha z_{ij-1}, \sigma^2\right), \quad i = 1, \ldots, m, \; j = 2, \ldots, n_i, \tag{3.3}$$

where β_0 is the intercept of the regression model and β_1 represents the treatment effect on the response variable. The parameter α is the autoregressive coefficient of the first order, the parameter standing for the dependence on the previous response (through latent variables). The model is completed by assuming

$$Z_{i1} \sim \mathcal{N}\left(\mu_0, \sigma^2\right), \quad i = 1, \ldots, m, \tag{3.4}$$

where μ_0 is considered constant and σ^2 unknown.

Model (3.1)-(3.4) represents a generalized linear model for longitudinal data. The normal regression model on the latent variables leads to a probit model for the binary variables. The underlying autoregressive structure makes the model belonging to

the conditional models class for longitudinal data as described in Diggle et al. (2002). Specifically, the latent variables are represented as a transition model since each variable Z_{ij} of subject i at visit j only depends on the past value Z_{ij-1}. In the above construction, β_0 represents a population baseline probability of tumor recurrence, and β_1 is a treatment-specific offset, with negative values representing a decreased probability of recurrence. Finally, α is the autoregressive effect of lagged responses.

It is straightforward to show that, marginally (of course conditioning on the parameters), for every individual i,

$$\begin{aligned}
\mathbb{E}[Z_{ij}] &= \frac{1-\alpha^{j-1}}{1-\alpha}\left(\beta_0+\beta_1 x_i\right)+\alpha^{j-1}\mu_0, \quad j=1,\ldots,n_i,\\
\mathrm{Var}[Z_{ij}] &= \frac{1-\alpha^{2j}}{1-\alpha^2}\sigma^2, \quad j=1,\ldots,n_i,\\
\mathrm{Cov}[Z_{ij},Z_{ik}] &= \alpha^{|j-k|}\,\mathrm{Var}[Z_{i\min(j,k)}], \quad j,k=1,\ldots,n_i,\ j\neq k.
\end{aligned}$$

Moreover, for each individual i, the probability of tumor recurrence, conditional on the parameters, is

$$\begin{aligned}
\mathbb{P}\left(Y_{i1}=1\right) &= \int_{\mathbb{R}} \mathbb{P}\left(Y_{i1}=1|z_{i1}\right) f\left(z_{i1}\right)\, dz_{i1} = \\
&= \int_{\mathbb{R}} \mathbb{I}_{[c,+\infty)}(z_{i1})\,\sqrt{\frac{\tau}{2\pi}}\exp\left\{\frac{-\tau\left(z_{i1}-\mu_0\right)^2}{2}\right\} dz_{i1} = \Phi[\sqrt{\tau}\,(\mu_0-c)],
\end{aligned} \tag{3.5}$$

where $\tau=1/\sigma^2$ and Φ denotes the standard gaussian d.f., and

$$\begin{aligned}
\mathbb{P}\left(Y_{i2}=1\right) &= \int_{\mathbb{R}^2} \mathbb{P}\left(Y_{i2}=1|z_{i2},z_{i1}\right) f\left(z_{i2}|z_{i1}\right) f\left(z_{i1}\right)\, dz_{i2}\, dz_{i1}\\
&= \frac{\tau}{2\pi}\int_{\mathbb{R}} dz_{i1}\int_c^{+\infty} dz_{i2}\exp\left\{\frac{-\tau[z_{i2}-(\beta_0+\beta_1 x_i+\alpha z_{i1})]^2}{2}\right\}\times\\
&\qquad\qquad \times\exp\left\{\frac{-\tau\left(z_{i1}-\mu_0\right)^2}{2}\right\},
\end{aligned}$$

and so on. From the above integrals it is clear that the likelihood has not a simple expression, if compared to the conditional likelihood

$$\mathcal{L}(\mathbf{y}|\mathbf{z},\mathbf{x}) = \prod_{\{y_{ij}=1\}} \mathbb{I}_{[0,+\infty)}(z_{ij}) \prod_{\{y_{ij}=0\}} \mathbb{I}_{(-\infty,0)}(z_{ij}).$$

The parameter of interest is the vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \tau, \mathbf{Z})'$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \alpha)'$, $\tau = 1/\sigma^2$ and $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ are the latent variables. Note that we treat the latent variables $\{Z_{ij}\}$ as random parameters, since they are unobserved variables.

We assign the following prior distributions:

$$\boldsymbol{\beta} \sim MN\left(\boldsymbol{\beta}_m, \mathbf{V}\right), \qquad \tau \sim gamma\left(\frac{n_0 S_0}{2}, \frac{n_0}{2}\right), \qquad \boldsymbol{\beta} \perp \tau \tag{3.6}$$

i.e. $\boldsymbol{\beta}$ follows a multivariate normal distribution with mean $\boldsymbol{\beta}_m$ and variance-covariance matrix $\mathbf{V}$, τ has gamma prior distribution with mean S_0 and variance $2S_0/n_0$. The conditional distribution of the latent variables Z_{ij} for $i = 1, \ldots, m, j = 2, \ldots, n_i$ given their previous Z_{ij-1} can be written as

$$f\left(z_{ij}|z_{ij-1}\right) = \sqrt{\frac{\tau}{2\pi}} \exp\left\{\frac{-\tau[z_{ij} - (\boldsymbol{\beta}'\mathbf{w}_{ij})]^2}{2}\right\}$$

where $\mathbf{w}_{ij} = (1, x_i, z_{ij-1})'$. The joint prior distribution will be expressed by the product

$$\pi\left(\boldsymbol{\theta}\right) = \pi\left(\boldsymbol{\beta}\right)\pi\left(\tau\right)\prod_{i=1}^{m} f\left(z_{i1}\right)\prod_{j=2}^{n_i} f\left(z_{ij}|z_{ij-1}\right),$$

where $\pi\left(\boldsymbol{\beta}\right)$, $\pi\left(\tau\right)$ denote the prior pdfs of the parameters $\boldsymbol{\beta}$ and τ in (3.6).

Due to mathematical convenience, we decide to choose conditionally-conjugate priors, i.e. all full conditional distributions have the same functional form of the priors.

