



INSTITUT FÜR STATISTIK  
SONDERFORSCHUNGSBEREICH 386



Czado, Song:

## State Space Mixed Models for Longitudinal Observations with Binary and Binomial Responses

Sonderforschungsbereich 386, Paper 232 (2001)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



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

Claudia Czado

SCA Zentrum Mathematik  
Technische Universität München  
D-80290 München  
Germany

Peter X.-K. Song

Department of Mathematics and Statistics  
York University  
Toronto, Ontario, M3J 1P3  
Canada

January 19, 2001

## 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 dynamic components. 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 dynamic pattern of data. We develop different Markov chain Monte Carlo algorithms to carry out statistical inference for models with binary and binomial responses. In a simulation experiment we investigate the mixing and convergence properties of these algorithms. In particular, we demonstrate that a joint state variable update is preferable over individual updates. In addition, different prior choices are studied. Finally, we illustrate the applicability of the proposed state space mixed models for longitudinal binomial response data in the analysis of the Tokyo rainfall data (Kitagawa 1987).

**Keywords and phrases:** Binary time series, longitudinal data, Markov chain Monte Carlo, probit, regression, seasonality, state space models.

# 1 INTRODUCTION

Longitudinal discrete data arise from many applied fields, and the development of models for such data has drawn a lot of attention in the literature. In this paper we consider an observed time series of discrete observations,  $\{Y_t, t = 1, \dots, T\}$ , where  $Y_t$  may be either binary or binomial, in the presence of  $p$  time-varying covariates  $X_{t1}, \dots, X_{tp}$ . For expository convenience, our discussion in this section focuses only on the binary case. In the analysis of the binary longitudinal data, 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 of the data.

Amongst available models for time series of binary observations in the literature, 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). 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 given by

$$Y_t | \boldsymbol{\theta}_t \sim \text{Bernoulli}(\mu_t)$$

where the conditional mean or the conditional probability of success  $\mu_t = P(Y_t = 1 | \boldsymbol{\theta}_t)$  follows the observation equation,

$$(1.1) \quad \mu_t = h(G_t' \boldsymbol{\theta}_t)$$

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 some time-varying covariates. We refer to the linear function  $G_t' \boldsymbol{\theta}_t$  with random coefficients  $\boldsymbol{\theta}_t$  to be as the dynamic linear predictor.

In the second process, states  $\{\boldsymbol{\theta}_t\}$  are assumed to follow a  $q$ -dimensional Markov process, governed by the state equation,

$$(1.2) \quad \boldsymbol{\theta}_t = H_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\epsilon}_t,$$

where  $H_t$  is a  $q \times q$ -dimensional transition matrix and the error vector  $\boldsymbol{\epsilon}_t$  has zero mean. When  $q = 1$ , two common state processes used in the literature are the random walk, a nonstationary process with  $H_t \equiv 1$ , and the stationary AR(1) process with  $H_t = \gamma \in (-1, 1)$ , a parameter known as the autocorrelation coefficient.

Generalized from the classical state space models for Gaussian longitudinal data (Jones, 1993), the binary state space models preserves the desirable model structure. That is, the observation equation (1.1) characterizes the dynamic nature of the mean pattern for the observed process while the state equation (1.2) models the serial dependence by a Markov process. Therefore the state space models are known to be both flexible and easy to interpret. Also see the development of state space models for longitudinal counts by, for example, Azzalini (1982), Chan and Ledolter (1995) and Jørgensen et al. (1996).

In this paper we propose a new class of binary state space models that extends the model presentation (1.1) by allowing deterministic linear predictors to enter the observation equation, in order to investigate the effects of trend, seasonal and other given covariates. By a deterministic

linear predictor as opposed to the dynamic linear predictor, we mean a linear function of time-varying covariates with deterministic (or nonrandom) coefficients. This extension is appealing as it opens the door for making statistical inferences such as testing hypothesis for some covariates, which is of scientific interest in many medical or biological studies. Moreover, the inclusion of both deterministic and dynamic components enables us to assess how and to what extent the covariates impact on the mean behavior of the observed process under a full strength of dynamics governed by state space models. Such presence of the two types of linear predictors in the framework of state space models has been considered in other settings, for instance, in the analysis of longitudinal count data by, for example, Zeger (1988), Chan and Ledolter (1995) and Jørgensen et al. (1999).

The difference between the models with and without deterministic predictors can be easily illustrated by the example of Tokyo rainfall data. The 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  $\mu_t$  of rainfall for each calendar day over a year. A state space model, which is currently being used for the data analysis, is as follows,

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

Here the process  $\theta_t$  may be thought of essentially as certain underlying meteorological pattern such as moisture most directly responsible for rainfall. Kitagawa (1987) and Fahrmeir (1992) studied the parameter estimation for  $\theta_t$  with the logistic link function, and Carlin and Polson (1992) developed an MCMC estimation for  $\theta_t$  and Song (2000) proposed Monte Carlo Kalman filter and smoother for  $\theta_t$  with the probit link function. The fitted mean curves  $\hat{\mu}_t = h(\hat{\theta}_t)$  shown in their papers clearly indicate the existence of a seasonal pattern. The nonstationarity due to seasonality appears typical in meteorological settings. In all current analyses mentioned before, such nonstationarity is not modeled, as the state process takes a random walk that behaves as independent increments. Although it is known that a random walk is nonstationary, it is unrealistic to postulate the underlying meteorological process responsible for rainfall follows the nonstationarity of a random walk. The unsuitability can be seen from several aspects of the random walk. For instance, the variance of a random walk increases steadily as time increases, implying that the underlying meteorological system has relatively little variation at the beginning of a year but becomes a lot more volatile at the end. In contrast, we believe that the nonstationarity in the rainfall data would be present more at level of the first moment in trend, seasonal and other forms, with however a bounded second moment. For the seasonality has been noted in the previous analyses, inevitably one faces to a question: Whether or not there is significant evidence in the data that the seasonality is responsible for the nonstationarity. This can not be answered by the existing approaches due to the way of modeling the data. Our approach can provide a solution to the question, by letting the observation equation accommodate seasonal covariates and in the meantime modifying the state process into a stationary AR(1) process. For more details of our analysis of the rainfall data see Section 5.

In general, we assume that the state process  $\boldsymbol{\theta}_t$  has zero mean and that given the  $\boldsymbol{\theta}_t$  the conditional expectation  $\mu_t$  of the observed process is  $\mu_t = h(\boldsymbol{\eta}_t + G_t' \boldsymbol{\theta}_t)$  where the deterministic

predictor  $\eta_t$  takes the form of the classical 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 a linear function of covariates. As far as the statistical inference concerns, we may simply write both trend and seasonal components together to be  $\mathbf{X}_t'\alpha$ , where  $\mathbf{X}_t = (X_{t1}, \dots, X_{tp})'$  and parameter vector  $\alpha$  consists of  $p$  regression coefficients to be estimated.

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

$$(1.3) \quad \mu_t = h(\mathbf{X}_t'\alpha + \theta_t)$$

where the dynamic component takes a simple form of a univariate state variable  $\theta_t$ , which may be regarded as a time-specific effect. With the logistic link function, the  $\theta_t$  describes the random deviation of the log-odds  $\log\{\mu_t/(1-\mu_t)\}$  differing from the expected deterministic pattern  $\mathbf{X}_t'\alpha$  over time. Because of the similarity of model representation (1.3) to generalized linear mixed models (Diggle et al., 1994), we refer the proposed models to be as *state space mixed models*.

The structure of this paper is as follows. In Section 2 we discuss the state space 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 verify the proposed MCMC algorithm. Section 5 presents the details of the analysis of the Tokyo rainfall data using the proposed method. Finally we make some concluding remarks in Section 6.

## 2 BINARY STATE SPACE MIXED MODELS

In this section state space models for binary response variables are investigated. Such models permit a latent variable representation which is utilized in an MCMC algorithm developed for parameter estimation.

### 2.1 MODEL FORMULATION

For a binary longitudinal data  $(Y_t, \mathbf{X}_t), t = 1, \dots, T$ , we adopt the so-called threshold approach (e.g. Albert and Chib, 1993) to model the serial dependence for the binary response vector  $\mathbf{Y}_T^* = (Y_1, \dots, Y_T)'$ . Here we assume that the unobservable latent threshold variable vector  $\mathbf{Z}_T^* = (Z_1, \dots, Z_T)'$  allows for the following linear state space formulation

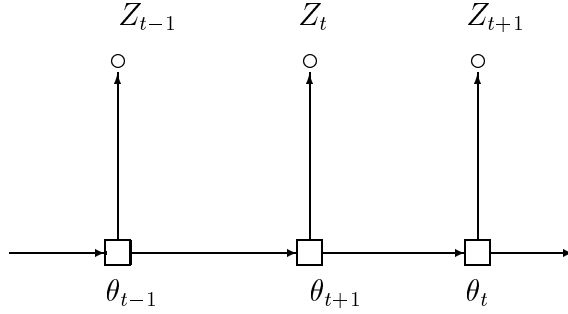
$$(2.1) \quad Z_t = -\mathbf{X}_t'\alpha - \theta_t + u_t, \quad t = 1, \dots, T,$$

$$(2.2) \quad \theta_t = \gamma\theta_{t-1} + \epsilon_t, \quad t = 1, \dots, T,$$

where  $\alpha$  is a  $p$ -dimensional regression parameter and  $\{\theta_t, t = 0, \dots, T\}$  denotes the collection of state variables. It is further assumed that

$$u_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1) \text{ and } \epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_t^2),$$

where  $N(\mu, \sigma^2)$  denotes a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Note that the variance parameters  $\{\sigma_t^2 > 0, t = 1, \dots, T\}$  in the state equation (2.2) are assumed to be unknown, time-varying, and bounded. Therefore the state process governed by (2.2) may accommodate more flexible patterns of variation for the data than a stationary AR(1) process that is a special case of (2.2) with the  $\sigma_t^2$  fixed constant. In addition we require mutual independence between two sets of innovations  $\{u_t, t = 1, \dots, T\}$  and  $\{\epsilon_t, t = 1, \dots, T\}$ . This implies that given  $\theta_t$ ,  $Z_t$  is conditionally independent of the other  $Z_t$ 's and  $\theta_t$ 's, and such a conditional independence may be easily visualized via a graphic representation given by



As initial condition we assume

$$\theta_0 \sim N(0, \sigma_0^2).$$

Finally the latent threshold variables  $Z_t$  are related to the observed binary responses  $Y_t$  through the following latent variable representation:

$$(2.3) \quad Y_t = 1 \iff Z_t \leq 0, t = 1, \dots, T.$$

Representation (2.3) ensures that the marginal distribution of  $Y_t$  given both state variable  $\theta_t$  and covariate vector  $\mathbf{X}_t$  follows a probit model, i.e.,  $\mu_t = P(Y_t = 1 | \theta_t, \mathbf{X}_t) = \Phi(\mathbf{X}_t' \alpha + \theta_t)$  where  $\Phi(\cdot)$  denotes the cumulative distribution function of  $N(0, 1)$ . This shows that a combination of (2.1) and (2.3) is equivalent to (1.3) with  $h(\cdot) = \Phi(\cdot)$ .

