Czado, Claudia; Song, Peter X.-K.

Working Paper

State space mixed models for longitudinal obsservations with binary and binomial responses

Discussion Paper, No. 503

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Czado, Claudia; Song, Peter X.-K. (2006) : State space mixed models for longitudinal obsservations with binary and binomial responses, Discussion Paper, No. 503, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München,
https://doi.org/10.5282/ubm/epub.1868

This Version is available at:
https://hdl.handle.net/10419/31027

STATE SPACE MIXED MODELS FOR LONGITUDINAL OBSERVATIONS WITH BINARY AND BINOMIAL RESPONSES

Claudia Czado
SCA Zentrum Mathematik
Technische Universität München, Germany

Peter X.-K. Song
Department of Statistics and Actuarial Science
University of Waterloo, Canada

June 20, 2006

Abstract

We propose a new class of state space models for longitudinal discrete response data where the observation equation is specified in an additive form involving both deterministic and random linear predictors. These models allow us to explicitly address the effects of trend, seasonal or other time-varying covariates while preserving the power of state space models in modeling serial dependence in the data. We develop a Markov chain Monte Carlo algorithm to carry out statistical inference for models with binary and binomial responses, in which we invoke de Jong and Shephard's (1995) simulation smoother to establish an efficient sampling procedure for the state variables. To quantify and control the sensitivity of posteriors on the priors of variance parameters, we add a signal-to-noise ratio type parameter in the specification of these priors. Finally, we illustrate the applicability of the proposed state space mixed models for longitudinal binomial response data in both simulation studies and data examples.

Keywords and phrases: Longitudinal data, Markov chain Monte Carlo, probit, random effects, regression, seasonality, signal-to-noise ratio.

1 INTRODUCTION

In this paper we consider a time series of discrete observations, $\{Y_t, t = 1, \ldots, T\}$, where Y_t may be either binary or binomial, associated with p time-varying covariates $x_{t1}, \ldots, x_{tp}$. The primary objective is to model both the mean of the observed process as a function of the covariates, and the pattern of serial correlation in the data.

Among available models for time series of dichotomous observations, the class of state space models or parameter-driven models (Cox, 1981) seems to have gained a great deal of popularity. See for example Fahrmeir (1992), Carlin and Polson (1992), and Song (2000), among others. A binary state space model consists of two processes: In the first observed process $\{Y_t\}$, the conditional distribution of Y_t given the q-dimensional state variable $\boldsymbol{\theta}_t$ is Bernoulli, namely $Y_t|\boldsymbol{\theta}_t \sim$ Bernoulli(μ_t), where the conditional mean or the conditional probability of success $\mu_t = P(Y_t = 1|\boldsymbol{\theta}_t)$ follows the observation equation,

$$\mu_t = h\left(G_t'\boldsymbol{\theta}_t\right) \tag{1.1}$$

with a given link function $h^{-1}(\cdot)$ as in generalized linear models (e.g. McCullagh and Nelder, 1989) and a known q-dimensional vector G_t comprised of the time-varying covariates. In the second process, the state variables $\{\boldsymbol{\theta}_t\}$ are assumed to follow a q-dimensional Markov process, governed by the state equation,

$$\boldsymbol{\theta}_t = H_t\boldsymbol{\theta}_{t-1} + \epsilon_t, \tag{1.2}$$

where H_t is a $q \times q$-dimensional transition matrix and the error vector ϵ_t has zero mean. For the special case of one-dimensional state process with $q = 1$, two common models used in the literature are the random walk and Box and Jenkins' AR(1) process. The random walk is a non-stationary process with the transition matrix being 1, $H_t \equiv 1$, while the AR(1) process is stationary with the transition matrix being a constant, $H_t = \gamma \in (-1, 1)$, which is a parameter representing the autocorrelation coefficient. One essential difference between these two types of processes is rooted in their variances: the AR(1) has a bounded variance but the random walk has an unbounded variance that effectively increases in time.

The above class of binary state space models has been widely used for the analysis of longitudinal discrete data. For example, it has been applied to analyze the Tokyo rainfall data by many authors (e.g. Kitagawa, 1987; Fahrmeir, 1992; Carlin and Polson, 1992; Song, 2000). The rainfall data, reported by Kitagawa (1987), consist of the daily number of occurrences of rainfall in Tokyo area during years 1983 and 1984. The central question of the analysis is to model the probability μ_t of rainfall for each calendar day over a year. The proposed state space model in these papers is given by

$$\mu_t = h(\theta_t) \quad \text{and} \quad \theta_t = \theta_{t-1} + \epsilon_t, \text{with } \epsilon_t \overset{iid}{\sim} N(0, \sigma^2), \tag{1.3}$$

where θ_t may be thought of essentially as a certain underlying meteorological variable such as moisture most directly responsible for rainfall. One obvious limitation of this model formulation is that it does not allow to examine directly the seasonal rainfall cycle.

In many practical studies, assessing covariate effects is of primary interest. To address such a need, we may extend the state space model (1.3) by allowing seasonal covariates to reflect the

nature of meteorological periodicity. That is, the conditional expectation μ_t of the observed process takes the form:

$$\mu_t = h(\eta_t + G_t'\boldsymbol{\theta}_t),\ t = 1, \ldots, T,$$

where η_t is the deterministic predictor that may take the form of the decomposition model (Brockwell and Davis, 1996), $\eta_t = m_t + s_t$. Here m_t and s_t represent the trend and seasonal components, respectively, and both may be modeled further as linear functions of covariates. For convenience, we may combine the trend and seasonal components into one expression, $\mathbf{x}_t'\alpha$, where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp})'$ and parameter vector α consists of p regression coefficients to be estimated.

The inclusion of the deterministic predictor, $\eta_t = \mathbf{x}_t'\alpha$, is appealing as it would enable us to quantify the relationship between the probability of success and covariates and to allow hypothesis testing on covariates, both of which are very often of scientific interest. It is noted that state space models that contain both deterministic and random predictors (η_t and $G_t'\boldsymbol{\theta}_t$, respectively) have been proposed in other settings, for instance, in the analysis of longitudinal count data by Zeger (1988), Chan and Ledolter (1995), Jørgensen et al. (1999), and Fahrmeir and Lang (2001a; 2001b).

In this paper we develop a Markov chain Monte Carlo (MCMC) estimation procedure for the proposed state space models with the observation equation given by

$$\mu_t = h(\mathbf{x}_t'\alpha + \theta_t),\ t = 1, \ldots, T, \tag{1.4}$$

where the state variable θ_t follows a univariate AR(1) process. Clearly, θ_t represents a time-specific effect on the observed process. Because of the similarity of model representation (1.4) to the generalized linear mixed models (e.g. Diggle et al., 2002, Chapter 9), we refer the proposed models to as *the state space mixed models*. The MCMC-based inference in the state space modeling has been widely studied in the literature; see for example, Carlin et al. (1992), Carter and Kohn (1994), Frühwirth-Schnatter (1994), Gamerman (1998), Pitt and Shephard (1999), among others. In the part of sampling the latent θ_t process, we invoke an efficient sampler, i.e. the simulation smoother proposed by de Jong and Shephard (1995), to facilitate our MCMC procedure.

The organization of this paper is as follows. In Section 2 we discuss the formulation of the state space mixed models for binary time series and the MCMC algorithm for estimation. Section 3 concerns the models for binomial time series and the MCMC estimation. A simulation experiment is given in Section 4 to evaluate the proposed MCMC algorithm. Section 5 presents two data analysis examples to illustrate the proposed models and methods. Finally we make some concluding remarks in Section 6. All technical details are listed in two appendices.

2 BINARY STATE SPACE MIXED MODELS

We start with state space mixed models for binary response variables, in which a latent variable representation is utilized to develop MCMC algorithms for parameter estimation.

2.1 MODEL FORMULATION

Consider a binary longitudinal data $(Y_t, \mathbf{x}_t), t = 1, \ldots, T$, where the binary response vector is denoted by $\mathbf{Y}_T^* = (Y_1, \cdots, Y_T)'$. In this paper, we adopt the so-called threshold approach (e.g. Albert and Chib, 1993) where Y_t is generated through dichotomization of an underlying continuous process Z_t, given by the one-to-one correspondence

$$Y_t = 1 \Longleftrightarrow Z_t \leq 0, t = 1, \cdots, T. \tag{2.1}$$

With the unobservable or latent threshold variable vector $\mathbf{Z}_T^* = (Z_1, \cdots, Z_T)'$, a binary state space mixed model can be rewritten as follows,

$$Z_t = -\mathbf{x}_t'\alpha - \theta_t + u_t, \ t = 1, \cdots, T, \tag{2.2}$$

$$\theta_{t+1} = \gamma\theta_t + \epsilon_t, \ t = 0, \cdots, T. \tag{2.3}$$

We further assume that both error terms are independent and normally distributed,

$$u_t \overset{iid}{\sim} N(0, 1) \text{ and } \epsilon_t \overset{iid}{\sim} N(0, \sigma^2),$$

and hence the expressions (2.2) and (2.3) together represent a linear Gaussian state space model. As in most of hierarchical model specifications, we assume mutual independence between u_t's and ϵ_t's. This implies that given θ_t, Z_t is conditionally independent of the other Z_t's and θ_t's. In addition, the initial condition is assumed to be $\theta_0 = 0$ and $\theta_1 \sim N(0, \frac{\sigma_0^2}{1-\gamma^2})$, which is the unconditional distribution of the AR(1) process and is a standard choice for near unit root AR(1) processes (e.g. Schotman, 1994, pp. 590-591). In this paper, we focus on the stationary case, i.e. $|\gamma| < 1$.