Based on the priors and the likelihood given, the posterior distribution of $\boldsymbol{\theta}$ is given by

$$\begin{aligned}
&\pi\left(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}\right) \propto \pi\left(\boldsymbol{\theta}\right)\mathcal{L}(\mathbf{y}|\mathbf{z}, \mathbf{x}) \\
&= \pi\left(\boldsymbol{\beta}\right)\pi\left(\tau\right)\prod_{i=1}^{m} f\left(z_{i1}\right)\prod_{j=2}^{n_i} f\left(z_{ij}|z_{ij-1}\right) \times \prod_{\{y_{ij}=1\}} \mathbb{I}_{[0,+\infty)}(z_{ij}) \prod_{\{y_{ij}=0\}} \mathbb{I}_{(-\infty,0)}(z_{ij}).
\end{aligned} \tag{3.7}$$

As far as the posterior predictive distributions are concerned, first of all, we will compute the probability of recurrence of the tumor at the $n_i + 1$-st time for every subject i, which is present in the study with n_i outcomes:

$$\mathbb{P}(Y_{in_i+1} = 1|\mathbf{Y}, \mathbf{x}) = \int_{\Theta\times\mathbb{R}} \mathbb{I}_{[0,+\infty)}(z_{in_i+1}) f(z_{in_i+1}|\mathbf{Y}, \mathbf{x}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{Y}, \mathbf{x})\, d\boldsymbol{\theta} dz_{in_i+1}.$$

Secondly, we aim at estimating the probability that a new subject k (in the T or P group) will have a recurrence in the number of tumors in all the 12 quarterly measurement times. For any $j = 2, \ldots, 12$, if $\mathbf{Y}_k^{(j-1)} = (Y_{k1}, Y_{k2}, \ldots, Y_{kj-1})$, then

$$\mathbb{P}\left(Y_{kj} = 1 | \mathbf{Y}, \mathbf{Y}_k^{(j-1)}, \mathbf{x}, x_k\right)$$
$$= \int_{\Theta^{(j-1)} \times \mathbb{R}} \mathbb{I}_{[0,+\infty)}(z_{kj}) f(z_{kj} | \mathbf{Y}, \mathbf{Y}_k^{(j-1)}, \boldsymbol{\theta}^{(j)}, \mathbf{x}, x_k) \pi\left(\boldsymbol{\theta}^{(j-1)} | \mathbf{Y}, \mathbf{Y}_k^{(j)}, \mathbf{x}, x_k\right) \, d\boldsymbol{\theta}^{(j-1)} dz_{kj},$$

where $\boldsymbol{\theta}^{(j-1)} = (\boldsymbol{\beta}, \tau, \mathbf{Z}, Z_{k1}, \ldots, Z_{kj-1})'$ and $\Theta^{(j-1)}$ denotes its space. An obvious modification of the last expression is due when $j = 1$.

4 Analysis

In this section we explore the differences among the Bayesian estimates of the parameters across different prior specifications, investigating also the effect of prior dependence among the $\boldsymbol{\beta}$-components. All estimates were computed via a Gibbs sampler algorithm, whose full conditional distributions are given in the Appendix. Moreover, all results obtained in Matlab were checked using WinBUGS (Lunn at al. 2000; see also `http://www.mrc-bsu.cam.ac.uk/bugs/`).

4.1 Results

For each simulation in Matlab we run a chain of 751 000 iterations with a burn-in period of 1 000 and thinning period of 30 iterations. Therefore, our estimates are based on an actual number of 25 000 iterations. Of course, we evaluated several diagnostics implemented in the CODA package (Cowles and Carlin, 1996) in R and found no evidence against convergence. Furthermore, for every chain we run, we also calculated the batch standard errors obtaining very small values. This is a useful tool to determine the accuracy on estimations, as suggested by Gilks et al. (1996).

First of all, we assumed $\mu_0 = 0$ and the prior (3.6) with $\boldsymbol{\beta}_m = \mathbf{0}$, $\mathbf{V} = \mathbf{I}$, $n_0 = 4$, $S_0 = 3/2$, *i.e.*

$$\boldsymbol{\beta} \sim MN(\mathbf{0}, \mathbf{I}), \qquad \tau \sim gamma(3, 2), \qquad \boldsymbol{\beta} \perp \tau \tag{4.1}$$

where $\mathbf{I}$ is the identity matrix of size 3. Posterior means and standard deviations of the parameters α, β_0, β_1 and τ are summarized in Table 5 (first row). The posterior mean of α, is 0.6269 confirming that the assumption of an autoregressive structure is reasonable. The posterior distributions of β_0 and β_1 are both mostly concentrated on the negative numbers. This means that there is a baseline probability of recurrence of less than 50% and that the treatment (thiotepa) contributes to reduce the probability of a recurrence in the number of tumors of the bladder cancer study.

Figure 1 shows the posterior kernel density plots of the parameters, obtained via the rule of thumb suggested by Silverman (1986) as the default option of R. From such plot for β_1 it is clear that this distribution lies mainly on the negative reals, indicating evidence of the treatment effect. Turning now our attention to predictions, we considered four patients in the P-group and three patients in the T-group. For the reader's sake, data concerning these patients are included here (Table 1).

Patient	Month											
ID	3	6	9	12	15	18	21	24	27	30	33	36
10	1	0	0	0	1	0	1	1	·	·	·	·
17	0	0	0	0	0	0	0	0	0	·	·	·
24	0	0	0	0	0	0	0	1	1	1	·	·
34	1	0	0	0	0	0	0	0	0	0	0	0
60	0	1	0	1	0	1	0	0	0	·	·	·
71	0	0	0	0	0	0	0	0	·	·	·	·
74	1	0	0	0	0	0	0	0	0	0	0	·

Table 1: Data for some Placebo and Treatment patients

For every patient we evaluated the probability of recurrence at a future time, obtaining the estimates in Table 2.

As expected, Patient 10 from group P has a quite high probability of recurrence at the next (9-th) check, since he/she presented four recurrences during the clinical trial. Subject 17 at the 10-th check, as well as Patient 34 at the 13-th check, shows a small probability of recurrence, because of their medical history. Subject 24 reveals a probability of recurrence of around 0.45, because of the responses of the last visits.

Predictive probabilities of recurrence						
$Y_{10,9}$	$Y_{17,10}$	$Y_{24,11}$	$Y_{34,13}$	$Y_{60,10}$	$Y_{71,9}$	$Y_{74,12}$
0.4439	0.0617	0.4562	0.0618	0.0666	0.0514	0.0502

Table 2: Estimates of the predictive probabilities in a new measurement for subject 10,17,24,34 (placebo) and 60,71,74 (treatment).