We denote the corresponding history vectors by  $\mathbf{Y}_t^* = (Y_1, \dots, Y_t)'$ ,  $\mathbf{Z}_t^* = (Z_1, \dots, Z_t)'$ ,  $\theta_t^* = (\theta_0, \dots, \theta_t)'$  and  $\sigma_t^{2*} = (\sigma_1^2, \dots, \sigma_t^2)'$ .

## 2.2 BAYESIAN INFERENCE USING MONTE CARLO MARKOV CHAIN METHODS

For the Bayesian approach, we assume independent prior distributions for the parameters  $(\alpha, \theta_T^*, \sigma_T^{2*}, \gamma)$ , indicated in a joint density of the form  $\pi(\alpha, \theta_T^*, \sigma_T^{2*}, \gamma) = \pi(\alpha) \times \pi(\theta_T^*) \times \pi(\sigma_T^{2*}) \times \pi(\gamma)$ .

Now we develop MCMC algorithms that allow us to draw samples from the posterior distribution  $[\alpha, \theta_T^*, \sigma_T^{2*}, \gamma, \mathbf{Z}_T^* | \mathbf{Y}_T^*]$ . As usual,  $[u|w]$  denotes the conditional distribution of  $u$  given  $w$ . We follow Tanner and Wong's (1987) Gibbs Sampling approach with data augmentation as what follows shows that the conditional distributions of  $[\mathbf{Z}_T^* | \mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^{2*}, \gamma]$ ,  $[\alpha | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \theta_T^*, \sigma_T^{2*}, \gamma]$ ,  $[\theta_T^* | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \sigma_T^{2*}, \gamma]$ ,  $[\sigma_T^{2*} | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \theta_T^*, \gamma]$  and  $[\gamma | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \theta_T^*, \sigma_T^{2*}, \alpha]$  appear indeed tractable when appropriate prior distributions are chosen. The rest of this section gives the detailed

derivations of these distributions in turn, and readers unfamiliar with MCMC methods can for example consult Casella and George (1992) and Gilks et al. (1996) for an introduction to the Gibbs Sampler.

**LATENT VARIABLE UPDATE:**

Since the latent variables  $Z_t$ 's are conditionally independent given  $\theta_T^*$ , we can immediately reduce the update of  $[\mathbf{Z}_T^* | \mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^{2*}, \gamma]$  to the sequentially individual updates of  $[Z_t | \mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^{2*}, \gamma]$  for  $t = 1, \dots, T$ . Each of these univariate conditional distribution is equivalent to  $[Z_t | \mathbf{Y}_T^*, \alpha, \theta_T^*]$  because given  $\theta_T^*$  the information contained in  $\sigma_T^{2*}$  and  $\gamma$  has no influence on the  $\mathbf{Z}_T^*$ . Moreover we have  $[Z_t | \mathbf{Y}_T^*, \alpha, \theta_T^*] = [Z_t | Y_t, \alpha, \theta_t], t = 1, \dots, T$  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  $Y_t = 0$ ). It is very likely to draw samples in the extreme tails, therefore we suggest to use the probability integral method for the generation of truncated univariate normal random variables in the numerical implementation (e.g. Robert, 1995).

**REGRESSION PARAMETER UPDATE:**

The fact that  $\mathbf{Y}_T^*$  is completely determined with given  $\mathbf{Z}_T^*$  produces the the following reduction,  $[\alpha | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \theta_T^*, \sigma_T^{2*}, \gamma] = [\alpha | \mathbf{Z}_T^*, \theta_{T1}^*]$ , where  $\theta_{T1}^* = (\theta_1, \dots, \theta_T)'$ . Let  $N_p(\mu, \Sigma)$  denote a  $p$ -variate normal distribution with mean vector  $\mu$  and positive definite variance-covariance matrix  $\Sigma$ . Under a multivariate normal  $N_p(\alpha_0, \Sigma_0)$  prior for  $\alpha$ , an application of standard Bayesian techniques (e.g. Lee, 1996, p. 187) immediately results that  $[\alpha | \mathbf{Z}_T^*, \theta_{T1}^*]$  is  $p$ -variate normal with mean vector

$$(2.4) \quad \alpha_m = -(\mathbf{X}'\mathbf{X} + \Sigma_0^{-1})^{-1} \{ \Sigma_m^{-1} \alpha_0 + \mathbf{X}'(\mathbf{Z}_T^* + \theta_{T1}^*) \},$$

and variance-covariance matrix

$$(2.5) \quad \Sigma_m = (\Sigma_0^{-1} + \mathbf{X}'\mathbf{X})^{-1},$$

where  $\mathbf{X} = (X_{ij})_{i=1, \dots, T; j=1, \dots, p}$  is the  $T \times p$  matrix of full column rank. For a flat improper prior of  $\alpha$  (2.4) and (2.5) can be simplified by replacing parameters  $\alpha_0 = 0$  and  $\Sigma_0^{-1} = 0$ .

**STATE VARIABLE UPDATE:**

We propose to jointly update the state variable vector  $\theta_T^*$  in a single step based on the  $(T + 1)$ -variate distribution  $[\theta_T^* | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \sigma_T^{2*}, \gamma]$ . To derive this distribution we first consider the conditional distribution  $[\theta_T^* | \sigma_T^{2*}, \gamma]$ . By the Markovity of the state process  $\theta_T^*$ , we can write

$$(2.6) \quad [\theta_T^* | \sigma_T^{2*}, \gamma] = \left( \prod_{t=1}^T [\theta_t | \theta_{t-1}, \sigma_t^2, \gamma] \right) [\theta_0 | \sigma_0^2].$$

The exponent on the right-hand side of (2.6) takes the form of

$$(2.7) \quad \sum_{t=1}^T \frac{1}{\sigma_t^2} (\theta_t - \gamma \theta_{t-1})^2 + \frac{1}{\sigma_0^2} \theta_0^2,$$

which can be rewritten as a quadratic form  $\theta_T^{*'} \Sigma_{\gamma, \sigma}^{-1} \theta_T^*$  with  $\Sigma_{\gamma, \sigma}^{-1} = P_\gamma D_\sigma P_\gamma'$ , where

$$P_\gamma = \begin{pmatrix} 1 & -\gamma & 0 & \cdots & 0 \\ 0 & 1 & -\gamma & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \text{ and } D_\sigma = \begin{pmatrix} \sigma_0^{-2} & 0 & \cdots & 0 \\ 0 & \sigma_1^{-2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T^{-2} \end{pmatrix}.$$

It follows immediately that  $[\theta_T^* | \sigma_T^{2*}, \gamma]$  is multivariate normal with zero mean vector and variance covariance matrix given by  $\Sigma_{\gamma, \sigma} = P_\gamma^{-1'} D_\sigma^{-1} P_\gamma^{-1}$ .

We now apply Corollary 1 given in the Appendix to determine that the conditional distribution of  $[\theta_T^* | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \sigma_T^{2*}, \gamma]$  is  $(T+1)$ -variate normal. This is because the conditional distribution of  $[\theta_T^* | \sigma_T^{2*}, \gamma]$  is multivariate normal as shown above and the conditional distribution of  $[\mathbf{Z}_T^* | \alpha, \theta_T^*]$  is multivariate normal with mean vector  $-\mathbf{X}\alpha - \theta_{T1}^*$  and variance-covariance matrix equal to the  $T$ -dimensional identity matrix  $I_T$ . The resulting mean vector and variance-covariance matrix are given by, respectively,

$$\Sigma_{\gamma, \sigma} A' (I_T + A \Sigma_{\gamma, \sigma} A')^{-1} (\mathbf{Z}_T^* + \mathbf{X}\alpha), \text{ and } \Sigma_{\gamma, \sigma} - \Sigma_{\gamma, \sigma} A' (I_T + A \Sigma_{\gamma, \sigma} A')^{-1} A \Sigma_{\gamma, \sigma},$$

where

$$A = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}_{T \times (T+1)}.$$

For the implementation of this sampling algorithm, it is useful to note that the elements  $\{s_{ij}, i = 0, \dots, T, j = 0, \dots, T\}$  of the matrix  $\Sigma_{\gamma, \sigma}$  can be computed recursively as

$$s_{00} = \sigma_0^2 \\ \text{for } j \geq i : s_{ij} = \begin{cases} s_{i-1, j} \gamma & \text{if } i \neq j \\ s_{i, j-1} \gamma + \sigma_i^2 & \text{if } i = j. \end{cases}$$

Thus we have  $\Sigma_{\gamma, \sigma} = (s_{ij})_{i=1, \dots, T+1; j=1, \dots, T+1}$ .