It follows from the one-to-one relationship (2.1) that the marginal distribution of Y_t given both state variable θ_t and covariate vector $\mathbf{x}_t$ follows a probit model of the form:

$$\mu_t = P(Y_t = 1|\theta_t, \mathbf{x}_t) = \Phi(\mathbf{x}_t'\alpha + \theta_t)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0, 1)$. In other words, a combination of (2.2) and (2.1) give rise to (1.4) with the probit link function $h(\cdot) = \Phi(\cdot)$.

The parameters of primary interest are α and γ, based on which we may make inference on the covariates and the autocorrelation structure. However, to make forecasting or to conduct model diagnostics, we also need to estimate the state variables θ_t's and the variance σ^2 in the state equation (2.3).

We now introduce several notations useful to establish the following derivations. We denote the history vectors by $\mathbf{Y}_t^* = (Y_1, \cdots, Y_t)'$, $\mathbf{Z}_t^* = (Z_1, \cdots, Z_t)'$, and $\boldsymbol{\theta}_t^* = (\theta_1, \cdots, \theta_{t+1})'$ Also, we denote the conditional distribution of u given w by $[u|w]$, and the marginal distribution of u by $[u]$, where u and w represent two generic random variables. As usual, $N_p(\mu, \Sigma)$ denotes a p-variate normal distribution with mean vector μ and positive definite variance-covariance matrix Σ.

For the MCMC approach, we assume the following prior distributions $[\alpha, \sigma^2, \gamma] = [\alpha] \times [\sigma^2|\gamma] \times [\gamma]$, where

$$\alpha \sim N_p(\alpha_0, \Sigma_0), \tag{2.4}$$

$$\sigma^2|\gamma \sim IG(a, b(\gamma)) \quad \text{with} b(\gamma) = c^2[(1-\gamma^2)(a-1)]^{-1}, \tag{2.5}$$

$$\gamma \sim \text{Uniform}(-1, 1). \tag{2.6}$$

The hyper-parameters α_0, Σ_0, a and c are pre-specified. Here the $IG(a, b)$ denotes the inverse gamma distribution with density given by

$$f(\sigma^2) = \frac{1}{b^a \Gamma(a)} (\sigma^2)^{a+1} \exp(-\frac{1}{\sigma^2 b}),$$

with $E(\sigma^2) = [b(a-1)]^{-1}$ and $Var(\sigma^2) = [(a-1)^2(a-2)b^2]^{-1}$ if $a > 2$.

The motivation behind our choice of the hyper-parameter b in (2.5), as a function of c, γ and a, is given as follows. First, note that this prior implies

$$E(\sigma^2|\gamma) = \frac{1-\gamma^2}{c^2}. \tag{2.7}$$

Second, to balance the effect between u_t and θ_t, similar to the familiar signal-to-noise ratio, we compare the standard deviation of u_t (equal to 1) to the unconditional standard deviation of the AR(1) process, which equals to $\sigma_s = \sqrt{\frac{\sigma^2}{1-\gamma^2}}$. Hence, the resulting ratio between these two random sources is given by

$$c := \frac{1}{\sigma_s} = \sqrt{\frac{1-\gamma^2}{\sigma^2}},$$

which leads to the restriction $\sigma^2 = \frac{1-\gamma^2}{c^2}$. Finally, by comparing this restriction to (2.7), we see that the prior mean for σ^2 may be chosen in such a way that the parameter c tunes the trade-off of relative variabilities between the u_t and the θ_t.

A major gain from balancing the two sources of variations, as presented in (2.5)-(2.6), is to reduce the sensitivity of the posterior distributions on the hyper-parameters in the prior distributions, especially parameter b in the IG prior. When the prior for σ^2 is independent of γ, we conducted some simulation experiments in a similar setup as that considered in Section 4 and found that the posteriors turned out to be very sensitive to the hyper-parameter b in the IG prior (results are not shown in the paper). In particular, if the b was not close to the true value the MCMC runs would often not even converge. Therefore, in this paper we will concentrate on the conditional prior choice for σ^2 given in (2.5).

As seen above, in this paper we only consider proper priors. However, if one is willing to use improper (or flat) priors in the proposed model, refer to Sun et al. (2001) regarding necessary and sufficient conditions for the propriety of posterior distributions with improper priors in hierarchical linear mixed models. Following Theorem 2 in Section 2.3 of their paper, when one considers a situation similar to that of Case 1, two conditions are needed in order to ensure that the IG prior of the variance component σ^2 leads to a proper posterior. Note that in our model,

the rank t, the number of blocks r and the total dimension of the random effects q are all equal to 1. In this case, two sufficient conditions are (a) In the IG prior, $b > 0$ and (b) the sample size n satisfies $n > p$, where p is the number of regression parameters in α.

2.2 MONTE CARLO MARKOV CHAIN METHODS

It follows from the model specification presented in the previous section that the marginal likelihood function for the parameters $(\alpha, \gamma, \sigma^2)$ is,

$$L(\alpha, \gamma, \sigma^2 | \mathbf{Y}_T^*) = \int [y_t|\theta_t; \alpha][\theta_t|\theta_{t-1}; \gamma, \sigma^2][\theta_0] d\theta_0 d\theta_1 \cdots d\theta_T$$

where the integral is clearly $(T+1)$-dimensional, with $[y_t|\theta_t; \alpha]$ being a Bernoulli distribution and $[\theta_t|\theta_{t-1}; \gamma, \sigma^2]$ a conditional normal. A direct evaluation of this high-dimensional integral is intractable, which hampers us from directly calculating the maximum likelihood estimates for these parameters. Instead, we tackle this problem by taking a Bayesian perspective and utilizing Markov Chain Monte Carlo (MCMC) methods. MCMC methods allow us to draw sufficiently large number of joint samples from the posterior distribution $[\alpha, \boldsymbol{\theta}_T^*, \sigma^2, \gamma, \mathbf{Z}_T^* | \mathbf{Y}_T^*]$. Using these samples, we can easily make inference for these parameters via, for example, their marginal posteriors. Readers unfamiliar with MCMC methods can consult Gilks et al. (1996), Gamerman (1997), Liu (2001) and Robert and Casella (2004) for introductory material, algorithms, convergence diagnostics and applications.

For the Bayesian approach, we use the prior specifications given in (2.4)-(2.6) and follow Tanner and Wong's (1987) Gibbs Sampling approach with data augmentation. In particular we will update the regression parameters and state variables jointly using the simulation smoother of de Jong and Shephard (1995), since the model of (2.2)-(2.3) is a member of the general state space models considered by de Jong (1991). This joint update of α and θ_T^* is considerably faster than separate updates considered in an earlier work by the authors of this paper. The adapted simulation smoother for the proposed model and the other ordinary Gibbs updates are shown in Appendix A in detail.

2.3 MODEL SELECTION

To assess the goodness-of-fit for a proposed model and to compare among several candidate models, we adopt Spiegelhalter et al.'s (2002) deviance information criterion (DIC) for model selection within the MCMC framework. Like the likelihood ratio tests and Akaike's information criterion (Akaike, 1973), the DIC serves a measure that reasonably balances between model fit and model complexity. To calculate the DIC, we need the Bayesian deviance given by

$$D(\alpha, \boldsymbol{\theta}_T^*) = -2 \log L(\alpha, \boldsymbol{\theta}_T^*) = -2 \sum_{t=1}^{T} [Y_t \log(\mu_t) + (1 - Y_t) \log(1 - \mu_t)]$$

for a model with mean μ_t and parameters $(\alpha, \boldsymbol{\theta}_T^*)$, and the value of the true number of free parameters defined as

$$p_D = \mathrm{E}_{\alpha, \boldsymbol{\theta}_{T,\gamma,\sigma^2|Y_T^*}^*} D(\alpha, \boldsymbol{\theta}_T^*) - D(E_{\alpha|Y_T^*}(\alpha), E_{\boldsymbol{\theta}_{T}^*|Y_T^*}(\boldsymbol{\theta}_T^*)) =: \overline{D} - D(\bar{\alpha}, \bar{\boldsymbol{\theta}}_T).$$

Therefore, the deviance information criterion (DIC) takes the form

$$\text{DIC} = \overline{D} + p_D = D(\bar{\alpha}, \bar{\boldsymbol{\theta}}_T) + 2p_D,$$

where as usual, $\overline{D}$ explains the model fit and p_D indicates the model complexity. Spiegelhalter et al. showed in some asymptotic sense that the DIC is a generalization of Akaike's information criterion.

Computing the DIC is straightforward in an MCMC implementation. Monitoring both $(\alpha, \boldsymbol{\theta}_T^*)$ and $D(\alpha, \boldsymbol{\theta}_T^*)$ in MCMC updates, at the end one can estimate the $\overline{D}$ by the sample mean of the simulated values of D and the $D(\bar{\alpha}, \bar{\boldsymbol{\theta}}_T)$ by plugging in the sample means of the simulated posterior values of α and $\boldsymbol{\theta}_T^*$. A lower value of DIC indicates a better-fitting model.