Treated individuals (T-group) seem to present a smaller probability of recurrence at future checks.

On the other hand, for two new subjects assigned to the T and P groups, respectively, the estimated probabilities of recurrence are in Table 3. It can be observed that, at least during the period under consideration, the recurrence probability is decreasing in both subjects but it is definitely lower for the subject in the T-group. It seems there is a *natural* decrease in the number of recurrences but this is more marked when subjects are treated with thiotepa.

Predictive probabilities of recurrence											
1	2	3	4	5	6	7	8	9	10	11	12
0.3439	0.2875	0.1957	0.1507	0.1200	0.1081	0.1020	0.0914	0.0897	0.0863	0.0850	0.0844
0.4759	0.3856	0.3542	0.3212	0.2765	0.2501	0.2103	0.1987	0.1901	0.1898	0.1822	0.1801

Table 3: Estimates of the predictive probabilities for two new subjects in the T (first row) and P (second row) groups at every measurement times under prior (4.1).

Finally, we compute the posterior cross correlation for $(\beta_0, \beta_1, \alpha, \tau)$, obtaining values in Table 4. The correlation between regression parameters are all moderate and similar. Both parameters β_0 and β_1 are highly correlated with the precision parameter τ, but the autoregression parameter α is seen to have almost null posterior correlation with τ.

4.2 Sensitivity analysis for β_0, β_1 and α

In order to assess the impact of choosing different priors, we conducted a sensitivity analysis of the inference, varying the parameters of mean and precision of the prior

CROSS CORRELATION				
	β_0	β_1	α	τ
β_0	1.0000	0.2261	0.2842	0.7365
β_1	0.2261	1.0000	0.1769	0.4212
α	0.2842	0.1769	1.000	-0.0060
τ	0.7365	0.4212	-0.0060	1.0000

Table 4: Cross correlation of the parameters.

distributions of α, β_0 and β_1 without modifying the distribution of τ. We assume first

$$\boldsymbol{\beta} \sim MN(\mathbf{0}, 4\mathbf{I}), \qquad \tau \sim gamma(3,2), \qquad \boldsymbol{\beta} \perp \tau \tag{4.2}$$

and

$$\boldsymbol{\beta} \sim MN(\mathbf{0}, 25\mathbf{I}), \qquad \tau \sim gamma(3,2), \qquad \boldsymbol{\beta} \perp \tau. \tag{4.3}$$

Moreover, the values in Table 4 justify the choice of the prior (3.6) with

$$\boldsymbol{\beta} \sim MN(\mathbf{0}, \mathbf{V}), \quad \mathbf{V} = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}, \quad \tau \sim gamma(3,2), \quad \boldsymbol{\beta} \perp \tau. \tag{4.4}$$

Here the marginal prior distributions for β_0, β_1 and α are all standard normal, but the parameters are no longer independent. The prior distribution of τ remains the same, τ being independent on $(\beta_0, \beta_1, \alpha)$.

We see that Bayesian estimates of the parameters present very close values in all cases. All the Bayesian estimates obtained under (4.2)-(4.4) are summarized in Table 5. Furthermore, if we compare those Bayesian estimates with the ones resulting from (4.1) (independent standard normal priors for the $\boldsymbol{\beta}$-components), no significative differences are detected. Kernel density plots represented in Figures 2, 3 and 4 show the same behavior as the one represented in Figure 1. We can conclude that the Bayesian estimates of the parameters are robust across reasonably different specifications of the parameters of the $\boldsymbol{\beta}$ prior.

Prior	β_0	β_1	α	τ
$\mathbf{V}=\mathbf{I}$	-0.3961	-0.2054	0.6269	1.4851
	(0.1302)	(0.1147)	(0.0667)	(0.7015)
$\mathbf{V}=4\mathbf{I}$	-0.3879	-0.1992	0.6259	1.5293
	(0.1203)	(0.1077)	(0.0667)	(0.7079)
$\mathbf{V}=25\mathbf{I}$	-0.3898	-0.2011	0.6270	1.5126
	(0.1224)	(0.1103)	(0.0676)	(0.7098)
$\mathbf{V}$ as in (4.4)	-0.3856	-0.1985	0.6242	1.5456
	(0.1188)	(0.1076)	(0.0670)	(0.7149)

Table 5: Posterior means and standard deviations of the parameters for priors as in (3.6).

4.3 Sensitivity analysis for τ

So far we have assigned a gamma with mean $E[\tau] = 3/2$ and variance $\mathrm{Var}[\tau] = 3/4$ as prior distribution for τ. Now we want to investigate the effect of an increased/decreased precision τ of the latent variables Z_{ij} on the parameters estimates, retaining the independence between τ and $(\beta_0, \beta_1, \alpha)$. First, the prior distributions considered for α, β_0 and β_1 are independent standard normal distributions (as in (4.1)). We fix different values for $\mathbb{E}[\tau]$ and $\mathrm{Var}[\tau]$; the estimates obtained are in Table 6.

It is worth mentioning that, in case (ii), the mean of $1/\tau$ (which expresses the variance of the latent variables) is infinite. Comparing (i), (ii), (iii) (as a function of the precision), we see that the estimate of α is really robust, while the estimates of β_0 and β_1 reveal some sensitivity, although in all cases the respective posterior distributions remain mostly concentrated on the negative numbers. The case of precision τ is different though. We find the posterior precision to be highly sensitive to prior assumptions. Similar findings relating to scale parameters in models involving related terms have been reported elsewhere (e.g. Gelman, 2006). We return to this particular point later.

We computed the predictive probabilities of recurrence for the same patients in the study considered before, that is, Patient 10,17,24,34 from the P-group, and 60,71,74

(treatment) from the T-group, when $\boldsymbol{\beta}$ and τ are a priori independent, with $E[\tau] = 1$ and $\mathrm{Var}[\tau] = 1$; see Table 8 (first row). There are no significative differences with values in Table 2, corresponding to independent standard gaussian distribution for the $\boldsymbol{\beta}$-components.