To evaluate the effectiveness of this joint update algorithm for the entire vector  $\theta_T^*$ , in comparison to alternatively updating single component  $\theta_t$  sequentially in a simulation experiment in Section 4, we list here relevant formulas required in such an alternative updating scheme. In fact our simulation study uses the AR(1) process for the state variables with  $\sigma_t^2 = \sigma^2$  for all  $t$ . It is straight forward to show that the conditional distribution

$$[\theta_r | \theta_t, t \neq r, \mathbf{Z}_T^*, \gamma, \sigma^2, \sigma_0^2, \alpha] \text{ is } N(\mu_{r*}, \sigma_{r*}^2),$$

where the mean and variance are given by, respectively,

$$\mu_{0*} = \frac{\theta_1 \gamma \sigma_0^2}{\gamma^2 \sigma_0^2 + \sigma^2} \text{ and } \sigma_{0*}^2 = \frac{\sigma_0^2 \sigma^2}{\gamma^2 \sigma_0^2 + \sigma^2}, \text{ for } r = 0;$$



$$\begin{aligned}\mu_{r*} &= \frac{\gamma(\theta_{r-1} + \theta_{r+1}) - \sigma^2(Z_r + \mathbf{X}'_r \alpha)}{\sigma^2 + 1 - \gamma^2} \text{ and } \sigma_{r*}^2 = \frac{\sigma^2}{\sigma^2 + 1 - \gamma^2}, \text{ for } r \neq 0, T; \\ \mu_{T*} &= \frac{\theta_{T-1}\gamma - \sigma^2(Z_T + \mathbf{X}'_T \alpha)}{\sigma^2 + 1} \text{ and } \sigma_{T*}^2 = \frac{\sigma^2}{\sigma^2 + 1}, \text{ for } r = T.\end{aligned}$$

### STATE VARIANCE UPDATE:

For this update we assume an inverse gamma prior for  $[\sigma_t^2]$  with density given by

$$(2.8) \quad \pi(\sigma_t^2) = \frac{1}{b_t^{a_t} \Gamma(a_t) (\sigma_t^2)^{a_t+1}} \exp\left(-\frac{1}{b_t \sigma_t^2}\right)$$

with  $a_t > 0$  and  $b_t > 0$  known hyper parameters, denoted by  $\sigma_t^2 \sim IG(a_t, b_t)$ . A straightforward calculation gives that the density of  $[\theta_t | \theta_{t-1}, \gamma]$  is

$$(2.9) \quad \pi(\theta_t | \theta_{t-1}, \gamma) = \frac{\Gamma(a_t + .5)}{\Gamma(a_t) \sqrt{2\pi} b_t^{a_t}} \left\{ \frac{1}{b_t} + \frac{(\theta_t - \gamma\theta_{t-1})^2}{2} \right\}^{-(a_t+.5)}.$$

Since the prior distributions are independent we can write

$$[\sigma_t^2 | \theta_t, \theta_{t-1}, \gamma] = \frac{[\theta_t | \theta_{t-1}, \sigma_t^2, \gamma][\sigma_t^2]}{[\theta_t | \theta_{t-1}, \gamma]}.$$

Substituting the expressions (2.8) and (2.9) into the last expression yields that the conditional distribution of  $[\sigma_t^2 | \theta_t, \theta_{t-1}, \gamma]$  is inverse gamma  $IG(a_t^*, b_t^*)$  with

$$a_t^* = a_t + .5 \text{ and } b_t^* = \left\{ \frac{1}{b_t} + \frac{(\theta_t - \gamma\theta_{t-1})^2}{2} \right\}^{-1}.$$

In the special case of the stationary AR(1) process with a constant  $\sigma_t^2 = \sigma^2$  for all  $t$ , the assumption of the prior  $\sigma^2 \sim IG(a, b)$  gives rise to the posterior  $[\sigma^2 | \theta_T^*, \gamma]$  being  $IG(a^*, b^*)$  with

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

For a less informative but proper prior choice, hyper parameters  $a$  and  $b$  are chosen in such a way that  $IG(a, b)$  is widely dispersed. When a flat improper prior for  $\sigma^2$  is used, the posterior  $[\sigma^2 | \theta_T^*, \gamma]$  is  $IG(a^*, b^*)$  with

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

In our simulation experiment presented in Section 4, we found that the less informative priors for  $\sigma^2$  tend to generate large values of estimates for  $\sigma^2$  by the MCMC algorithm. Note that unduly large (or small) values of  $\sigma^2$  would lead to a domination of the dynamic component in the model, which could largely reduce the quality of inference on the deterministic component.

Therefore it seems desirable to control values of  $\sigma^2$  in between two reasonable numbers  $l$  and  $u$  if possible. The resulting prior choice for  $\sigma^2$  would apparently be a truncated  $IG(a, b)$  distribution on  $[l, u]$  or more simply a uniform distribution truncated on  $[l, u]$ . In our simulation the latter prior choice is used.

### STATE CORRELATION UPDATE:

A causal state process given by (2.2) requires that  $\gamma \in (-1, 1)$ . So we assume a uniform prior distribution on  $(-1, 1)$  for  $\gamma$ . First writing the exponent of  $[\gamma|\theta_T^*, \sigma_T^{2*}]$  and then turning it into a quadratic form, we find that  $[\gamma|\theta_T^*, \sigma_T^{2*}]$  is univariate normal truncated to  $(-1, 1)$  with mean  $\mu_\gamma$  and variance  $\sigma_\gamma^2$  given by, respectively,

$$\mu_\gamma = \frac{\sum_{t=1}^T \theta_t \theta_{t-1} / \sigma_t^2}{\sum_{t=1}^T \theta_{t-1}^2 / \sigma_t^2} \text{ and } \sigma_\gamma^2 = \frac{\sum_{t=1}^T \theta_{t-1}^2}{\sigma_t^2}.$$

In particular, when (2.2) is stationary with  $\sigma_t^2 = \sigma^2$  for all  $t$ , the above mean and variance expressions can be simplified as follows,

$$\mu_\gamma = \frac{\sum_{t=1}^T \theta_t \theta_{t-1}}{\sum_{t=1}^T \theta_{t-1}^2} \text{ and } \sigma_\gamma^2 = \sigma^2 \left( \sum_{t=1}^T \theta_{t-1}^2 \right)^{-1}.$$

## 3 BINOMIAL STATE SPACE MIXED MODELS

We now consider longitudinal data with binomial response variables. As seen from what follows, the MCMC algorithm developed for the data analysis can easily be established by a simple extension from the previous section. In this section we assume that  $n_t$  Bernoulli trials  $Y_{it}, i = 1, \dots, n_t$  give rise to the binomial response  $Y_t = \sum_{i=1}^{n_t} Y_{it}$ . The latent variable representation (2.3) now specifies the correspondence in a componentwise fashion as follows:

$$(3.1) \quad Y_{it} = 1 \iff Z_{it} \leq 0, t = 1, \dots, T, i = 1, \dots, n_t.$$

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

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

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

$$(3.2) \quad Z_{it} = -\mathbf{X}_i' \alpha - \theta_t + u_{it}, \quad i = 1, \dots, n_t, t = 1, \dots, T,$$

$$(3.3) \quad \theta_t = \gamma \theta_{t-1} + \epsilon_t, \quad t = 1, \dots, T,$$

where

$$u_{it} \stackrel{\text{i.i.d.}}{\sim} N(0, 1) \text{ and } \epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2).$$

Similarly the mutual independence between the sets of innovations,  $\{u_{it}, i = 1, \dots, n_t, t = 1, \dots, T\}$  and  $\{\epsilon_t, t = 1, \dots, T\}$  is imposed.

It is apparent that the only difference between the binary and binomial state space mixed models appears in the dimension on the observed processes. In effect, the binomial case can be regarded as an aggregation of a number of independent binary copies, both driven by the same state process. The immediate implication of this observation is that we need to only modify the updating procedures for  $\theta_t$  and  $\alpha$ , because only these two are directly affected by such dimension expansion.

#### MODIFIED REGRESSION PARAMETER UPDATE:

To modify the regression parameter update, we first define an expanded design matrix  $\overline{\mathbf{X}}$  of  $N \times p$  dimension as  $\overline{\mathbf{X}} = B\mathbf{X}$ , where  $B = \text{blockdiag}(1_{n_1}, 1_{n_2}, \dots, 1_{n_T})$ , a block diagonal matrix with the  $i$ -th block  $1_{n_i}$  being an  $n_i$ -dimensional vector of ones. As before,  $\mathbf{X} = (X_{ij})_{i=1, \dots, T; j=1, \dots, p}$ . Clearly  $B'B$  is a  $T \times T$  diagonal matrix with diagonal elements  $\{n_t, t = 1, \dots, T\}$ . One can now proceed the procedure of updating regression parameter in an exactly same way as in the binary case with the  $\overline{\mathbf{X}}$  in place of the  $\mathbf{X}$ .