3 BINOMIAL STATE SPACE MIXED MODELS

We now consider longitudinal data with binomial response variables. This section will show how the MCMC algorithm developed for binary responses can be extended to binomial responses. To begin, we assume that aggregating n_t independent Bernoulli trials $Y_{it}, i = 1, \cdots, n_t$ gives rise to the binomial response $Y_t = \sum_{i=1}^{n_t} Y_{it}$. The latent variable representation (2.1) now specifies the correspondence in a component-wise fashion as follows:

$$Y_{it} = 1 \Longleftrightarrow Z_{it} \leq 0, t = 1, \cdots, T, i = 1, \cdots, n_t. \tag{3.1}$$

At a given time T, let $N = \sum_{t=1}^{T} n_t$ be the total number of Bernoulli trials, and denote the history vectors by

$$\mathbf{Y}_T^* = (Y_{11}, \cdots, Y_{n_1,1}, \cdots, Y_{1T}, \cdots, Y_{n_T,T})' \text{ and } \mathbf{Z}_T^* = (Z_{11}, \cdots, Z_{n_1,1}, \cdots, Z_{1T}, \cdots, Z_{n_T,T})'.$$

Likewise, assume that the latent vector $\mathbf{Z}_T^*$ follows component-wise state space formulation,

$$Z_{it} = -\mathbf{x}_t'\alpha - \theta_t + u_{it}, \ i = 1, \cdots, n_t, t = 1, \cdots, T, \tag{3.2}$$

$$\theta_t = \gamma\theta_{t-1} + \epsilon_t, \ t = 1, \cdots, T, \tag{3.3}$$

where $u_{it} \overset{iid}{\sim} N(0,1)$ and $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$. Similarly mutual independence between the two sets of innovations $\{u_{it}, i = 1, \cdots, n_t, t = 1, \cdots, T\}$ and $\{\epsilon_t, t = 1, \cdots, T\}$ is imposed.

It is noted that the only difference between the binary and binomial models appears in the dimension of the observed processes. In effect, the binomial case can be regarded as an aggregation of a number of independent binary copies, each driven by the same state process. The immediate implication of this fact is that we need only to modify the updating procedures for the α and θ_t's, since only these two are effectively affected by such a dimension expansion. Refer to Appendix B for the detail regarding updates of state variables θ_t and the regression parameter α with the utility of de Jong and Shephard's (1995) simulation smoother technique.

On the other hand, since the updates of both autocorrelation γ and variance σ^2 do not involve the latent variables Z_{it}'s, the sampling procedures developed for the binary case in Appendix A remain valid in the binomial case.

4 A SIMULATION EXPERIMENT

To investigate the performance of the MCMC algorithm and different prior specifications for the proposed model we conducted a simulation study with the same covariate setup as for the infant sleep data investigated later. It is based on the following binary state space mixed model:

$$Z_t = -\alpha_0 - \alpha_1 x_{1t} - \alpha_2 x_{2t} - \theta_t + u_t,\ t = 1, \cdots, 121, \tag{4.1}$$

$$\theta_t = \gamma\theta_{t-1} + \epsilon_t,\ t = 1, \cdots, 121, \tag{4.2}$$

where x_{t1} and x_{t2} are the covariates defined in Section 5.2. The true parameter values are chosen to be similar to those obtained in the data analysis. In particular, the true values are $\alpha_0 = .1$, $\alpha_1 = -.45$, $\alpha_2 = .25$ and $\gamma = .96$. The independent innovations are $u_t \overset{iid}{\sim} N(0,1)$ and $\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ with the true value $\sigma = .45$. The starting condition was set as $\theta_0 \sim N(0, \frac{\sigma^2}{1-\gamma^2})$.

The specific purposes of this simulation are (i) to evaluate the effectiveness and the variability of the proposed MCMC algorithm based on two data sets generated from (4.1)-(4.2), and (ii) to assess how the different hyper-parameter choices in the priors of (σ^2, γ) influence the length of burn-in and other convergence issues. In particular, we investigated two levels of $c = 1$ and $c = .65$. Here $c = 1$ reflects an equal sharing of the variability between the regression error u_t and the AR(1) innovation ϵ_t, while $c = .65$ allows the variability from the AR(1) process to dominate over that of the regression error term. In addition, throughout this simulation we used the independent normal priors with zero mean and variance$= 500^2$ for the regression parameters $\alpha_j, j = 0, 1, 2$. The MCMC algorithms were run for 20000 iterations with every 20th iteration recorded. By examining the time plots we decided to choose a burn-in of 10 recorded iterations in the thinned chain. The MCMC calculations were implemented in the R software package. Figure 4.1 shows that there is low autocorrelation for the recorded MCMC iterates in all the cases considered.

[Figure 4.1 about here]

Figure 4.2 shows the marginal posterior density estimates of the parameters $\alpha_0, \alpha_1, \alpha_2, \gamma$ and σ for the two data sets and the two choices of c, where the vertical lines correspond to their true values. From the results we can see that in all cases the true values are well inside 90% credible intervals. A higher c value will result in a lower posterior mode estimate for σ, while the estimate of γ turns to be higher. The variation in the posterior mode estimation of the regression effects is mainly attributed to the variability of the MCMC algorithm itself. The dichotomization can result in different versions of binary data if the associated underlying processes may be very similar. This is why in general estimating parameters in models for binary data is difficult simply because the data itself does not provide rich information about the underlying processes.

[Figure 4.2 about here]

Figure 4.3 shows a comparison of the performances for the two data sets and the two choices of c in updating the state variables $\theta_t, t = 1, \cdots, 121$. The true underlying state values are indicated

by the solid lines, while the dashed and dotted lines represent posterior mean estimates and the 90% credible intervals, respectively. We see that the true values are within the 90% credible intervals except for one or two time points, and the posterior mean values for data 2 happen to be better confined in the credible intervals than data 1.

[Figure 4.3 about here]

Also the DIC values were calculated. For data 1 we had a DIC value of 108.5 (108.3) at $c = 1(c = .65)$, while for data 2 the DIC was 134.4 (133.1) at $c = 1(c = .65)$. This shows that a prior value of $c = .65$ gives a slightly better performance.

In summary, based on this simple simulation study, we have already seen that the Bayesian analysis is sensitive to the choice of the priors for (σ, γ). The magnitude of the sensitivity would depend on many factors of the proposed model (such as the level of autocorrelation), but the influence of the hyper-parameters is certainly predominant.

5 DATA ANALYSIS

5.1 ANALYSIS OF THE RAINFALL DATA

We now illustrate the application of the proposed model to analyze the Tokyo rainfall data, which has been briefly discussed in Section 1. Let Y_t be the number of occurrences of rainfall for a given calendar day t during the years 1983-1984. So, $Y_t \sim \text{Binomial}(2, p_t), t \neq 60$ and $Y_t \sim \text{Binomial}(1, p_t), t = 60$ (February 29, 1984). Therefore, we set $T = 366$ in our analysis. Most previous analyses assumed a random walk for the state process and ignored the non-stationarity of seasonality, although certain periodic patterns were clearly revealed in their studies.

The proposed binomial state space mixed model directly addresses seasonal and monthly effects through covariates $\mathbf{x}_t = (\cos 1_t, \sin 1_t, \cos 4_t, \sin 4_t, \cos 12_t, \sin 12_t)'$, where

$$\cos m_t = \cos\left(\frac{2\pi m t}{T}\right) \text{ and } \sin m_t = \sin\left(\frac{2\pi m t}{T}\right), \ m = 1, \ldots, T.$$

So the latent variables $\{Z_{it}\}$ follow

$$\begin{aligned} Z_{it} &= -\alpha_0 - \alpha_1 \cos 1_t - \alpha_2 \sin 1_t - \alpha_3 \cos 4_t - \alpha_4 \sin 4_t - \alpha_5 \cos 12_t - \alpha_6 \sin 12_t \\ &\quad -\theta_t + u_{it}, \ i = 1, 2; t = 1, \ldots, 366, \end{aligned} \tag{5.1}$$

and the state variables $\{\theta_t\}$ follow the stationary AR(1) model. In each of all the cases described below, a total of 20,000 iterations of the MCMC algorithm were run with every 10th iteration recorded. A burn-in of 100 recorded iterations was used for the posterior density estimation.

To compare several possible combinations of seasonal covariates as well as the effect of the ratio parameter c, we invoked the DIC information criterion introduced in Section 2.3. Totally, we ran

the MCMC in 9 different settings, each with a set of chosen covariates and a value of the c. The results are summarized in Table 5.1. From this table we see that the DIC values of the second submodel are steadily higher than those of the two other models, so the second combination $\mathbf{x}_t = (\cos 1_t, \sin 1_t, \cos 4_t, \sin 4_t)'$ should not be considered further. Comparing the DIC values between the first submodel and the full model (5.1), the full model appeared to have a more stable performance over different levels of the c and reached the minimum at $c = 5$. We repeated this DIC calculation a few times to monitor the possible variation due to the MCMC algorithm itself, the DIC always appeared to favor the full model. Therefore, we selected the full model in the further analysis of the data.

Table 5.1: Model Fit $\overline{D}$, effective number of parameters p_D and DIC for the Rain Fall Data.