Moreover, under the same prior, in Table 9 we display the predictive probabilities of recurrence for two new subjects, one in the P-group, the other in the T-group. We started from an ignorance situation in which we had no opinion about tumor recurrence within the first 3-month period. We modelled ignorance by taking $\mu_0 = c = 0$ in (3.5) so that we started with a recurrence probability at the first visit equal to 0.5. Of course, we expect that information will be available in other experiments conducted with physicians and different choices of μ_0 and c will be possible. From Table 9 we can see that our choice was rather conservative: after an estimated recurrence probability at the first visit near 0.5 for both groups, we can see it drops dramatically to, at least, 0.29 for the T-group and 0.36 for the P-group.

Since the hypothesis of independence between $\boldsymbol{\beta}$ and τ could be restrictive, as an alternative to (3.6), we assumed the following prior

$$\boldsymbol{\beta}|\tau \sim MN(\mathbf{0}, \frac{1}{\tau}\mathbf{V}) \text{ and } \tau \sim gamma\,(n_0 S_0/2, n_0/2)\,. \tag{4.5}$$

As before, we choose different values of $\mathbb{E}[\tau]$ and $\mathrm{Var}[\tau]$ and computed the Bayesian estimates; see Table 6 (iv), (v), (vi). In particular, we computed, under the same prior as in Table 6 (vi), the probabilities of recurrence for new patients, as well as for some patients already included in the study. See Table 8 and 9. In all these computational experiments we observe some common features. The predicted probabilities and regression coefficients do not change drastically across different scenarios. But the precision τ is considerably affected by prior assumptions, as before.

Given the previous discussion, we decided to follow Gelman's recommendations for prior distributions on variance parameters in hierarchical generalized linear models (Gelman, 2006). Specifically, we assigned a noninformative diffuse uniform prior for the standard deviation σ, i.e. we assume

$$\boldsymbol{\beta} \sim MN(\mathbf{0}, \mathbf{I}), \quad \sigma \sim U(0, 100), \quad \boldsymbol{\beta} \perp \sigma; \tag{4.6}$$

see Table 6 (vii) for the estimates.

Prior	β_0	β_1	α	τ
(i): $\mathbb{E}[\tau]=1$, $Var[\tau]=3/4$, $\boldsymbol{\beta} \perp \tau$	-0.3713	-0.1930	0.6253	1.5690
	(0.0989)	(0.0988)	(0.0669)	(0.5906)
(ii): $\mathbb{E}[\tau]=1$, $Var[\tau]=1$, $\boldsymbol{\beta} \perp \tau$	-0.2596	-0.1342	0.6247	3.5551
	(0.0924)	(0.0778)	(0.0671)	(1.8139)
(iii): $\mathbb{E}[\tau]=4$, $Var[\tau]=1/4$, $\boldsymbol{\beta} \perp \tau$	-0.8550	-0.4406	0.6302	0.2615
	(0.1506)	(0.2034)	(0.0669)	(0.0319)
(iv) $\mathbb{E}[\tau]=3/2$, $Var[\tau]=3/4$, $\boldsymbol{\beta} \not\perp \tau$	-0.3941	-0.2010	0.6361	1.4329
	(0.11031)	(0.1055)	(0.0651)	(0.6501)
(v) $\mathbb{E}[\tau]=4$, $Var[\tau]=1/4$, $\boldsymbol{\beta} \not\perp \tau$	-0.7638	-0.3435	0.6313	0.2425
	(0.1417)	(0.2135)	(0.0655)	(0.0304)
(vi): $\mathbb{E}[\tau]=4$, $Var[\tau]=1$, $\boldsymbol{\beta} \not\perp \tau$	-0.2088	-0.1082	0.6269	4.6208
	(0.0018)	(0.0027)	(0.0045)	(1.1730)
(vii): $\sigma \sim U(0,100)$, $\boldsymbol{\beta} \perp \sigma$ as in (4.6)	-0.9521	-0.7682	0.6451	0.1307
	(0.6001)	(0.4552)	(0.0649)	(0.1515)

Table 6: Posterior means and standard deviations of the parameters across different prior specification.

Finally, we include a treatment-specific parameter α_2 in the autoregression of the latent variables. We assume the following distribution of Z_{ij} for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$ "given the past"

$$Z_{ij}|Z_{ij-1} \sim N\left(\beta_0 + \beta_1 x_i + \alpha_1 z_{ij-1} + \alpha_2 x_i z_{ij-1}, \frac{1}{\tau}\right), \tag{4.7}$$

and β_0, β_1,α_1,α_2 and τ a priori independent, with standard gaussian distributions for all parameters except τ, which is given a gamma distribution with $\mathbb{E}[\tau] = 3/2$ and $\text{Var}[\tau] = 3/4$. Such model allows for an autoregressive interaction effect, i.e. different linear trends for T and P groups. Table 7 shows the inferences. The posterior distribution of α_2 is not significantly away from 0, which suggests that its addition to the model does not produce substantial changes with respect to what was already

reported. In other words, the data suggest that there is no such interaction effect. See Tables 8 and 9 for predictive probabilities.

As a final remark on predictive probabilities of recurrence for a new subject, Table 3 and 9 show that both sets of predictions are somewhat sensitive to the introduced prior changes, as to be expected from the changes in the parameters discussed earlier. Nevertheless, we still observe the decreased pattern, and the highest recurrence probability values for the P-group. However the different τ-prior specification has an effect on predictive probabilities at the first three visits, but later the sensitivity with respect to the prior assumption fades away.

β_0	β_1	α_1	α_2	τ
-0.7946	-0.3418	0.6333	0.0205	1.6031
(0.6470)	(0.3889)	(0.0903)	(0.1960)	(0.6652)

Table 7: Posterior means and standard deviations of the parameters for model (4.7).

Predictive probabilities of recurrence							
Prior/Model	$Y_{10,9}$	$Y_{17,10}$	$Y_{24,11}$	$Y_{34,13}$	$Y_{60,10}$	$Y_{71,9}$	$Y_{74,12}$
as in Table 6(ii)	0.4351	0.0652	0.4550	0.0674	0.0645	0.0514	0.0489
as in Table 6(vi)	0.4364	0.0663	0.4522	0.0642	0.0620	0.0485	0.0508
(4.7)	0.4567	0.0673	0.4628	0.0654	0.0628	0.0498	0.0518

Table 8: Estimates of the predictive probabilities in a new measurement for subject 10,17,24,34 (placebo) and 60,71,74 (treatment).