#### MODIFIED STATE VARIABLE UPDATE:

The state variable update for the binomial case requires care, to avoid large matrix inversions and related calculations. First, we proceed as in the binary case to derive that  $[\theta_T^* | \mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \sigma_T^{2*}, \gamma]$  is  $(T + 1)$ -dimensional normal with mean vector

$$(3.4) \quad \mu_{post} = \Sigma_{\gamma, \sigma} A' B' (I_N + BA \Sigma_{\gamma, \sigma} A' B')^{-1} (\mathbf{Z}_T^* + \overline{\mathbf{X}} \alpha),$$

and variance-covariance matrix

$$(3.5) \quad \Sigma_{post} = \Sigma_{\gamma, \sigma} - \Sigma_{\gamma, \sigma} A' B' (I_N + AB \Sigma_{\gamma, \sigma} A' B')^{-1} BA \Sigma_{\gamma, \sigma}.$$

Set  $A_s = A \Sigma_{\gamma, \sigma} A'$ . It follows from Sherman-Morrison-Woodberry formula (see Ortega and Rheinboldt, 1970, p.50) that

$$(3.6) \quad \begin{aligned} (I_N + BA_s B')^{-1} &= I_N - B(I_T + A_s B' B)^{-1} A_s B' \\ &= I_N - B(B' B)^{-1} \{(B' B)^{-1} + A_s\}^{-1} A_s B'. \end{aligned}$$

Using (3.6) we can simplify (3.4) and (3.5), respectively, to

$$\mu_{post} = \Sigma_{\gamma, \sigma} A' \left[ I_T - \{(B' B)^{-1} + A_s\}^{-1} \right] B' (\mathbf{Z}_T^* + B\mathbf{X}\alpha)$$

and

$$\Sigma_{post} = \Sigma_{\gamma, \sigma} - \Sigma_{\gamma, \sigma} A' \left[ I_T - \{(B' B)^{-1} + A_s\}^{-1} \right] B' B A \Sigma_{\gamma, \sigma}.$$

Finally, since both state correlation and state variance updates do not involve the latent variables, the corresponding procedures of updates for the binary case remain valid in the binomial case.

## 4 A SIMULATION EXPERIMENT

To investigate the performance of the different MCMC algorithms and prior specifications for the proposed models we conducted a simulation study based on a binary state space mixed model given as follows:

$$(4.1) \quad Z_t = -\alpha_0 - \alpha_1 x_t - \theta_t + u_t, \quad t = 1, \dots, 200,$$

$$(4.2) \quad \theta_t = \gamma \theta_{t-1} + \epsilon_t, \quad t = 1, \dots, 200,$$

where  $x_t$  is a time trend equally spaced between  $-2$  and  $2$ , and the true values are  $\alpha_0 = 1$ ,  $\alpha_1 = .5$  and  $\gamma = .9$ . The independent innovations are  $u_t \stackrel{\text{iid}}{\sim} N(0, 1)$  and  $\epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$  with the true value  $\sigma = .5$  ( $\sigma^2 = .25$ ). For convenience, we chose as starting condition  $\theta_0 \sim N(0, \sigma^2)$ , so here  $\sigma_0^2 = \sigma^2$ . Figure 4.1 shows a longitudinal data of size  $T = 200$  generated from the above model settings. Note that the deterministic linear predictor is

$$\eta_t = \alpha_0 + \alpha_1 x_t, \quad t = 1, \dots, 200.$$

[Figure 4.1 about here]

The specific purposes of this simulation are (1) to evaluate the effectiveness of the proposed MCMC algorithm between two different procedures of the state variable update, individual  $\theta_t$  updates versus joint  $\theta_t$  update, and (2) to assess how the different prior choices for  $\sigma^2$  influence the length of burn-in and other convergence issues. In particular, two priors for  $\sigma^2$  were chosen in the simulation. They are, the flat improper prior and the uniform prior on interval  $[.1, 1]$ . In addition, throughout the simulation we used the flat improper prior for regression parameters  $\alpha_j, j = 0, 1$ . The four cases considered in the simulation are tabulated as follows,

Case	Prior for $\sigma^2$	Update for $\theta_t$
<b>ind</b>	Flat improper	Individual $\theta_t$ updates
<b>improper</b>	Flat improper	Joint $\theta_t$ update
<b>indtrunc</b>	Uniform $ [.1, 1]$	Individual $\theta_t$ updates
<b>trunc</b>	Uniform $ [.1, 1]$	Joint $\theta_t$ update

The MCMC algorithms were run for 100,000 iterations with every 50th iteration recorded. By examining the time plots we decided to choose a burnin of 50 recorded iterations. The MCMC calculations were implemented in Matlab.

Figure 4.2 shows posterior density estimates of the parameters  $\alpha_0, \alpha_1, \gamma$  and  $\sigma$ , 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. The four estimated posterior modes in most plots are similar, except for the **indtrunc** case where we observe some multimodality for the  $\gamma$  estimation. This indicates that the truncation at 1 has been utilized somewhat differently in the **indtrunc** case. In addition for this case the left tails of the posterior densities for  $\alpha_0$  and  $\alpha_1$  seem to be much heavier than those for the other cases.

[Figure 4.2 about here]

Next we compare the performance of the four algorithms in the aspects of updating the state variables  $\theta_t, t = 1, \dots, 200$  and the latent variables  $Z_t, t = 1, \dots, 200$ . In the simulation study, we know the true underlying values, which are given as the dashed line in Figures 4.3 and 4.4. In both figures the solid lines represent posterior mean estimates. We observe that the posterior mean estimates behave quite similar and the true values are within the 90% credible intervals indicated by two upper and lower dotted lines. Above in Figure 4.2 the **indtrunc** case has been identified to produce a different tail behavior for density estimates of  $\alpha_0$  and  $\alpha_1$ , and now it creates a slightly heavy right tail for  $\{\theta_t\}$  but a reasonably consistent tail pattern for  $\{Z_t\}$ .

[Figures 4.3 and 4.4 about here]

In addition to the graphical checks provided by Figures 4.3 and 4.4 we calculated an estimate of the mean squared error, given by

$$MSE(\theta) = \sum_{t=1}^T (\theta_t^{post} - \theta_t^{true})^2$$

$$MSE(Z) = \sum_{t=1}^T (Z_t^{post} - Z_t^{true})^2,$$

where  $\theta_t^{post}(\theta_t^{true})$  denote the posterior mean estimates (true values). Table 4.1 contains these estimates.

Case	$MSE(\theta)$	$MSE(Z)$
<b>ind</b>	188.94	153.84
<b>improper</b>	157.35	123.12
<b>indtrunc</b>	268.97	112.41
<b>trunc</b>	153.16	115.15

Table 4.1: Estimated Mean Squared Errors for State and Latent Variables

From this we see that the individual  $\theta_t$  updates give higher MSE estimates for the state variables than the corresponding joint updates. For the latent variables individual  $\theta_t$  updates with improper prior for  $\sigma^2$  give the largest estimates, while the other updating schemes perform about equally good. In summary, joint state variable updates are preferable.

Finally we plot the autocorrelation function of the sampled MCMC chains for the four parameters, shown in Figure 4.5. It is clearly seen that the autocorrelation functions for  $\alpha_0$  and  $\alpha_1$  decay significantly more slowly for the procedures with individual state variable updates compared to the procedures with a joint state variable update. This implies that the algorithms using the individual state variable update give much slower rate of mixing and hence take more iterations for convergence. Similar patterns occur in the autocorrelation functions for both parameters of  $\gamma$  and  $\sigma$ .

[Figure 4.5 about here]

In summary, we conclude that the algorithm using the joint state variable update together with the truncated  $\sigma^2$  prior performed superiorly to the other three algorithms. We also notice that the algorithm using joint state variable update but with the flat improper  $\sigma^2$  prior worked reasonably well, which would serve as an alternative to the best when the choice of truncation interval for the informative prior becomes difficult.

## 5 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.

As mentioned in Section 1, most previous analyses assumed a random walk for the state process and ignored the nonstationarity of seasonality, although the seasonal pattern was revealed by these studies.

We now apply the proposed binomial state space mixed model to investigate seasonal and monthly effects with chosen covariates  $\mathbf{X}_t = (\cos 1_t, \sin 1_t, \cos 4_t, \sin 4_t, \cos 12_t, \sin 12_t)'$ , where

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

So the latent variables  $\{Z_t\}$  follow

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

and in contrast to Kitagawa (1987) the state variables  $\{\theta_t\}$  here follow the stationary AR(1) process with a common and bounded variance  $\sigma^2$ . Note again that we take the probit link instead of the logistic link used in Kitagawa (1987) and Fahrmeir and Tutz (1994).

Our experience with the prior choice in the simulation study led us to favor informative prior for  $\sigma^2$ . So we used the uniform prior on the truncation interval  $[.05, 1]$ . A total of 10,000 iterations of this MCMC algorithm adopted to the binomial model were run with every 10th iteration recorded. A burnin of 100 recorded iterations was used for the posterior density estimates.

Figure 5.1 displays the estimated posterior densities for the regression parameters  $\alpha_i, i = 0, \dots, 6$ , the standard error parameter  $\sigma$  and the autocorrelation parameter  $\gamma$ . For comparison, we also fit the rainfall data using the probit regression model with the same deterministic component as in the binomial model but the serial dependence is neglected, and the corresponding point estimates of the regression parameters are indicated by the vertical dotted lines in

these plots. By the 90% credible intervals, it is evident that the effects of the yearly effect  $\cos(1_t)$ , the seasonal covariate  $\cos(4_t)$  and monthly covariate  $\cos(12_t)$  are present. Meanwhile, the estimate of the autocorrelation coefficient  $\gamma$  is around .4, and such a medium sized  $\gamma$  is clearly not in favor of the random walk model (with  $\gamma = 1$ ) for the state variables. In conclusion, the nonstationarity of the rainfall data seems attributed to the seasonality of the meteorological system that can not be modeled by the process of independent increments. We also found the estimated  $\sigma^2$  is around .16.

[Figure 5.1 about here]

In this analysis we also computed the pointwise estimation of the rain probability  $p_t$  at day  $t$ ,  $t = 1, \dots, 366$ . Figure 5.2 shows the posterior mean estimates  $\bar{p}_t$ ,  $t = 1, \dots, 366$  based on the last 900 recorded iterations, represented by the solid line, together with its 90% credible bounds. The green line indicates the posterior mean estimates of the probabilities computed by only using the deterministic component. This line without the disturbance of the state variables  $\theta_t$ 's amplifies the seasonal pattern, which has been fitted to the data. By comparing this green line to the solid line, we may tell how and to what extent the the time-specific random effect  $\theta_t$  affects the deterministic mean pattern.