Covariate Combination	Ratio c	$\overline{D}$	p_D	DIC
$\mathbf{x}_t = (\cos 4_t, \sin 4_t, \cos 12_t, \sin 12_t)'$	1.0	702.31	79.83	782.14
	2.0	745.67	33.76	779.42
	5.0	763.21	11.53	774.74
$\mathbf{x}_t = (\cos 1_t, \sin 1_t, \cos 4_t, \sin 4_t)'$	1.0	705.01	82.02	787.03
	2.0	737.09	50.98	788.07
	5.0	774.18	22.27	796.45
$\mathbf{x}_t = (\cos 1_t, \sin 1_t, \cos 4_t, \sin 4_t, \cos 12_t, \sin 12_t)'$	1.0	696.50	78.57	775.07
	2.0	727.19	47.25	774.43
	5.0	755.25	18.39	773.63

Figure 5.1 displays the estimated posterior densities for the regression parameters $\alpha_i, i = 0, \cdots, 6$, the standard error parameter σ and the autocorrelation parameter γ, at different values $c = 1, 2$ and 5. It is clear that the densities of the regression coefficients $\alpha_j, j = 0, 1, \ldots, 6$ appeared to be less to the c than the densities of the σ and γ. For comparison, in Figure 5.1 we also plot the the corresponding naive point estimates, indicated by the vertical lines, of the model parameters with the serial dependence being neglected. Apparently, the hierarchical modeling considered in this paper for auto-correlated data did produce differences from the marginal probit regression modeling of independent observations. Furthermore, the 90% credible intervals of the regression parameters based on the full model suggest that a yearly covariate $\cos(1_t)$, a seasonal covariate $\cos(4_t)$ and two monthly covariates $\sin(12_t)$ and $\cos(12_t)$ turn out to be important explanatory variables, at all of the c levels considered. Meanwhile, the average estimate of the autocorrelation coefficient γ over the three c levels is around .41, and such a medium sized γ is supportive to the AR model, rather than the random walk model with $\gamma = 1$, for the modeling of the state variables. In conclusion, the proposed model has identified several important seasonal covariates to explain the variability in the rainfall probability over one year period. In addition, we found the average estimate of σ is around .32 over the three c levels.

[Figure 5.1 about here]

In this analysis we also computed the pointwise estimation of the rain probability p_t at day t, $t = 1, \ldots, 366$. Here we only considered the case with $c = 5$ that corresponds to the smallest DIC.

The solid line in Figure 5.2 shows the posterior mean estimates $\bar{p}_t$, $t = 1, \ldots, 366$, which was computed by only using the deterministic component of the full model, together with its 90% credible bounds indicated by the dotted lines. The broken line, which is mostly tangled with the solid line, represents the posterior mean estimates of the probabilities computed by using both deterministic component and random component θ_t.

[Figure 5.2 about here]

To better visualize how and to what extent the the time-specific random effect θ_t affects the deterministic mean pattern, we plot the posterior mean estimates for the state variables θ_t, $t = 1, \ldots, 366$ in Figure 5.3, with 90% pointwise credible bounds. Although the random effects cannot be easily seen in Figure 5.2 after the transformation with the probit link function, they do not seem to be constant and ignorable. Summarizing what we learned from Figures 5.1-5.3, we conclude that the proposed model fits the data well, and it provides a better analysis than the marginal analysis with no random component in that no serial correlation is accounted for.

[Figure 5.3 about here]

5.2 ANALYSIS OF INFANT SLEEP DATA

A dichotomous time series $\{y_t\}$ of infant sleep status, reported by Stoffer et al. (1988), were recorded in a 120 minute EEG study. Here, $y_t = 1$ if during minute t the infant was judged to be in REM sleep and otherwise $y_t = 0$. Associated are two time-varying covariates: the number of body movements due not to sucking during minute t (x_{t1}) and the number of body movements during minute t (x_{t2}). This time series alone has been previously analyzed by Carlin and Polson (1992) and Song (2000) using a simple probit state space model without covariate effects,

$$\mu_t = \Phi(\theta_t), \text{ and } \theta_t = \gamma\theta_{t-1} + \epsilon_t, \ t = 1, \ldots, 120,$$

where the initial state $\theta_0 \sim N(0, 1)$. The state process $\{\theta_t\}$ may be regarded as an underlying continuous "sleep state" governed by a first-order stationary Markov process. Their primary goal is to estimate the probability of being in REM sleep status.

In the present paper, however, our objective of this data analysis is different from theirs. Our interest is to investigate whether or not the probability of being in REM sleep status is significantly related to the two types of body movements x_{t1} and x_{t2}. If so, the use of a deterministic predictor $\alpha_0 + \alpha_1 x_{t1} + \alpha_0 x_{t2}$ together with the random component θ_t would better interpret and predict this probability of REM sleep. Therefore our observation equation takes an extended form as follows,

$$\mu_t = \Phi(\alpha_0 + \alpha_1 x_{t1} + \alpha_2 x_{t2} + \theta_t),$$

and the state equation is AR(1), the same as in the previous analyses.

We then applied the MCMC algorithm with the conditional joint prior for (σ, γ) at different values of c. A total of 20,000 iterations with every 20th iteration recorded were run. Various graphical examinations showed that a burn-in of recorded 10 iterations was sufficient for this case. Figure 5.4 displays the marginal posterior density estimates for the regression parameters $(\alpha_0, \alpha_1, \alpha_2)$, the autocorrelation (γ) and the variance parameter (σ) together with the autocorrelations of the MCMC iterates for the different c values. This shows that the mixing of the MCMC algorithms is fast for the regression parameters and the correlation parameter, but less so for σ. For σ the autocorrelations are lower for a lower c value. For the posterior density estimates we see that the results for the regression and correlation parameter are less sensitive to the prior choice of c than for σ, where the posterior mode estimates of σ increase as c decreases. It is of further evident that a very high autocorrelation is present in the AR(1) process for state variables and thus in this binary time series.

[Figure 5.4 about here]

Table 5.2: Posterior Mean and Quantiles for the Infant Sleep Data

Posterior	c	α_0	α_1	α_2	γ	σ
Mean	2.00	0.0624	-0.4315	0.2625	0.960	0.393
	1.00	0.1140	-0.432	0.2419	0.958	0.473
	.65	0.0540	-0.4225	0.2341	0.952	0.571
5% Quantile	2.00	-1.7610	-0.7765	-0.0137	0.915	0.181
	1.00	-2.1040	-0.830	-0.0281	0.904	0.252
	.65	-2.1023	-0.8122	-0.0594	0.887	0.339
10% Quantile	2.00	-1.2050	-0.7070	0.0462	0.927	0.213
	1.00	-1.3708	-0.717	0.0213	0.919	0.282
	.65	-1.5310	-0.7333	0.0163	0.909	0.373
50% Quantile	2.00	0.0755	-0.4291	0.2665	0.966	0.355
	1.00	0.0879	-0.427	0.2399	0.965	0.434
	.65	-0.0419	-0.4146	0.2333	0.961	0.543
90% Quantile	2.00	1.3204	-0.1632	0.4861	0.986	0.590
	1.00	1.6967	-0.144	0.4536	0.986	0.727
	.65	1.8113	-0.1389	0.4692	0.985	0.804
95% Quantile	2.00	1.8083	-0.0822	0.5303	0.988	0.667
	1.00	2.2723	-0.064	0.5352	0.988	0.849
	.65	2.6322	-0.0555	0.5523	0.987	0.925

Posterior means and quantiles are given in Table 5.2. It is clear that the influence of the number of body movements (x_2) is marginal, since the corresponding 90% credible interval for α_2 contains the zero value. In contrast the influence of the number of body movements not due to sucking (x_1) is detected. The negative value of the posterior mean for α_1 shows that a higher number of body movements not due to sucking will reduce the probability of the infant being in REM sleep. This conclusion is intuitively meaningful.

The top panel of Figure 5.5 shows the posterior mean estimates for the state variables $\{\theta_t\}$ with 90% pointwise credible intervals for $c = .65$ chosen according to the smallest DIC. This shows that the state process θ_t behaves as an underlying continuous "sleep state". We also observe that the posterior mean estimates of $\bar{\theta}_t$'s have a zero mean value and stable variation over time, demonstrating no obvious lack of fit.

[Figure 5.5 about here]

Posterior mean estimates of the REM sleep state probabilities $p_t = \Phi(\alpha_0 + \alpha_1 x_{t1} + \alpha_2 x_{t2} + \theta_t)$ are presented in the bottom panel of Figure 5.5 for $c = .65$, together with the 90% credible bounds indicated by the dotted lines.

To further demonstrate the usefulness of the mixed state space models, we now investigate their predictive ability, in comparison to regular state space models. From the evidence presented below, it is clear that a pure binary state space model will have poor predictive ability simply because the state variables have zero expectation. On the other hand, the inclusion of covariates will in general improve the predictive power. To illustrate this, we consider three observation equations: (i) with both covariates x_{t1} and x_{t2}, (ii) with only x_{t1}, and (iii) with no covariates. We then run the same MCMC algorithm as before for each of the three cases with only the first 80 observations used. The out-of-sample predicted probabilities of REM sleep are computed by

$$\hat{\mu}_t = \Phi(\mathbf{x}_t'\bar{\alpha} + \hat{\theta}_t), \ t > 80$$

where $\hat{\theta}_t = \bar{\gamma}\hat{\theta}_{t-1}$. Here $\bar{\alpha}$ and $\bar{\gamma}$ denote the corresponding posterior mean estimates.

Figure 5.6 shows that the fitted probabilities for $t \leq 80$ and the predicted probabilities for $t > 80$. It is obvious that for all three cases a reasonable fit of the probabilities ($t \leq 80$) is indicated. Moreover, as expected, the pure state space model has little predictive ability, while the model with both covariates shows better predictive power by utilizing the information from the both covariates over the period $t > 80$.