5 Conclusions

We have presented a Bayesian generalized linear model for longitudinal binary sequences. It has been applied to a medical study conducted in the USA by the VACURG (Veterans Administration Cooperative Urological Research Group) about recurrence of Stage I bladder cancer. Data available from this study consist of the cumulative

Predictive probabilities of recurrence												
Prior/Model	1	2	3	4	5	6	7	8	9	10	11	12
as in Table 6(ii)	0.5001	0.2875	0.1932	0.1515	0.1266	0.1086	0.0978	0.0905	0.0909	0.0902	0.0868	0.0862
	0.5003	0.3576	0.2897	0.2472	0.2210	0.2070	0.1990	0.1946	0.1937	0.1905	0.1853	0.1802
as in Table 6(vi)	0.5004	0.2886	0.1915	0.1500	0.1240	0.1105	0.0998	0.0964	0.0939	0.0916	0.0848	0.0869
	0.5102	0.3876	0.2983	0.2574	0.2375	0.2184	0.2093	0.2074	0.1955	0.1934	0.1879	0.1867
(4.7)	0.4537	0.2326	0.1873	0.1463	0.1182	0.1033	0.0945	0.0837	0.0816	0.0763	0.0701	0.0698
	0.4672	0.3567	0.3636	0.3353	0.2863	0.2764	0.2852	0.2851	0.2277	0.2198	0.2076	0.1982

Table 9: Estimates of the predictive probabilities for two new subjects in the T (first row) and P (second row) groups at every measurement times.

counts of recurrent tumors at each visit for 82 patients, 36 assigned to treatment group (thiotepa) and 46 assigned to the placebo group. We have no further information about the patients in the clinical trial.

Generally, factors like gender, age or other clinical variables that can be related with the illness constitute very useful information about the evaluation of a treatment effect. In our case, the only covariate x_i we can assume in the model represents the treatment group assigned to patient i. Despite the lack of information in our case study, we can conclude that thiotepa treatment has a sensible effect on preventing recurrence of Stage I bladder cancer. In fact, the posterior mean of β_1 indicates that the treatment (thiotepa) contributes to reduce the probability of a recurrence in the number of tumors of the bladder cancer study. Since the estimate of β_1 assumes a negative value, thus latent variables for individuals belonging to the treatment group will have a lower probability of assuming positive values than latent variables corresponding to individuals in the placebo group. This fact leads to a lower probability of $Y_{ij} = 1$, i.e. a smaller probability of a recurrence in the number of tumors for individuals belonging to the treatment group.

Furthermore, the Markovian structure we introduce in the latent variables seems to be very useful to model the correlation of the longitudinal values. The parameter α, that represents the dependence of the latent variables on the previous one, is very robust as we can conclude from the sensitivity analysis that we have conducted. On the other hand, extra efforts are needed to specify the prior distribution for τ, the precision of the latent variables, by the extreme sensitivity of the posterior estimates to the prior specification. However, it could be very interesting to analyze the same

set of data with more covariates representing for instance age and other factors that could have an influence on the recurrence in bladder cancer tumors, or to apply the same model to a case study with more covariates.

Acknowledgments

Alessandra Guglielmi was partially funded by MIUR-Italy, grant 2006/134525. Fernando Quintana was partially funded by grants FONDECYT 1060729 and Laboratorio de Análisis Estocástico PBCT-ACT13.

APPENDIX

In this section we briefly illustrate the full conditionals required in the Gibbs sampler to provide the Bayesian estimates of $\boldsymbol{\theta}$. The calculations are straightforward; more details can be found in Giardina (2008). The full conditional of τ is

$$\tau|\boldsymbol{\theta}_{-\tau} \sim gamma\left(\frac{n_0^* S_0^*}{2}, \frac{n_0^*}{2}\right)$$

where

$$\frac{n_0^* S_0^*}{2} = \frac{n_0 S_0 + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(z_{ij} - \boldsymbol{\beta}' \mathbf{w}_{ij}\right)^2 + \sum_{i=1}^m \left(z_{i1} - \mu_0\right)^2}{2}, \; \frac{n_0^*}{2} = \frac{n_0 + \sum_{i=1}^m n_i}{2},$$

if the prior is as in (3.6), while

$$\frac{n_0^* S_0^*}{2} = \frac{n_0 S_0 + \left(\frac{\beta_0^2}{v_1} + \frac{\beta_1^2}{v_2} + \frac{\alpha^2}{v_3}\right) + \sum_{i=1}^m \left(z_{i1} - \mu_0\right)^2 + \sum_{i=1}^m \sum_{j=2}^{n_i} \left(z_{ij} - \beta_0 - \beta_1 x_i - \alpha z_{ij-1}\right)^2}{2},$$

and $n_0^*/2 = \left(n_0 + \sum_{i=1}^m n_i + 3\right)/2$ under (4.5). Of course, $\boldsymbol{\theta}_{-\tau}$ is the standard way to denote all components of $\boldsymbol{\theta}$ except for τ.

Because of the Markovian structure (see equation (3.3)), the full-conditional distribution of Z_{ij} for all $j = 2, \ldots, n_i - 1$ and any subject $i = 1, \ldots, m$, is a function of both the previous and the next latent variables Z_{ij-1} and Z_{ij+1}

$$Z_{ij}|\boldsymbol{\theta}_{-Z_{ij}} \sim \begin{cases} N_{[0,+\infty)}\left(\mu_{Z_{ij}}, \tau_{Z_{ij}}^{-1}\right) & \text{if } Y_{ij} = 1 \\ N_{(-\infty,0)}\left(\mu_{Z_{ij}}, \tau_{Z_{ij}}^{-1}\right) & \text{if } Y_{ij} = 0 \end{cases},$$

where $\mu_{Z_{ij}} = ((1-\alpha)(\beta_0+\beta_1 x_i)+\alpha(z_{ij+1}+z_{ij-1}))/(1+\alpha^2)$, $\tau_{Z_{ij}}^{-1} = (\tau(1+\alpha^2))^{-1}$ and $N_{[0,+\infty)}\left(\mu_{Z_{ij}}, \tau_{Z_{ij}}^{-1}\right)$ denotes the distribution proportional to the normal with mean $\mu_{Z_{ij}}$ and variance $\tau_{Z_{ij}}^{-1}$ restricted to the interval $[0,+\infty)$.