[Figure 5.2 about here]

Figure 5.3 gives the posterior mean estimates for the state variables  $\theta_t$ ,  $t = 1, \dots, 366$  with 90% pointwise credible bounds. The plot indicates that posterior mean estimates have a zero mean value and a stable variation over time, so the proposed model seems to be fitting the data well.

[Figure 5.3 about here]

Furthermore, we conclude that for the rainfall data a state space model with an AR(1) structure is more appropriate than a random walk structure, since we observe that the 90% credible interval for  $\gamma$  ( $[-.14, .65]$ ) is comfortably away from  $-1$  and  $1$  and that the variation of the state variables  $\theta_t$  remains evenly bounded. In addition, if the AR(1) structure was mistakenly used for the true random walk process, unduly large values for  $\sigma^2$  would be frequently seen in the course of its updating procedure, but this was indeed not the case.

## 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 dynamic components are more flexible to address some substantive issues in data analysis than the state space models with only dynamic components, providing the access of inferential methods used in regression

analysis. They are also useful to detect formally causes for nonstationarity, which is often of scientific interest in applications. In both the simulation study and the analysis of the rainfall data the proposed MCMC algorithm has shown to work satisfactorily with a reasonably fast convergence. We found that the improper priors led to reasonable results in the simulation and may be used when informative priors are not available.

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

The authors are currently looking for the possibility of implementing the proposed algorithm via the BUGS software (<http://www.mrc-bsu.cam.ac.uk/bugs>). With the hope of such success, the programming and computational burden for the application of these new models for data analyst would become less.

Conceptually, perhaps with some analytic efforts, the proposed models can be extended to analyze longitudinal data with polychotomous responses or with ordinal responses. These data types are often encountered in practice, and the development of models analogous to our binary state space mixed models would be needed. At the moment, we are investigating such extensions.

## ACKNOWLEDGEMENTS

The first author's research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistische Analyse diskreter Strukturen*. The second author's research was supported by the Natural Sciences and Engineering Research Council of Canada.

## REFERENCES

- Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Stat. Assoc.* **88**, 669–679.
- Azzalini, A. (1982). Approximate filtering of parameter driven processes. *J. Time Series Anal.* **3**, 38–44.
- 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.
- 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.
- Diggle, P.J., Liang, K.-Y. and Zeger, S.L. (1994). *The Analysis of Longitudinal Data*. 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 Tutz, G. (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag, Berlin.
- 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. (1996). State-space models for multivariate longitudinal data of mixed types. *Cand. J. Statist.* **24**, 385–402.
- 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.
- Jones, R.H. (1993). *Longitudinal Data with Serial Correlation*. Chapman and Hall, London.
- Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series (with comments). *J. Amer. Stat. Assoc.* **82**, 1032-1063.
- Lee, P.M. (1997). *Bayesian Statistics : An Introduction*. Arnold, London.
- 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.
- Song, P.X.-K. (2000). Monte Carlo Kalman filter and smoothing for multivariate discrete state space models. *Cand. J. Statist.* **28**, 641–652.
- Tanner, M.A. and Wong, W.H. (1987). The calculation of posterior distributions by data augmentation. *J. of Amer. Stat. Assoc.* **82**, 528–540.
- Zeger, S.L. (1988). A regression model for time series of counts. *Biometrika* **75**, 621–629.

## APPENDIX: PROOFS

**Lemma 1** *If the following holds for random vectors  $X \in \mathcal{R}^p$ ,  $Y \in \mathcal{R}^m$  and  $Z \in \mathcal{R}^s$*

$$X|Y, Z \sim N_p(AY + BZ + c, \Sigma_1)$$

$$Y|Z \sim N_m(\mu, \Sigma_2)$$

where  $A$  is a  $p \times m$ ,  $B$  a  $p \times s$ ,  $\Sigma_1$  a  $p \times p$  and  $\Sigma_2$  a  $m \times m$  matrix, while  $c \in \mathcal{R}^p$  and  $\mu \in \mathcal{R}^m$ . If  $\Sigma_1$  is independent of  $Y$  then the joint conditional distribution of  $(Y, X)$  given  $Z$  is as follows:

$$(A.1) \quad \begin{pmatrix} Y \\ X \end{pmatrix} | Z \sim N_{p+m} \left\{ \begin{pmatrix} \mu \\ A\mu + BZ + c \end{pmatrix}, \begin{bmatrix} \Sigma_2 & \Sigma_2 A' \\ A\Sigma_2 & \Sigma_1 + A\Sigma_2 A' \end{bmatrix} \right\}$$

*Proof:* First note that

$$(A.2) \quad \left[ \begin{pmatrix} Y \\ X \end{pmatrix} | Z \right] = [X|Y, Z] \times [Y|Z]$$

The exponential exponent in the density of (A.2) can therefore be written as

$$(A.3) \quad \begin{aligned} & (X - AY - BZ - c)' \Sigma_1^{-1} (X - AY - BZ - c) + (Y - \mu)' \Sigma_2^{-1} (Y - \mu) \\ &= [(X - A\mu - BZ - c) - A(Y - \mu)]' \Sigma_1^{-1} [(X - A\mu - BZ - c) - A(Y - \mu)] \\ &+ (Y - \mu)' \Sigma_2^{-1} (Y - \mu) \\ &= (X - A\mu - BZ - c)' \Sigma_1^{-1} (X - A\mu - BZ - c) + (Y - \mu)' [\Sigma_2^{-1} + A' \Sigma_1^{-1} A] (Y - \mu) \\ &- (Y - \mu)' A' \Sigma_1^{-1} (X - A\mu - BZ - c) - (X - A\mu - BZ - c)' \Sigma_1^{-1} A (Y - \mu) \end{aligned}$$

On the other hand, the exponential exponent in the density of (A.1) is given by

$$(A.4) \quad \{(Y - \mu)', (X - A\mu - BZ - c)'\} \begin{pmatrix} \Sigma_2 & \Sigma_2 A' \\ A\Sigma_2 & \Sigma_1 + A\Sigma_2 A' \end{pmatrix}^{-1} \begin{pmatrix} Y - \mu \\ X - A\mu - BZ - c \end{pmatrix},$$

where the inverse matrix can be written (see Rao, 1973, p.33) as follows

$$\begin{pmatrix} \Sigma_2 & \Sigma_2 A' \\ A\Sigma_2 & \Sigma_1 + A\Sigma_2 A' \end{pmatrix}^{-1} = \begin{pmatrix} \Sigma_2^{-1} + A' \Sigma_1^{-1} A & -A' \Sigma_1^{-1} \\ -\Sigma_1^{-1} A & \Sigma_1^{-1} \end{pmatrix}.$$

It is now straight forward matrix calculation to show that expressions (A.3) and (A.4) are identical. Therefore the distributions of (A.1) and (A.2) are equal.

**Corollary 1** *Under the same conditions as Lemma 1, the conditional distribution of  $Y$  given  $X$  and  $Z$  is*

$$Y|X, Z \sim N_m(\mu_{X,Z}, \Sigma_{X,Z})$$

where

$$\begin{aligned} \mu_{X,Z} &= \mu + \Sigma_2 A' (\Sigma_1 + A\Sigma_2 A')^{-1} (X - A\mu - BZ - c) \\ \Sigma_{X,Z} &= \Sigma_2 - \Sigma_2 A' (\Sigma_1 + A\Sigma_2 A')^{-1} A \Sigma_2 \end{aligned}$$

*Proof:* Follows immediately from Lemma 1 using standard results for conditional distributions of multivariate normal random variables.