Finally, we considered also the problem of model selection when $c = .65$ is chosen. For the same three cases as considered in the prediction study, we obtained the DIC values, for respective models, (i) 100.1, (ii) 100.3, and (iii) 99.57. These values do not differ much, so the models are quite close with regard to the selection of covariates. A very slight preference goes to the second model with the single covariate x_{t1}. This clearly indicates that the DIC is not suitable for the assessment of prediction capability for a model. This is because the first model with both covariates apparently have a better predictive ability than the other two models.

[Figure 5.6 about here]

6 CONCLUSIONS AND DISCUSSION

In this paper we proposed a class of state space mixed models for longitudinal discrete data, useful in finding statistical evidence for the relation between the mean of the observed process

and some covariates of interest. The models with both deterministic and random predictors are more flexible than models with no deterministic components, since they provide access to the inferential methods used in regression analysis. We show how the DIC criterion of Spiegelhalter et. al. (2002) can be applied in these models for the purpose of model selection based on goodness-of-fit, rather than based on the prediction capability. In addition, the predictive ability is improved substantially over a pure binary state space model by utilizing the information of covariates, when they are included in the proposed model explicitly. In both the simulation study and the two data analyses, the proposed MCMC algorithm based on de Jong and Shephard's (1995) simulation smoother technique was shown to work very efficiently with fast convergence.

To handle the sensitivity of posteriors to the priors for variance parameters, we introduce a signal-to-noise ratio type parameter in the specification of these priors. Such a parameter serves, to some extent, as a mirror to reflect the "honesty" of related inference and conclusions. As shown in both simulation studies and data examples, the posteriors can vary largely over different values of the ratio parameter c. Therefore, we strongly avocate, if relevant, the use of this ratio parameter in the specification of prior distributions and to report this value in the conclusions.

The proposed MCMC algorithm can be modified to deal with the logistic link function, in which the distribution for updating α will no longer be exact multivariate normal but in principle can still be done using the Metropolis-Hastings algorithm. The difference in conclusions using the two link functions is very mild, so we did not pursue any further development with the logistic link.

ACKNOWLEDGMENTS

The authors are very grateful to the two anonymous referees and the Editor for their constructive comments that led to a substantial improvement for the paper. The first author's research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistische Analyze diskreter Strukturen*. The second author's research was supported by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd Intl. Symp. on Information Theory*, (ed. B.Petrov and F.Csaki), Akademiai Kiado, Budapest.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Stat. Assoc.* **88**, 669–679.

Brockwell, P.J. and Davis, R.A. (1996). *Time Series: Theory and Methods*, 2nd Ed., Springer-Verlag, New York.

Carlin, B.P. and Polson, N.G. (1992). Monte Carlo Bayesian methods for discrete regression models and categorical time series. *Bayesian Statistics* **4**, 577–586.

Carlin, B.P., Polson, N.G. and Stoffer, D.S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association* **87**, 493-500.

Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* **81**, 541-553.

Casella, G. and George, E.I. (1992). Explaining the Gibbs Sampler. *Amer. Statistician* **46**, 167–174.

Chan, K.S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving observations. *J. Amer. Statist. Assoc.* **90**, 242–252.

Cox, D. R. (1981). Statistical analysis of time series, some recent developments. *Scand. J. Statist.* **8**, 93–115.

De Jong, P. (1991). The diffuse Kalman filter. *Ann. Statist.*, **2**, 1073–1083.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, **82**, 2, 339–350.

Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. (2002). *The Analysis of Longitudinal Data, 2nd Ed.,* Oxford University Press, Oxford.

Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models. *J. Amer. Stat. Assoc.* **87**, 501–509.

Fahrmeir, L. and Lang, S. (2001a). Bayesian inference for generalized additive mixed models based on Markov random field priors. *Appl. Statist.* **50**, 201-220.

Fahrmeir, L. and Lang, S. (2001b). Bayesian semiparametric regression analysis of multicategorical time-space data. *Ann. Inst. Statist. Math.* **53**, 11-30.

Fahrmeir, L. and Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag, Berlin.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* **15**, 183-202.

Gamerman, D. (1997). *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall/CRC, New York.

Gamerman, D. (1998). Markov chain Monte Carlo for dynamic generalized linear models. *Biometrika* **85**, 215-227.

Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. New York, Chapman and Hall.

Jørgensen, B., Lundbye-Christensen, S., Song, P.X.-K. and Sun, L. (1999). A state-space models for multivariate longitudinal count data. *Biometrika* **86**, 169–181.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of non-stationary time series (with comments). *J. Amer. Stat. Assoc.* **82**, 1032-1063.

Lee, P.M. (1997). *Bayesian Statistics : An Introduction*. Arnold, London.

Liu, J.S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer Verlag, New York.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd ed., Chapman and Hall, London.

Ortega, J.M. and Rheinboldt, W.C. (1970). *Iterative Solutions of Nonlinear Equations in Several Variables*. Academic Press, London.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*. 2nd Ed., New York, Wiley.

Robert, C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing* **5**, 121–125.

Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Verlag, New York.

Schotman, P.C. (1994) Priors for the AR(1) model: parametrisation issues and time series considerations. *Economet. Theory*,**10**, 579–595.

Song, P.X.-K. (2000). Monte Carlo Kalman filter and smoothing for multivariate discrete state space models. *Cand. J. Statist.* **28**, 641–652.

Spiegelhalter, D. J., Best, N.G., Carlin, B.P. and van der Linde, A.(2002). Bayesian measures of model complexity and fit.*J. Roy. Stat. Soc. B*, **64**, 583–639.

Stoffer, D.S., Scher, M.S., Richardson, G.A., Day, N.L. and Coble, P.A. (1988). A Walsh-Fourier analysis of the effects of moderate maternal alcohol consumption on neonatal sleep-state cycling. *J. Amer. Stat. Assoc.* **83**, 954–963.

Sun, D., Tsutakawa, R.K. and He, Z. (2001). Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica* **11**, 77–95.

Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *J. of Amer. Stat. Assoc.* **82**, 528–540.

Pitt, M.K. and Shephard, N. (1999). Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis* **20**, 63-85.

Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–629.

APPENDIX A: CONDITIONAL DISTRIBUTIONS FOR BINARY DATA

LATENT VARIABLE UPDATE:

Since the latent variables Z_t's are conditionally independent given θ_T^*, we can immediately reduce the update of $[\mathbf{Z}_T^*|\mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^{2*}, \gamma]$ to the individual updates of $[Z_t|\mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^{2*}, \gamma]$ for $t = 1, \cdots, T$. Each of these univariate conditional distribution is equivalent to $[Z_t|\mathbf{Y}_T^*, \alpha, \theta_T^*]$, since given θ_T^* the information contained in σ_T^{2*} and γ has no influence on the $\mathbf{Z}_T^*$. Moreover, we have $[Z_t|\mathbf{Y}_T^*, \alpha, \theta_T^*] = [Z_t|Y_t, \alpha, \theta_t], t = 1, \ldots, T$ due again to the conditional independence. It is easy to see that these distributions are univariate truncated normal with mean $-\mathbf{X}_t'\alpha - \theta_t$ and variance 1. Truncation interval is $(-\infty, 0]$(or $[0, \infty)$) when $Y_t = 1$(or $Y_t = 0$). We use the inversion method for the generation of truncated univariate normal random variables in the numerical implementation, proposed by Robert (1995).

STATE VARIABLES AND REGRESSION PARAMETER UPDATE:

The fact that $\mathbf{Y}_T^*$ is completely determined with given $\mathbf{Z}_T^*$ produces the following reduction, $[\alpha|\mathbf{Y}_T^*, \mathbf{Z}_T^*, \theta_T^*, \sigma_T^{2*}, \gamma] = [\alpha|\mathbf{Z}_T^*, \theta_T^*]$, which allows us to use the simulation smoother. De Jong and Shephard (1995) consider the following general state space model

$$\begin{aligned} y_t &= X_t\beta + Z_t\alpha_t + G_t\mathbf{u}_t \; t = 1, \cdots, n \\ \alpha_{t+1} &= W_t\beta + T_t\alpha_t + H_t\mathbf{u}_t \; t = 0, \cdots, n \end{aligned}$$

and $\mathbf{u}_y \sim N_m(0, \sigma^2 I_m)$ i.i.d. Comparing this model to (2.2)-(2.3), we identify $n \equiv T$, $y_t \equiv Z_t$, $X_t \equiv -\mathbf{x}_t'$, $\alpha \equiv \beta$, $Z_t \equiv -1$, $\alpha_t \equiv \theta_t$, $W_t \equiv 0$, $T_t \equiv \gamma$, $\mathbf{u}_t = (u_t, \epsilon_t)'$, $G_t \equiv (1, 0)$, $H_0 \equiv (0, \frac{\sigma}{1-\gamma^2})$ and $H_t \equiv (0, \sigma)$ for $t = 1, \cdots, T$. Here we used the initial condition $\theta_0 = 0$ and $\theta_1 \sim N(0, \frac{\sigma^2}{1-\gamma^2})$. In the simulation smoother draws are from $\eta = (\eta_0, \cdots, \eta_n)'$ for an appropriately chosen F_t. In our setup we use $F_t = H_t$. Further, since we want to update the state variables and the regression parameters jointly we apply the extensions discussed in Section 5 of de Jong and Shephard (1995). Using this approach and simplifying the simulation smoother for (2.2)-(2.3) the Kalman filter recursions for $t = 1, \cdots, T$ are given by

$$\begin{aligned} E_t &= (\mathbf{x}_t'\Sigma_0, Z_t + \mathbf{x}_t'\alpha_0) + A_t \\ A_{t+1} &= \gamma A_t - \frac{\gamma P_t}{P_t + 1} E_t \\ Q_{t+1} &= Q_t + \frac{1}{P_t + 1} E_t' E_t \\ P_{t+1} &= \frac{\gamma^2 P_t}{P_t + 1} + \sigma^2 \; , P_1 = \sigma^2. \end{aligned}$$