Analogously,

$$Z_{i1}|\boldsymbol{\theta}_{-Z_{i1}} \sim \begin{cases} N_{[0,+\infty)}\left(\mu_{Z_{i1}}, \tau_{Z_{i1}}^{-1}\right) & \text{if } Y_{i1}=1 \\ N_{(-\infty,0)}\left(\mu_{Z_{i1}}, \tau_{Z_{i1}}^{-1}\right) & \text{if } Y_{i1}=0 \end{cases},$$

where $\mu_{Z_{i1}} = (\mu_0+\alpha(z_{i2}-\beta_0-\beta_1 x_i))/(1+\alpha^2)$, $\tau_{Z_{i1}}^{-1} = (\tau(1+\alpha^2))^{-1}$, and

$$Z_{in_i}|\boldsymbol{\theta}_{-Z_{in_i}} \sim \begin{cases} N_{[0,+\infty)}\left(\mu_{Z_{in_i}}, \tau_{Z_{in_i}}^{-1}\right) & \text{if } Y_{in_i}=1 \\ N_{(-\infty,0)}\left(\mu_{Z_{in_i}}, \tau_{Z_{in_i}}^{-1}\right) & \text{if } Y_{in_i}=0 \end{cases},$$

$\mu_{Z_{in_i}} = \beta_0+\beta_1 x_i+\alpha z_{in_i-1}$ and $\tau_{Z_{in_i}}^{-1} = \tau^{-1}$.

On the other hand, from (3.7) we have

$$\log(\pi(\boldsymbol{\beta}|\boldsymbol{\theta}_{-\boldsymbol{\beta}})) \propto -\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\beta}_m)'\mathbf{V}^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_m) - \frac{\tau}{2}\sum_{i=1}^{m}\sum_{j=2}^{n_i}(z_{ij}-\beta'\mathbf{w}_{ij})^2,$$

so that the full-conditional distribution of $\boldsymbol{\beta}$ (of a conjugate type) is a multivariate normal distribution

$$\boldsymbol{\beta}|\boldsymbol{\theta}_{-\boldsymbol{\beta}} \sim MN(\boldsymbol{\beta}_m^*, \mathbf{V}^*)$$

where

$$\boldsymbol{\beta}_m^* = \mathbf{V}^*\left(\mathbf{V}^{-1}\boldsymbol{\beta}_m + \tau\sum_{i=1}^{m}\sum_{j=2}^{n_i}\mathbf{w}_{ij}z_{ij}\right) \text{ and } \mathbf{V}^* = \left(\mathbf{V}^{-1}+\tau\sum_{i=1}^{m}\sum_{j=2}^{n_i}\mathbf{w}_{ij}\mathbf{w}_{ij}'\right)^{-1}.$$

References

Byar, D. P., Blackard, C., and the Veterans Administration Cooperative Urological Research Group (1977) Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of Stage I bladder cancer. *Urology*, **10**, 556-561.

Cowles, M.K. and Carlin, B.P. (1996) Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review. *Journal of the American Statistical Association,* **91**, 883-904.

Davis, C.S. (1996) Non-parametric methods for comparing multiple treatment groups to a control group, based on incomplete non-decreasing repeated measurements. *Statistics in Medicine*, **15**, 2509-2521.

Davis, C. S. and Wei, L.J. (1988) Nonparametric Methods for Analyzing Incomplete Nondecreasing Repeated Measurements. *Statistics in Medicine*, **44**, 1005-1018.

Diggle P.J., Heagerty P, Liang K.Y., and Zeger S.L. (2002) *The analysis of Longitudinal Data*. Oxford: Oxford University Press.

Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis*, **3**, 515-534.

Giardina, F. (2008) *Bayesian First Order Autoregressive Latent Variable Models For Multiple Binary Sequences*. MSc Dissertation, Politecnico di Milano, Milan, Italy.

Gilks, W.R., Richardson, S., Roberts, G.O. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman and Hall.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000) WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, **10**, 325-337.

Quintana, F. A. and Müller, P. (2004) Optimal Sampling for Repeated Binary Measurements. *Canadian Journal of Statistics*, **32**, 73-84.

Quintana, F. A. and Newton, M. A. (1998) Assessing the order of dependence for partially exchangeable binary data. *Journal of the American Statistical Association*, **93**, 194-202.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman and Hall.