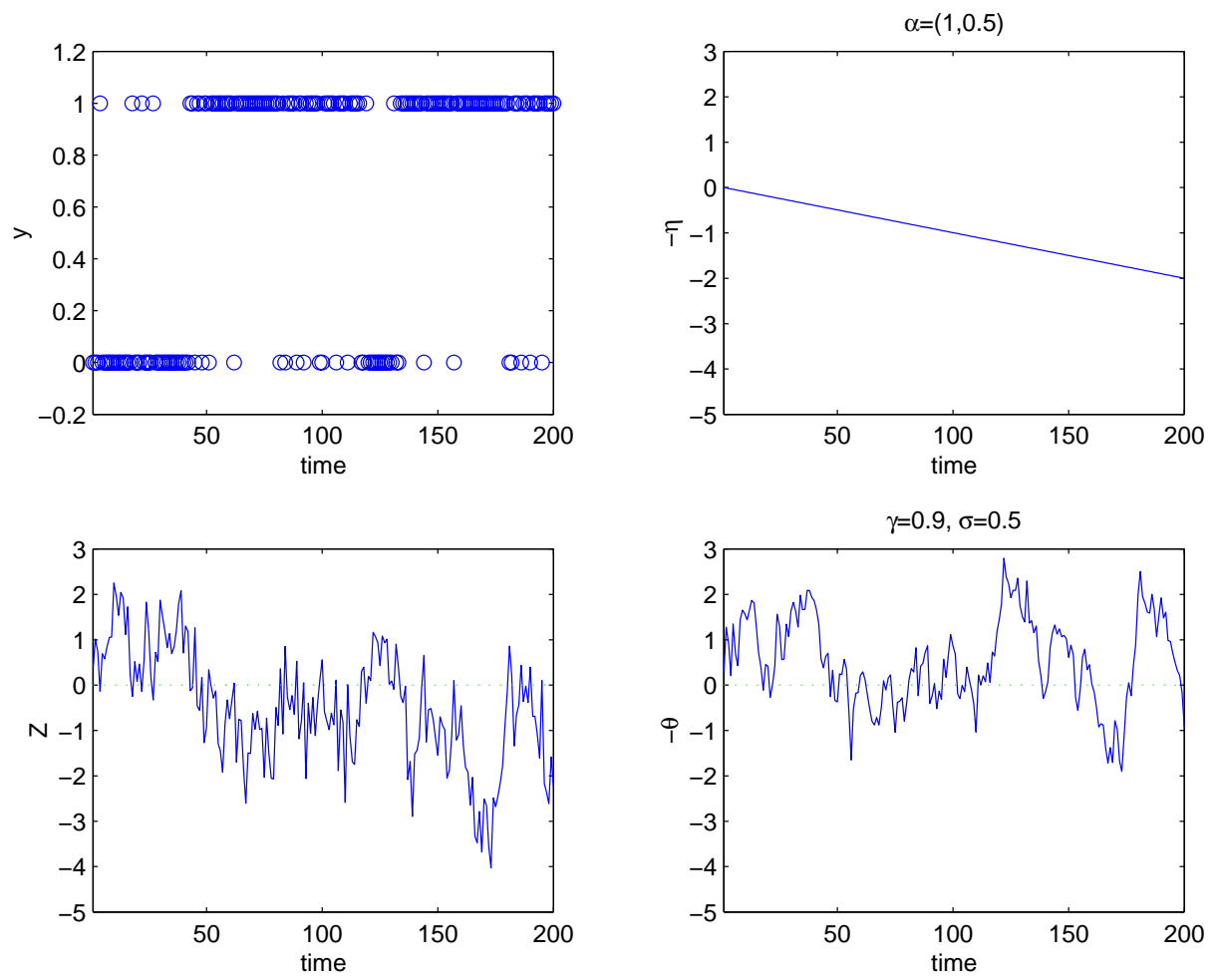


Figure 4.1: Simulated Data from a Regression State Space Model for Binary Response Variables

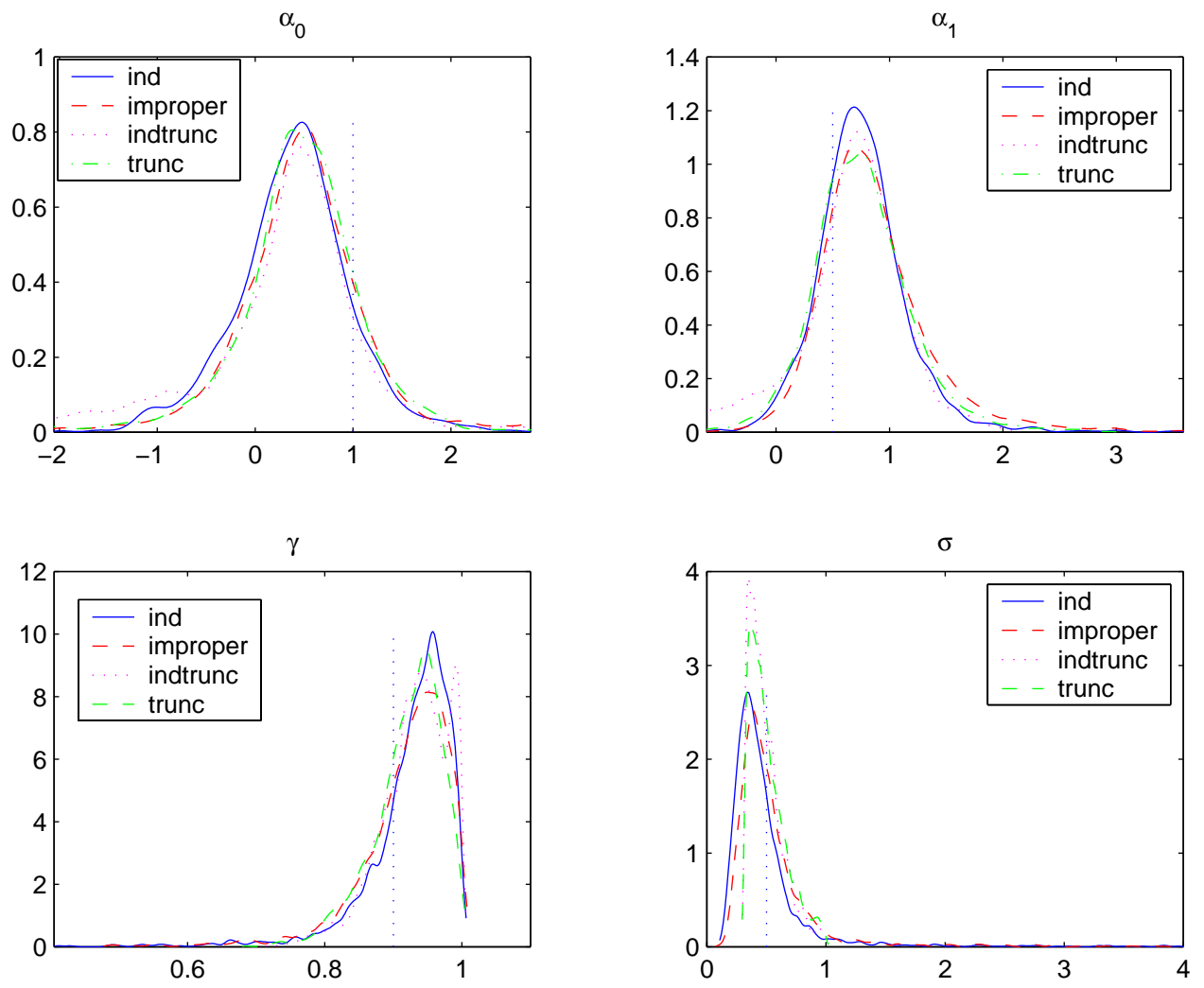


Figure 4.2: Posterior Density Estimates

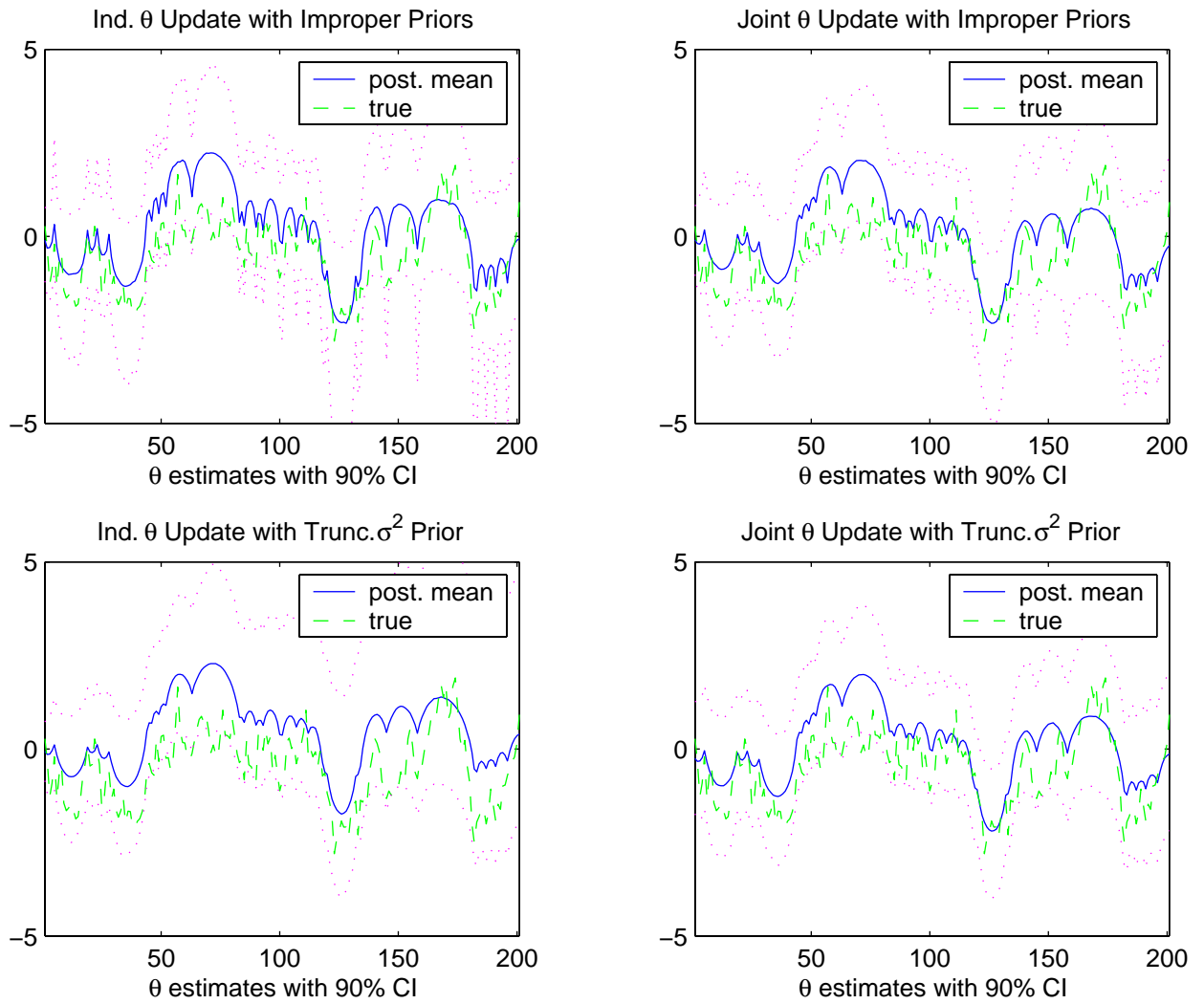


Figure 4.3: Pointwise Posterior Mean Estimates of the State Variables  $\theta_t$  with Pointwise 90 % Credible Intervals