Calculate S and s from the partition of $Q_{T+1} = \begin{pmatrix} S & -s \\ -s' & * \end{pmatrix}$ and draw $\delta \sim N_p((S+I_p)^{-1}s, (S+I_p)^{-1})$. For the smoothing recursions do for $t = T, T-1, \cdots 1$

$$r_t = 0, \; U_T = 0$$

$$\begin{aligned}
\text{draw } \epsilon_t &\sim N(0, \sigma^2(1-\sigma^2 U_t)) \\
e_t &= E_t \begin{pmatrix} \delta \\ 1 \end{pmatrix} \\
r_{t-1} &= \frac{1}{P_t+1}[\gamma r_t - e_t - \frac{\gamma U_t e_t}{1-\sigma^2 U_t}] \\
U_{t-1} &= \frac{1}{P_t+1}[1 + \frac{\gamma^2 U_t}{(P_t+1)(1-\sigma^2 U_t)}].
\end{aligned}$$

Finally calculate for $t = 0, 1, \cdots, T$

$$\eta_t = \sigma^2 r_t + \epsilon_t$$

and update the state variables for $t = 0, \cdots, T$ by

$$\theta_{t+1} = \gamma\theta_t + \eta_t$$

with $\theta_0 = 0$. This gives a draw from the conditional distribution $[\theta_T^*|\alpha, \mathbf{Z}_T^*, \gamma, \sigma^2]$. To get a draw from the conditional distribution $[\alpha|\mathbf{Z}_T^*, \theta_T^*, \gamma, \sigma] = N_p(\alpha_0 + \Sigma_0(S+I_p)^{-1}\Sigma_0')$ set

$$\alpha = \alpha_0 + \Sigma_0\delta$$

STATE VARIANCE UPDATE:

We will use the conditional $IG(a, b(\gamma)$ prior given in (2.5) for σ^2. A straightforward calculation gives that the density of $[\theta_T|\sigma^2, \gamma]$ is

$$\text{(A.1)} \qquad \pi(\theta_T^*|\sigma^2, \gamma) = \frac{1}{(2\pi\sigma^2)^{\frac{T+1}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(\theta_{t+1} - \gamma\theta_t)^2 - \frac{1-\gamma^2}{2\sigma^2}\alpha_1^2\right\}$$

Condition (A.1) implies that

$$\begin{aligned}
\pi(\sigma^2|\theta_T^*, \gamma) &\propto \pi(\theta_T^*|\sigma^2, \gamma)\pi(\sigma^2) \\
&\propto \frac{1}{(\sigma^2)^{\frac{T+1}{2}+a+1}} \exp\left\{-\frac{1}{\sigma^2}\left[\frac{1}{b(\gamma)} + \frac{1}{2}\left\{\sum_{t=1}^{T}(\theta_{t+1} - \gamma\theta_t)^2 + (1-\gamma^2)\alpha_1^2\right\}\right]\right\},
\end{aligned}$$

which shows that the conditional distribution of $[\sigma^2|\alpha, \gamma]$ is inverse gamma $IG(a^*, b^*)$ with

$$\text{(A.2)} \qquad a^* = \frac{T+1}{2} + a \text{ and } b^* = \left[\frac{1}{b(\gamma)} + \frac{1}{2}\left\{\sum_{t=1}^{T}(\theta_{t+1} - \gamma\theta_t)^2 + (1-\gamma^2)\alpha_1^2\right\}\right]^{-1}.$$

STATE CORRELATION UPDATE:

A causal state process given by (2.3) requires that $\gamma \in (-1, 1)$. So we assume a uniform prior distribution on $(-1, 1)$ for γ. First writing the exponent of $[\gamma|\theta_T^*, \sigma^2]$ and then turning it into a quadratic form, we find that $[\gamma|\theta_T^*, \sigma^2]$ is univariate normal truncated to $(-1, 1)$ with mean μ_γ and variance σ_γ^2 given by, respectively,

$$\text{(A.3)} \qquad \mu_\gamma = \frac{\sum_{t=1}^{T}\theta_t\theta_{t+1}}{\sum_{t=2}^{T}\theta_t^2} \text{ and } \sigma_\gamma^2 = \frac{\sigma^2}{\sum_{t=2}^{T}\theta_t^2}.$$

APPENDIX B: CONDITIONAL DISTRIBUTIONS FOR BINOMIAL DATA

Here we only present the related formulas to update the state variables θ_t and the regression parameter α, because the updates of the other variables are almost identical to those given in Appendix A for the binary case.

To begin, we stack the n_t 1-dimensional latent processes at time t into the following vector process:

$$\mathbf{Z}_t = -\mathbf{1}_{n_t}\mathbf{x}_t' - \mathbf{1}_{n_t}\theta_t + \mathbf{u}_t$$

where $\mathbf{Z}_t = (Z_{1t}, \ldots, Z_{n_t,t})'$ and $\mathbf{1}_{n_t} = (1, \ldots, 1)'$, and $\mathbf{u}_t = (u_{1t}, \ldots, u_{n_t,t})'$. Then in the application of de Jong and Shephard's (1995) simulation smoother, as done in Appendix A, relevant matrices can be readily identified from the above vector process. Some simple algebra leads to a set of terms needed in the Kalman filter recursions: for $t = 1, \ldots, T$,

$$\begin{aligned}
E_t &= (\mathbf{1}_{n_t}\mathbf{x}_t'\Sigma_0, \mathbf{Z}_t + \mathbf{1}_{n_t}\mathbf{x}_t'\alpha_0) + \mathbf{1}_{n_t}A_t \\
A_{t+1} &= \gamma A_t - \frac{\gamma P_t}{n_t P_t + 1}\mathbf{1}_{n_t}'E_t \\
Q_{t+1} &= Q_t + E_t'E_t - \frac{P_t}{n_t P_t + 1}E_t'\mathbf{1}_{n_t}\mathbf{1}_{n_t}'E_t \\
P_{t+1} &= \frac{\gamma^2 P_t}{n_t P_t + 1} + \sigma^2 \ , P_1 = \sigma^2.
\end{aligned}$$

Similar to Appendix A, we acquire matrix S and vector s by partitioning matrix $Q_{T+1} = \begin{pmatrix} S & -s \\ -s' & * \end{pmatrix}$, and then draw $\delta \sim N_p((S + I_p)^{-1}s, (S + I_p)^{-1})$. The smoothing recursions begin with $r_T = 0$, $U_T = 0$, for $t = T, T-1, \ldots, 1$, and then draw $\epsilon_t \sim N(0, \sigma^2(1 - \sigma^2 U_t))$ backwards, with

$$\begin{aligned}
e_t &= E_t \begin{pmatrix} \delta \\ 1 \end{pmatrix} \\
r_{t-1} &= \frac{1}{n_{n_t}P_t + 1}[\gamma r_t - \mathbf{1}_{n_t}'e_t - \frac{\gamma U_t e_t}{1 - \sigma^2 U_t}] \\
U_{t-1} &= \frac{1}{n_t P_t + 1}[n_t + \frac{\gamma^2 U_t}{(n_t P_t + 1)(1 - \sigma^2 U_t)}].
\end{aligned}$$

Finally, we calculate for $t = 0, 1, \cdots, T$

$$\eta_t = \sigma^2 r_t + \epsilon_t,$$

and update the state variables for $t = 0, \cdots, T$ by

$$\theta_{t+1} = \gamma\theta_t + \eta_t,$$

with $\theta_0 = 0$. Thus, this gives a draw from the conditional distribution $[\theta_T^*|\alpha, \mathbf{Z}_T^*, \gamma, \sigma^2]$. Similarly, we get a draw from the conditional distribution $[\alpha|\mathbf{Z}_T^*, \theta_T^*, \gamma, \sigma] = N_p(\alpha_0 + \Sigma_0(S + I_p)^{-1}\Sigma_0')$ by setting

$$\alpha = \alpha_0 + \Sigma_0\delta.$$

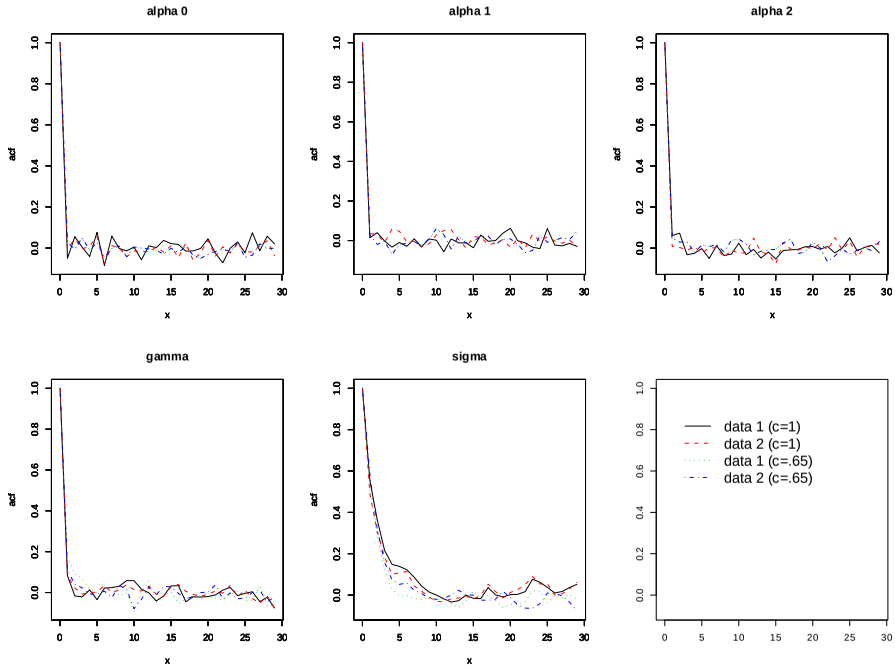

Figure 4.1: Autocorrelations within MCMC iterations of the simulation study.