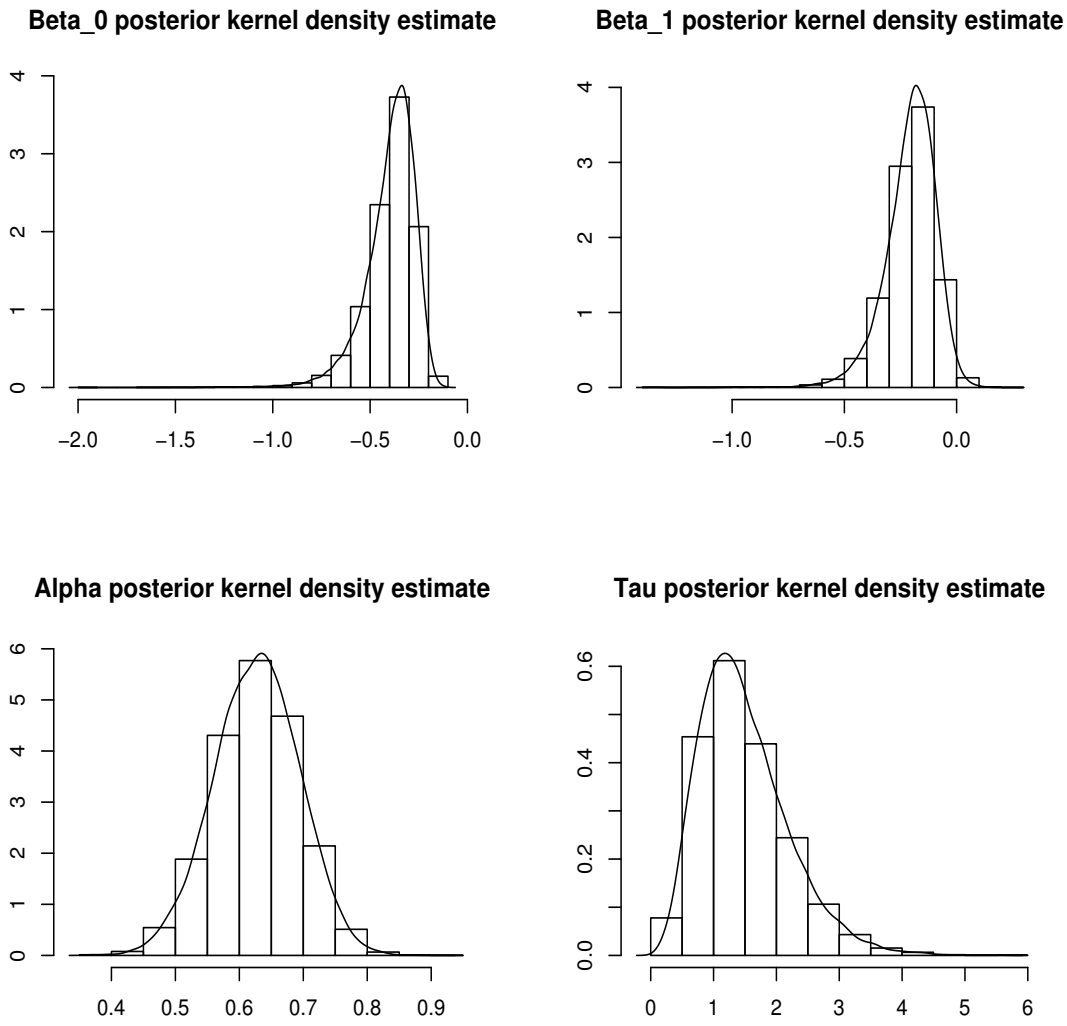

Figure 1: Posterior kernel density plots of the parameters under prior (4.1).

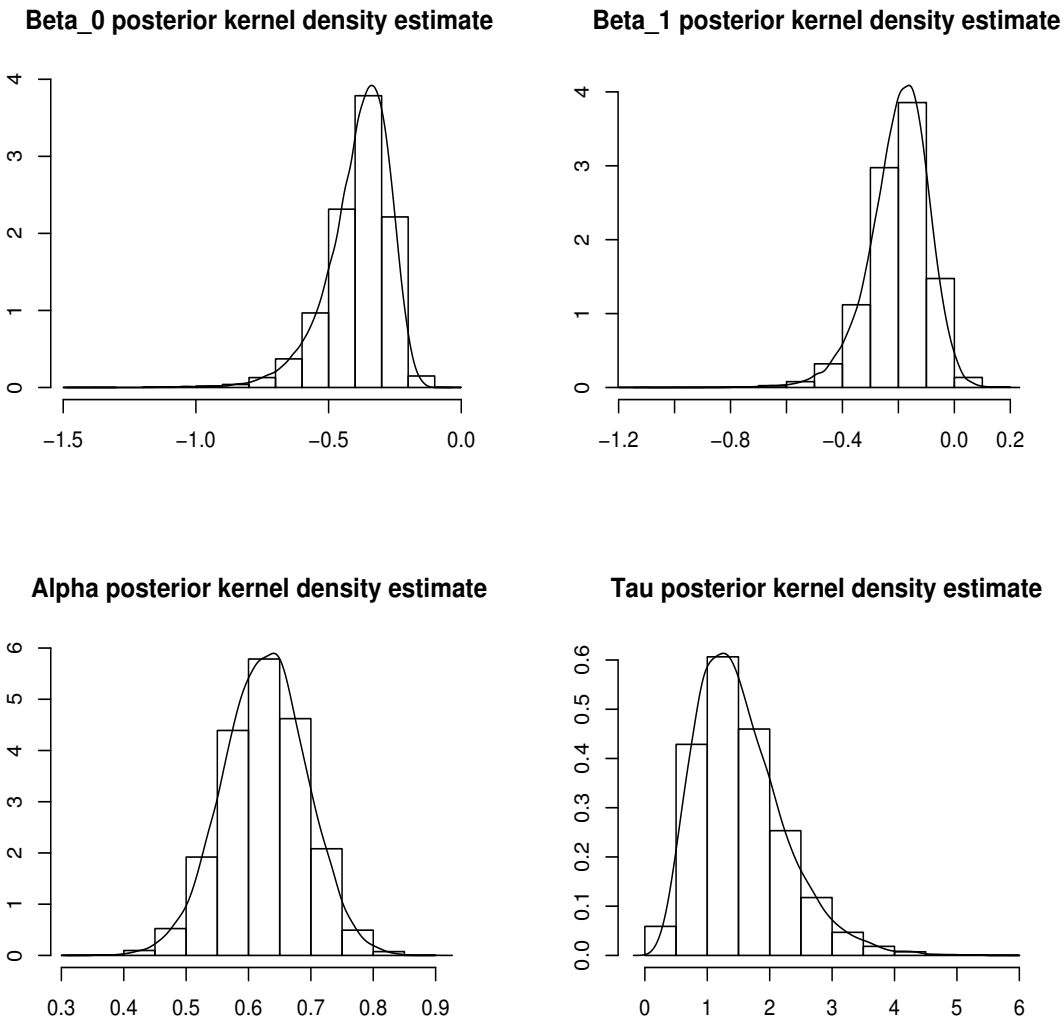

Figure 2: Posterior kernel density plots of the parameters under prior (4.2).