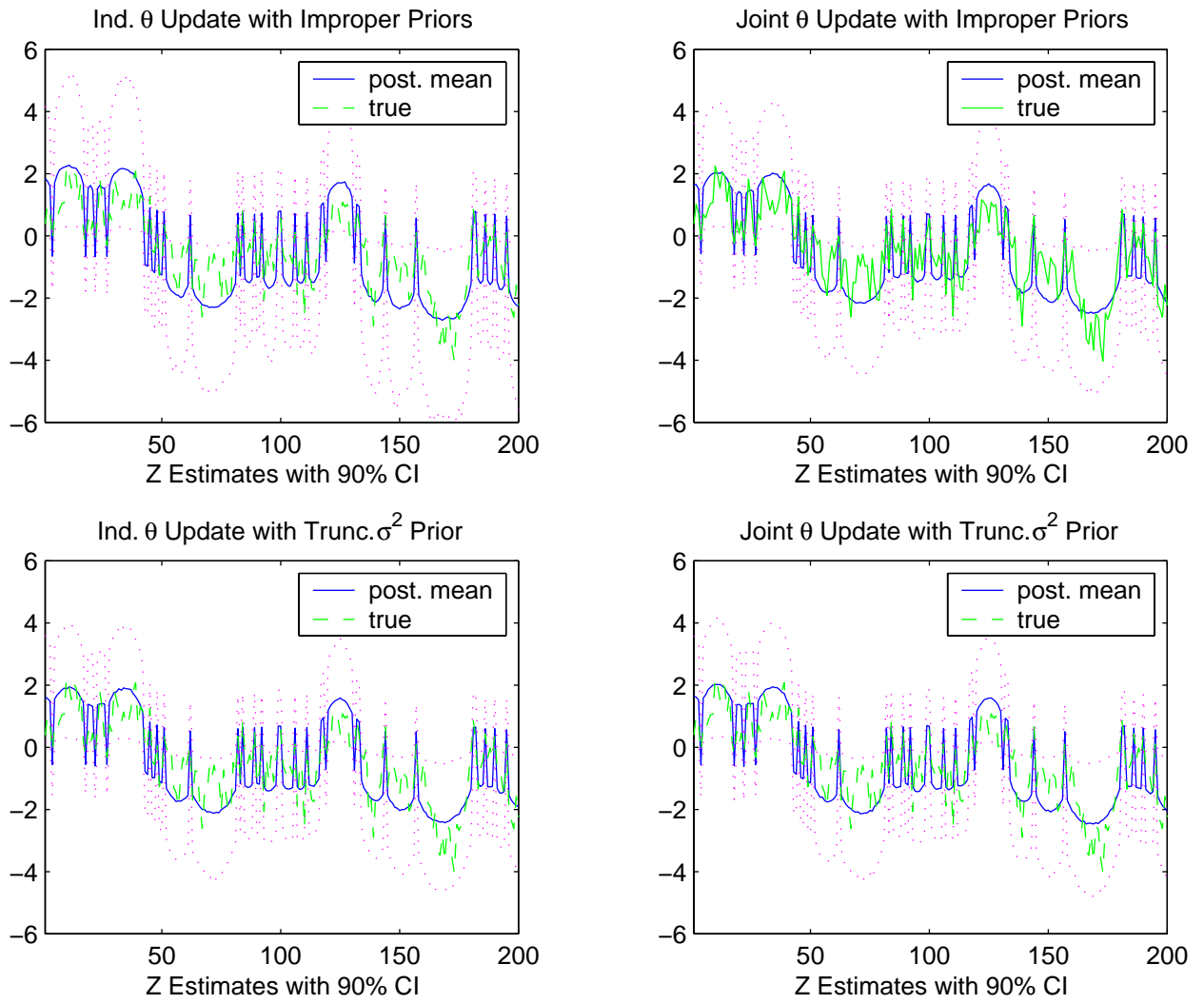


Figure 4.4: Pointwise Posterior Mean Estimates of the Latent Variables  $Z_t$  with Pointwise 90 % Credible Intervals