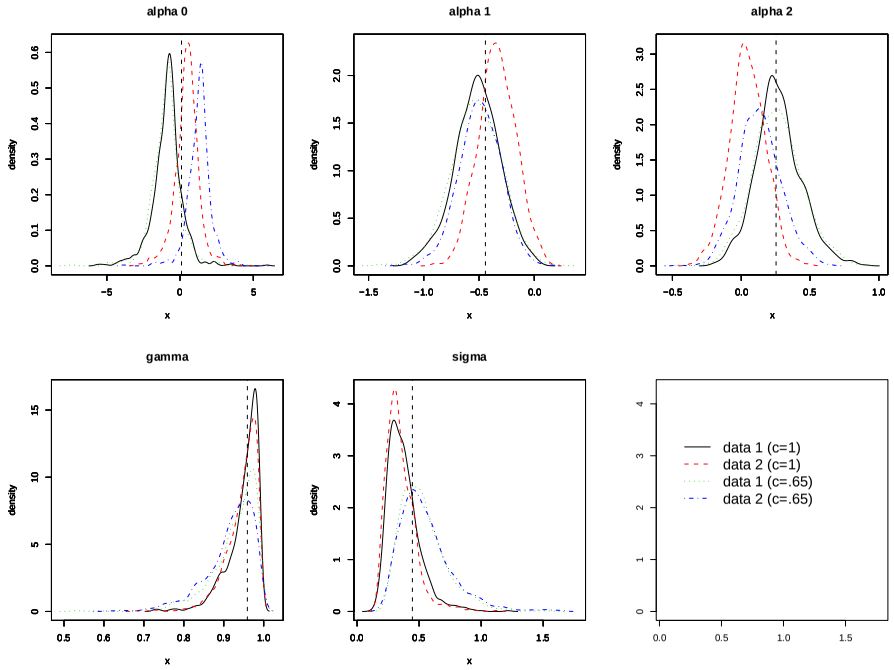

Figure 4.2: Posterior density estimates in the simulation study.

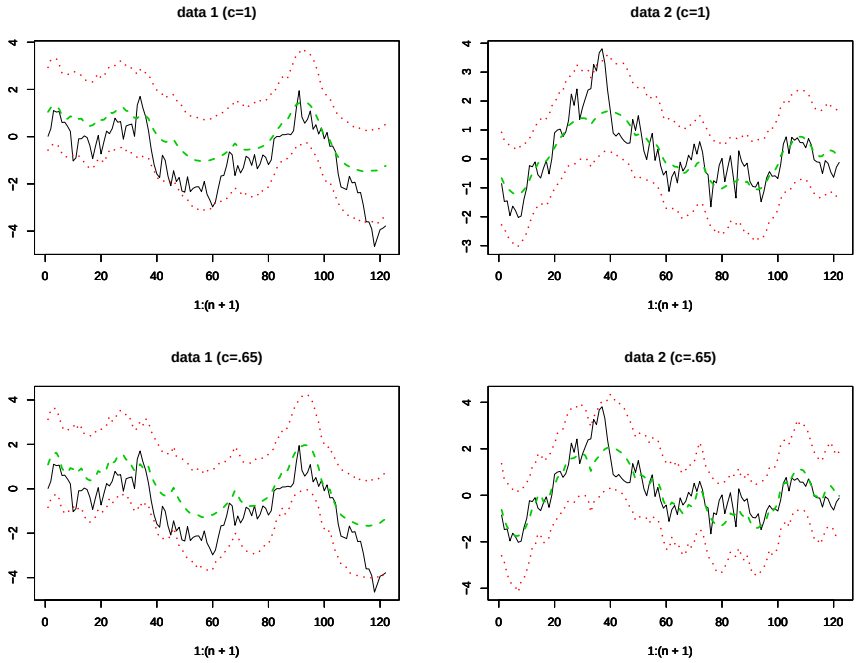

Figure 4.3: Pointwise posterior mean estimates of the state variables θ_t with pointwise 90 % credible intervals of the simulation study.

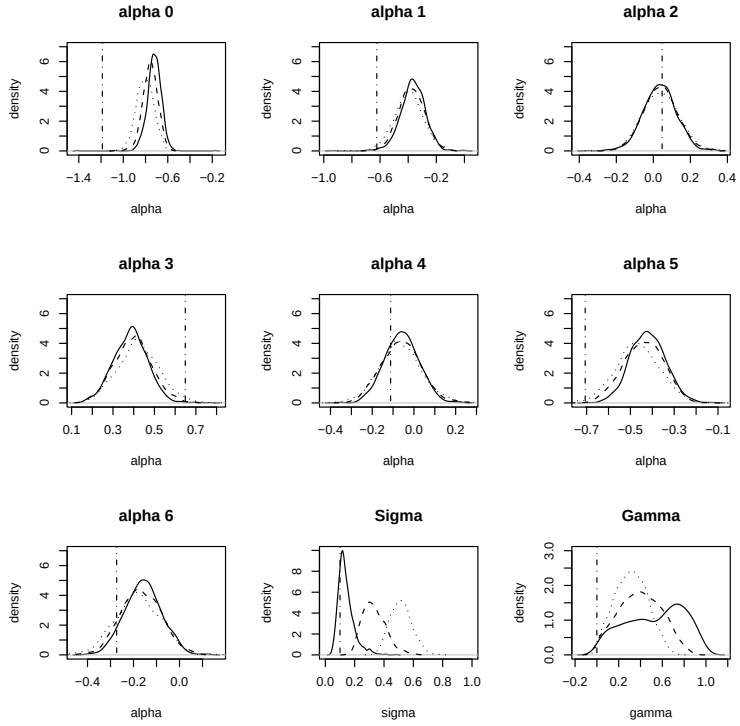

Figure 5.1: Posterior density estimates for the rainfall data with $c = (5, 2, 1)$, corresponding to the solid, the broken and the dotted lines.

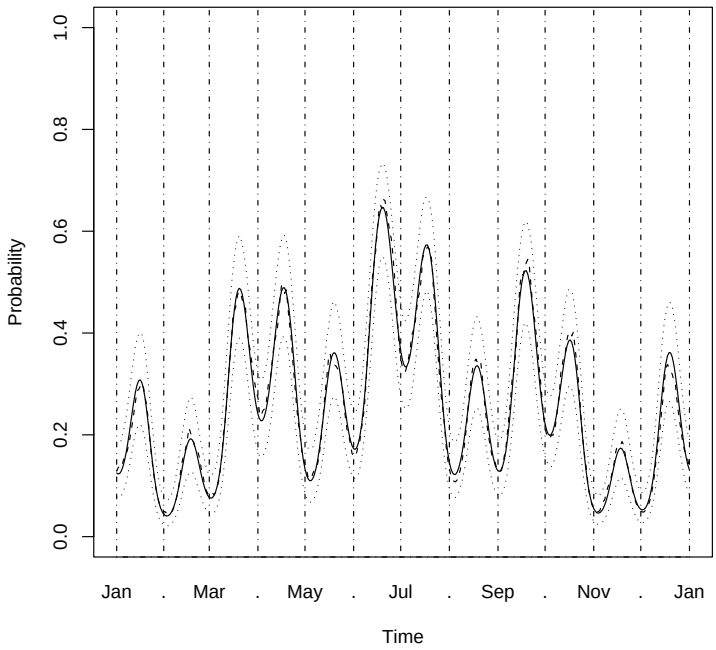

Figure 5.2: Pointwise posterior mean estimates of the probabilities of rainfall with pointwise 90 % credible intervals.

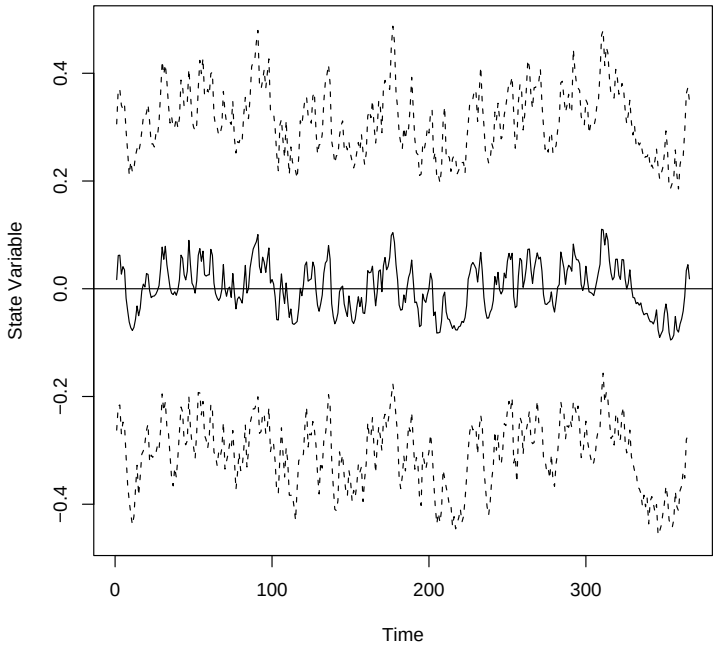

Figure 5.3: Pointwise posterior mean estimates of the state variables θ_t with pointwise 90 % credible intervals.