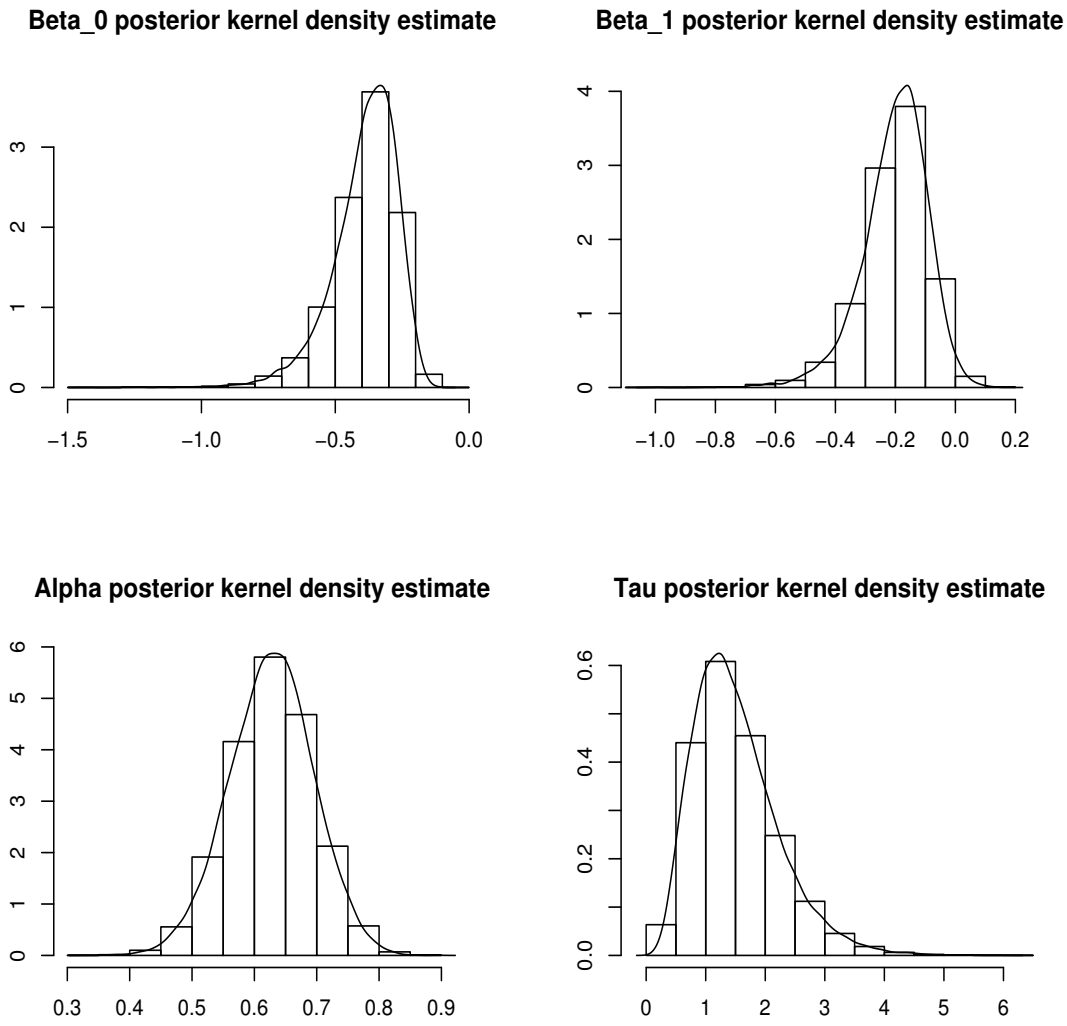

Figure 3: Posterior kernel density plots of the parameters under prior (4.3).

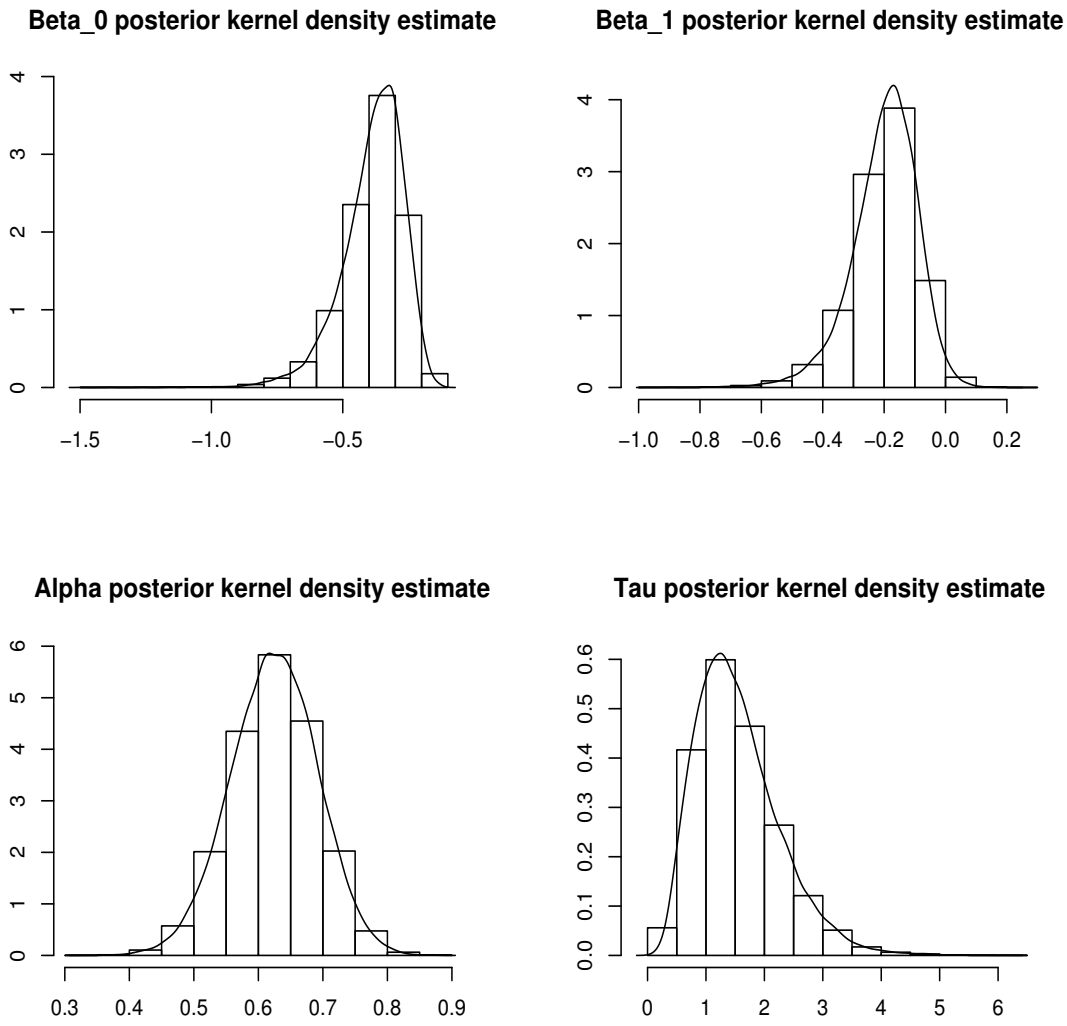

Figure 4: Posterior kernel density plots of the parameters under prior (4.4).