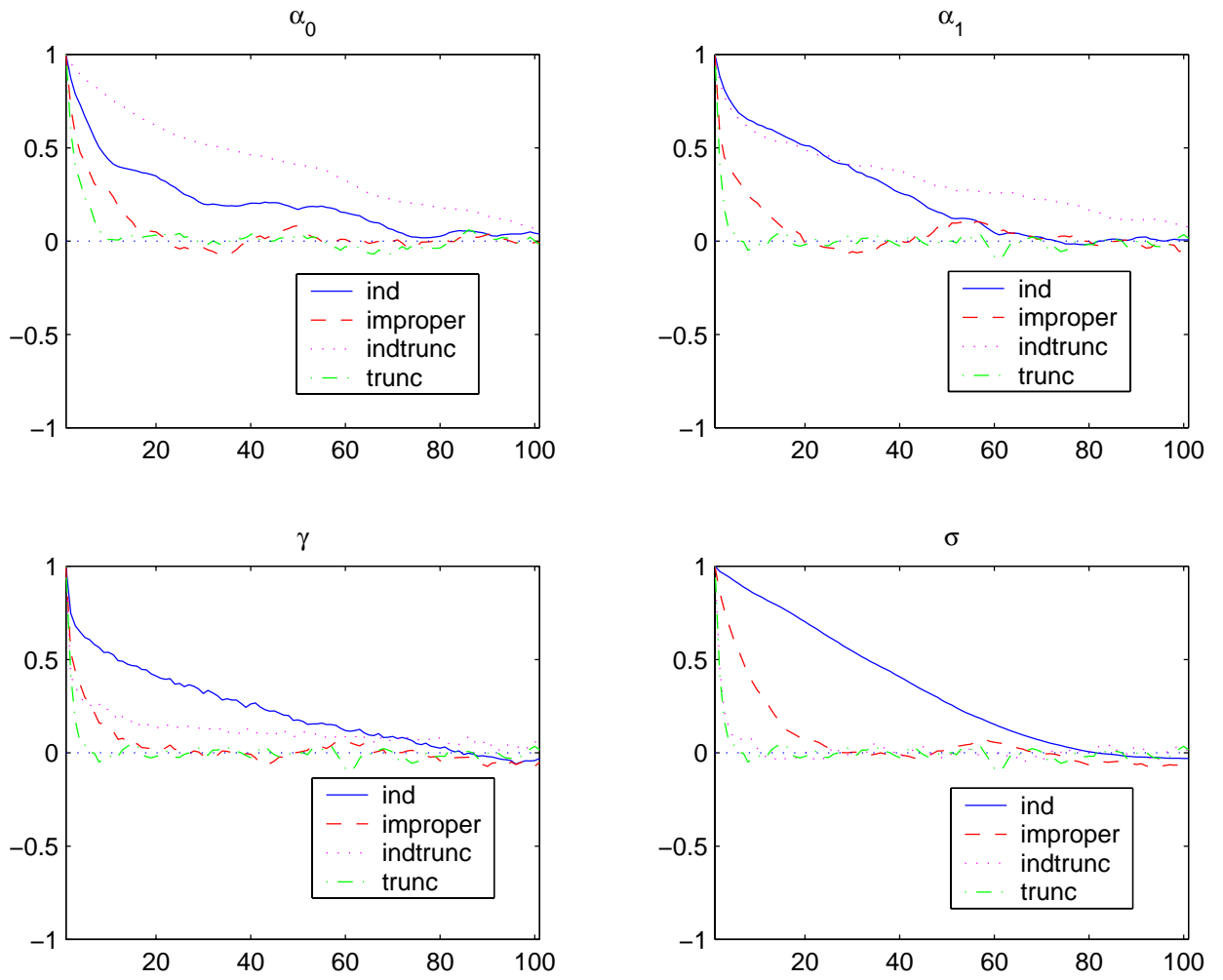


Figure 4.5: Autocorrelations within MCMC Iterations

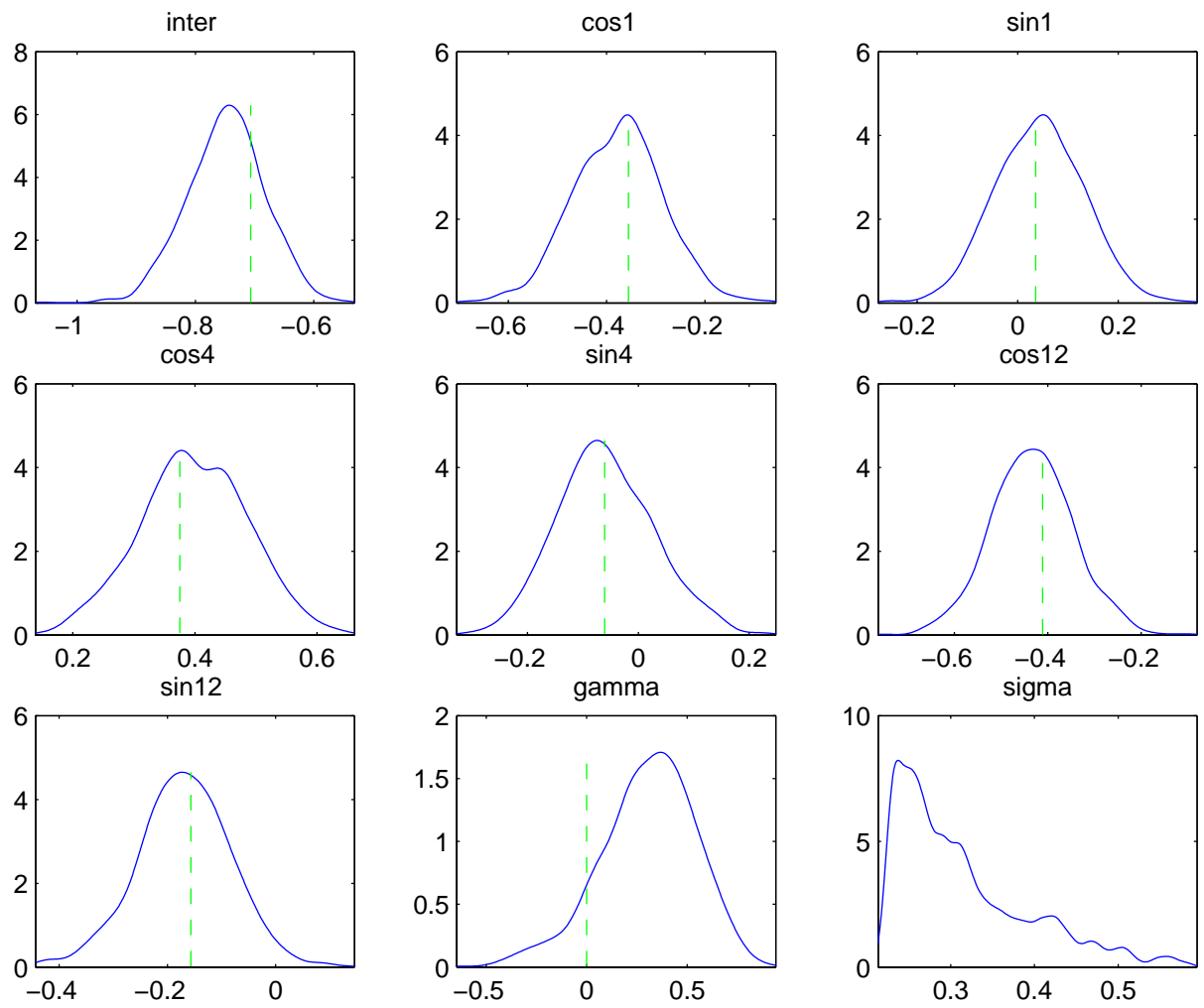


Figure 5.1: Posterior Density Estimates for the Rainfall Data



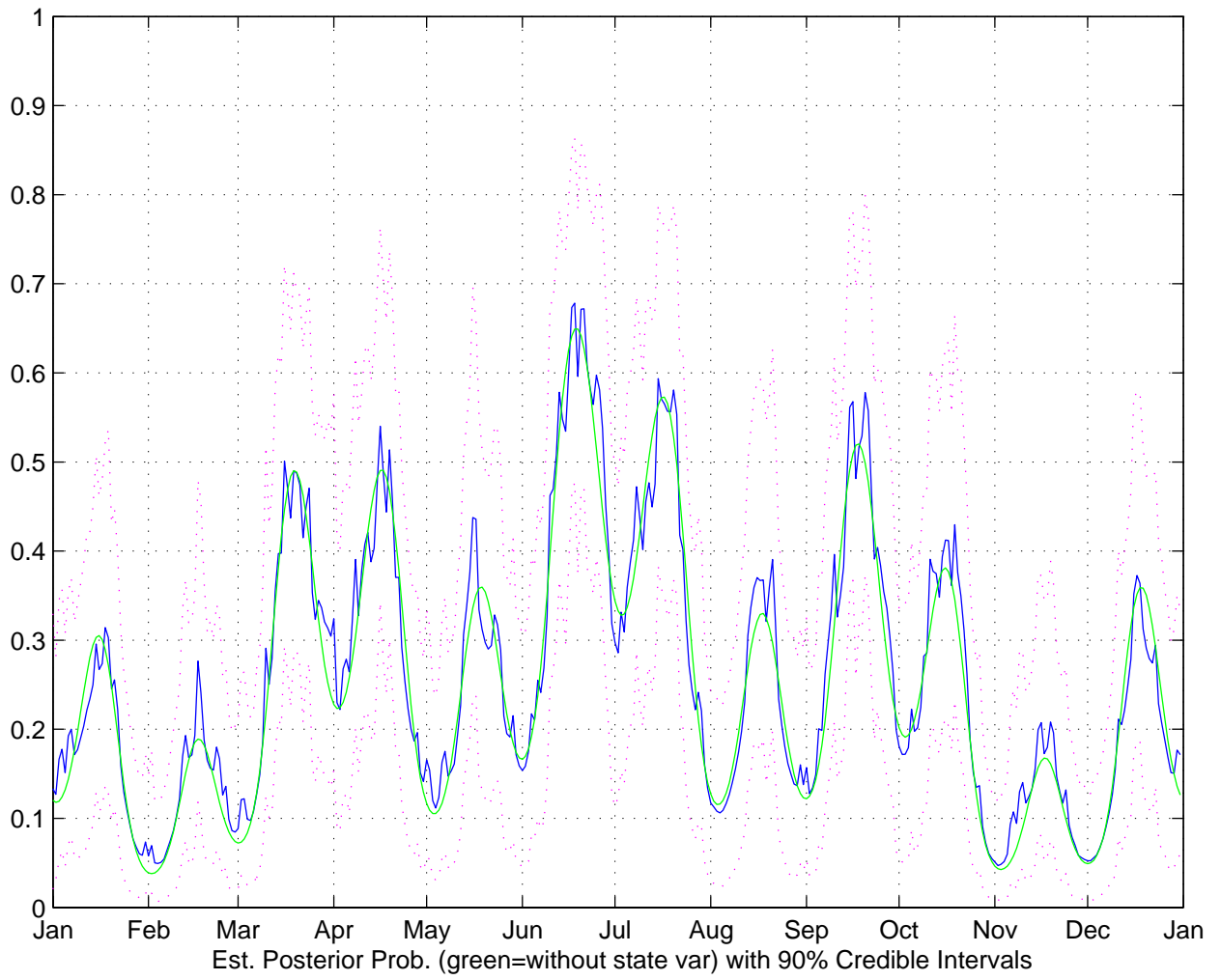


Figure 5.2: Pointwise Posterior Mean Estimates of the Success Probabilities with Pointwise 90 % Credible Intervals

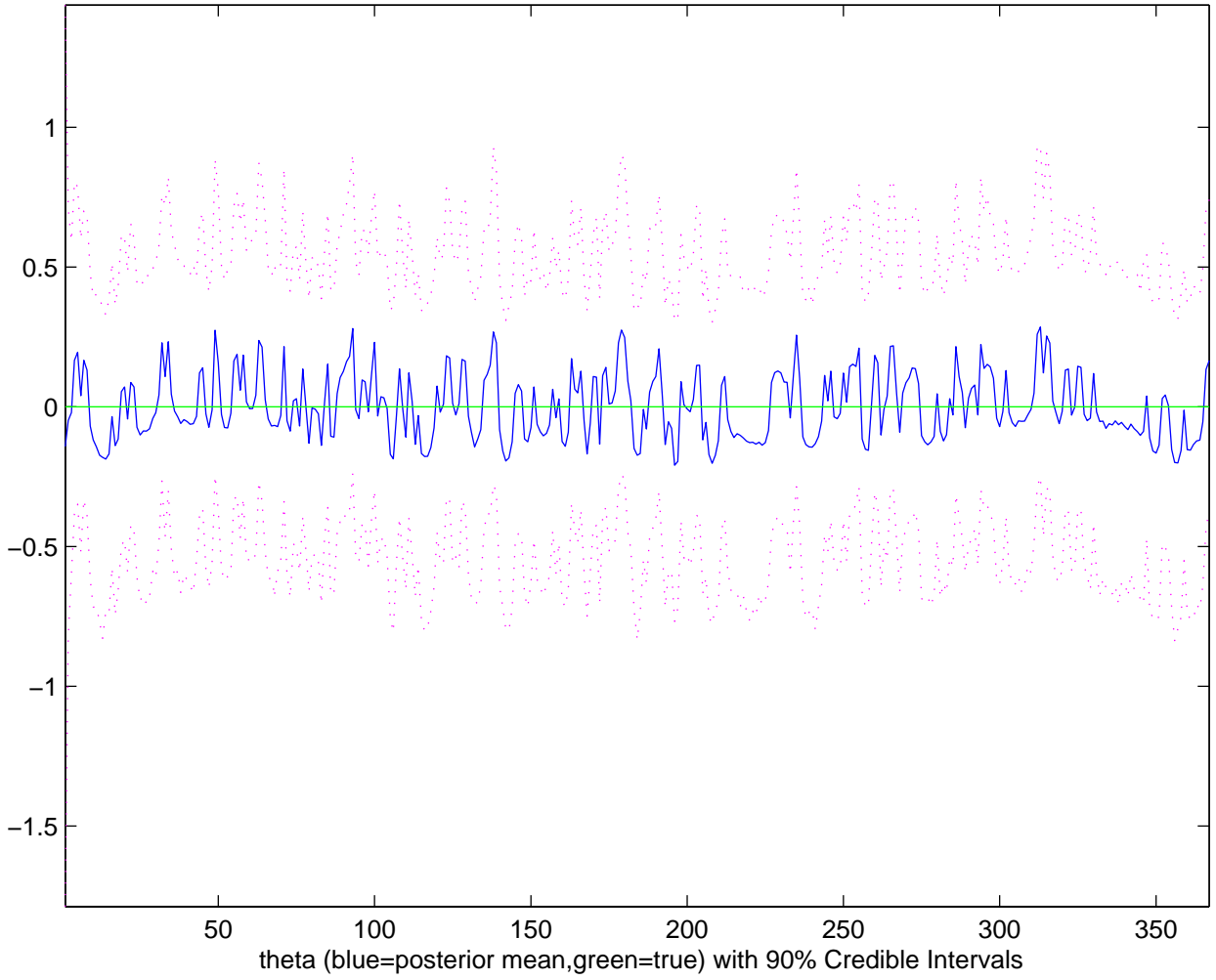


Figure 5.3: Pointwise Posterior Mean Estimates of the State Variables  $\theta_t$  with Pointwise 90 % Credible Intervals