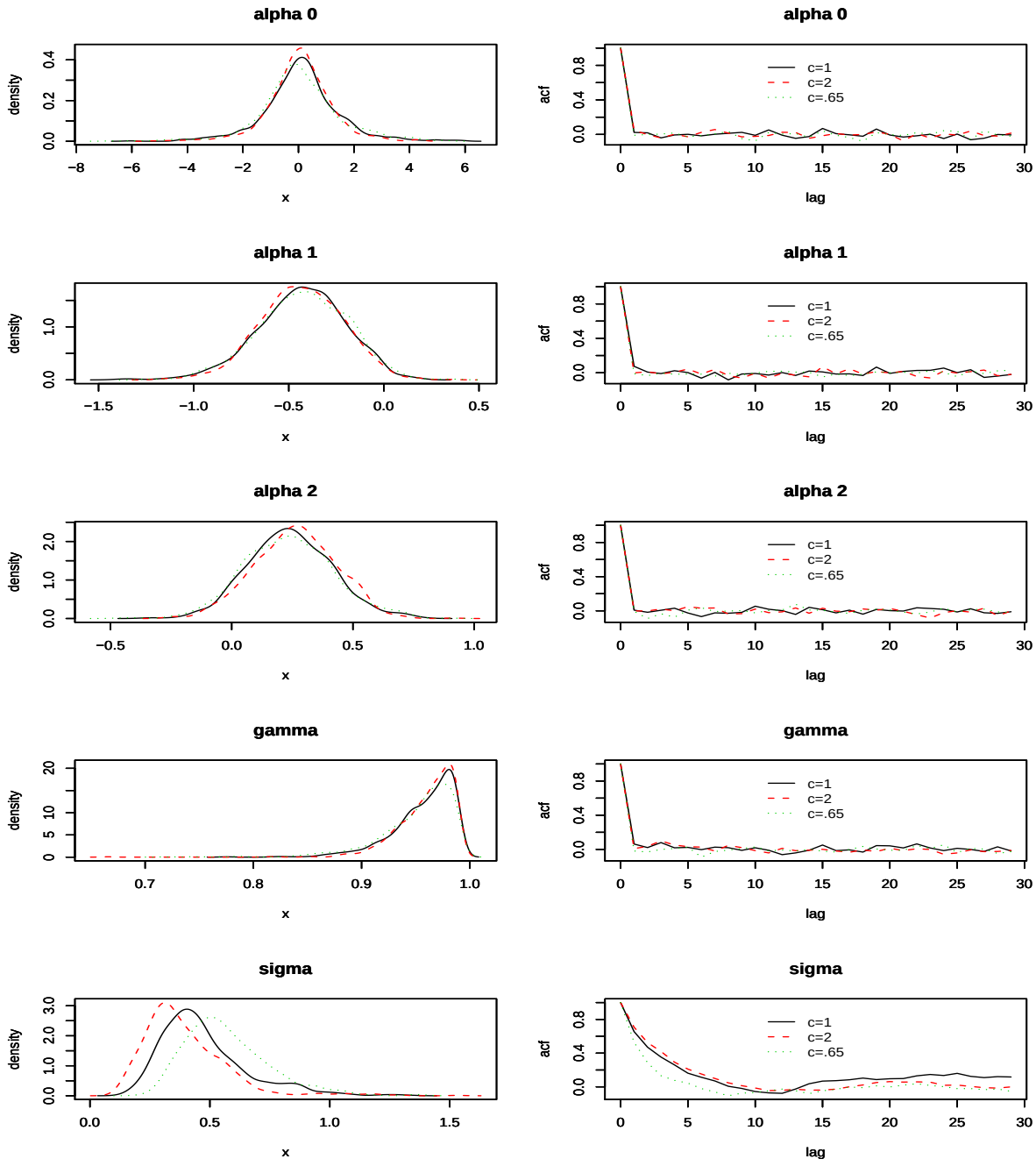

Figure 5.4: Posterior density estimates and estimated autocorrelations among the recorded MCMC iterates for the infant sleep data.

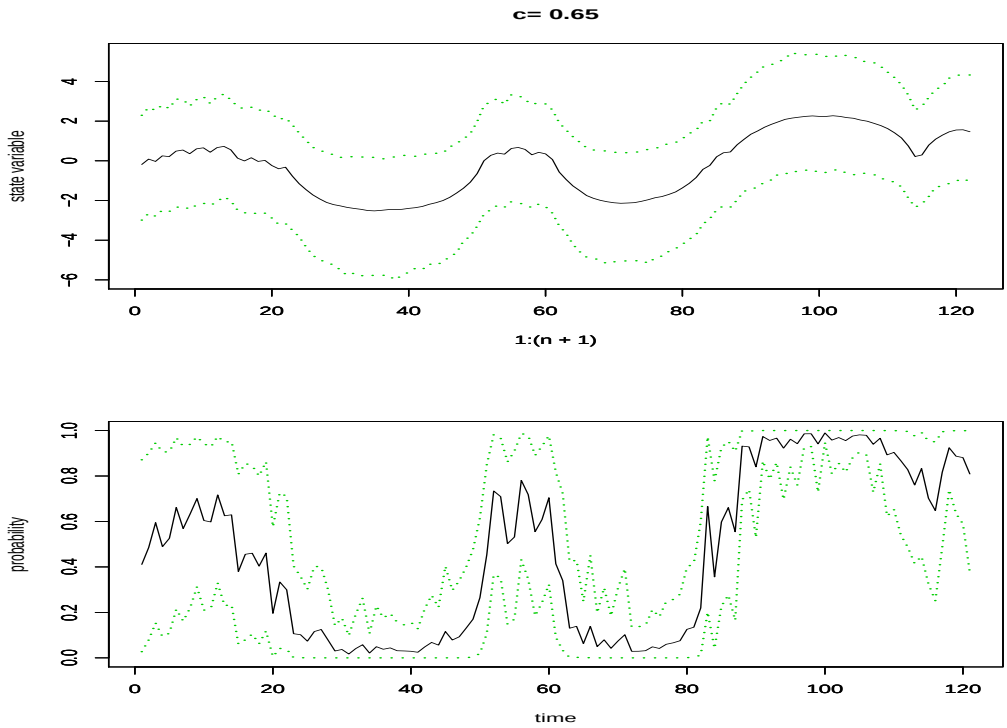

Figure 5.5: Posterior mean estimates of the state variables (top panel) and posterior mean estimates of the REM sleep state probabilities (bottom panel).

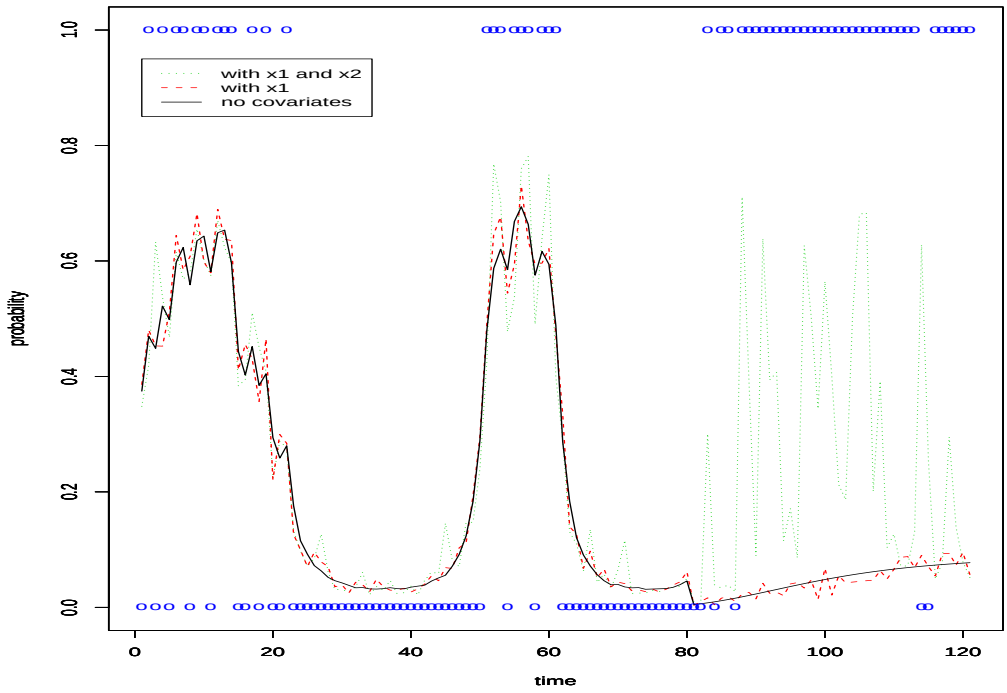

Figure 5.6: Pointwise posterior mean estimates ($t \leq 80$) and predictions ($t > 80$) of REM sleep state probabilities for the infant sleep data.