Individual Migraine Risk Management using Binary State Space Mixed Models

By CLAUDIA CZADO

SCA Zentrum Mathematik, Technische Universität München, D-80290 München, Germany
e-mail: cczado@ma.tum.de

Summary

In this paper binary state space mixed models of Czado and Song (2001) are applied to construct individual risk profiles based on a daily dairy of a migraine headache sufferer. These models allow for the modeling of a dynamic structure together with parametric covariate effects. Since the analysis is based on posterior inference using Markov Chain Monte Carlo (MCMC) methods, Bayesian model fit and model selection criteria are adapted to these binary state space mixed models. It is shown how they can be used to select an appropriate model, for which the probability of a headache today given the occurrence or nonoccurrence of a headache yesterday in dependency on weather conditions such as windchill and humidity can be estimated. This can provide the basis for pain management of such patients.

Some key words: Binary time series, longitudinal data, Markov chain Monte Carlo, probit, regression, state space models, model fit and selection.

1. Introduction

About half of all migraine patients believe that weather is a trigger for their headaches (Raskin 1988). Weather conditions such as cold, heat, bright sunshine, changes in pressure, warm dry winds and others have been suggested to increase the probability of a headache. However, many studies investigating these suggestions have been negative or inconclusive. Wilkinson and Woodrow (1979) and Diamond et. al. (1990) found no correlation between headache frequency and adverse weather conditions in London, England and Chicago, U.S.A., respectively. In contrast, Cull (1982) found that a sharp rise in barometric pressure in Scotland reduced the frequency of migraine attacks. On the other hand, the studies of Schulman et. al. (1980) in Boston, U.S.A. and of Nursall (1981) in southern Ontario, Canada, showed no relationship between migraine and pressure. Nursall (1981) however showed that the headache frequency increased as temperature and humidity increased. These studies are commonly based on daily patient diaries reporting on the occurrence or nonoccurrence of a migraine headache. However, more recently the presence of Chinook winds of the Canadian Rockies has been identified as a trigger for migraine headaches (Piorecky et. al. (1996) and Cooke et. al. (2000)) in some patient groups. One possible explanation of these results might be that the influence of weather conditions on migraine headaches varies from individual to individual. Therefore it is of interest to construct patient specific risk profiles from an individual patient dairy, which is the focus of this paper.

As an example of such data, we investigate in this paper the headache dairy of a 52 year old female, who is working part-time in a clerical position. She recorded daily between February

25 until November 30, 1995, if she experienced a headache that day or not. On 98 days out of the 279 recorded days she recorded a headache. She suffers from migraines without aura for 15 years. Since she believes that her headache is triggered by weather conditions, weather related information on a daily basis was also collected. These included information on humidity, windchill, temperature and pressure changes, wind direction, precipitation and cloud cover. Since she is working part-time a cyclical occurrence of migraine attacks can be suspected. The general problem of recursivity is considered by Cugini et. al. (1990). The data is part of a larger study on determinants of migraine headaches collected by the psychologist T. Kostecki-Dillon, York University, Toronto, Canada.

Early analyses ignored the correlation of multiple measurements on the same patient, while Piorecky et. al. (1996) utilized a generalized estimating approach (GEE) introduced by Zeger and Liang (1986) to adjust for this dependency. For the collected long time series, we prefer a model that allows for evolution over time and is likelihood based to investigate the influence of weather conditions on the frequency of migraine attacks. For this task we are looking for a model which can accommodate time dependent covariates such as given by the weather conditions together with a dynamic mechanism which models the dependency between successive days. Such models were introduced and studied by Czado and Song (2001), which use a threshold approach together with a state space approach.

While state space models are first studied for gaussian dynamic systems (e.g. West and Harrison (1989), Jones (1993)), they have become more popular for non gaussian dynamic systems (see for example Fahrmeir (1992), Carlin and Polson (1992) and Song (2000)). While Carlin and Polson (1992) also use a threshold approach, they do not allow for parametric covariate effects. For longitudinal count data state space models with parametric covariate effects have been considered by Zeger (1988), Chan and Ledolter (1995) and Jorgensen et. al (1999).

In this paper we utilize the binary state space mixed models as introduced and studied by Czado and Song (2001) for the headache dairy. We consider several model specifications for this data set and use Markov Chain Monte Carlo (MCMC) methods to facilitate the statistical inference. We also apply some Bayesian model selection criteria such as the Bayesian deviance information criteria by Spiegelhalter et. al (1998) and posterior predictive simulations (see for example Gelman and Meng (1996) and Gelman et. al (1996)) to help to assess model fit and to discriminate between models.

The analysis of the headache dairy of this patient reveals the presence of strong day effects together with weather effects. The presence of severe windchill increases the probability of a headache, while the effect of humidity is less severe. It is also shown that the presence or absence of a headache on the previous day also influences the presence or absence of a headache today. Patient specific risk profiles are constructed, which might help the patient to manage her migraine attacks more precisely.

The paper is organized as follows: Section 2 gives a short review of binary state space mixed models, while Section 3 discusses Bayesian model fit and model selection criteria. Section 4 presents the analysis of the headache data set. Conclusions and discussions are presented in Section 5.

2. Binary State Space Mixed Models

For a binary longitudinal data $(Y_t, \mathbf{X}_t), t = 1, \ldots, T$, Czado and Song (2001) adopted the so-called threshold approach (e.g. Albert and Chib, 1993) to model the serial dependence for the binary response vector $\mathbf{Y}_T^* = (Y_1, \cdots, Y_T)'$. They assume that the unobservable latent threshold variable vector $\mathbf{Z}_T^* = (Z_1, \cdots, Z_T)'$ allows for the following linear state space formulation

$$Z_t = -\mathbf{X}_t'\alpha - \theta_t + u_t, \ t = 1, \cdots, T, \tag{2.1}$$
$$\theta_t = \gamma\theta_{t-1} + \epsilon_t, \ t = 1, \cdots, T, \tag{2.2}$$

where α is a p-dimensional regression parameter and $\{\theta_t, t = 0, \cdots, T\}$ denotes the collection of state variables. It is further assumed that $u_t \overset{\text{i.i.d.}}{\sim} N(0, 1)$ and $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0, \sigma_t^2)$, where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The variance parameters $\{\sigma_t^2 > 0, t = 1, \cdots, T\}$ 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 σ_t^2 fixed constant. In addition, we require mutual independence between two sets of innovations $\{u_t, t = 1, \cdots, T\}$ and $\{\epsilon_t, t = 1, \cdots, T\}$. This implies that given θ_t, Z_t is conditionally independent of the other Z_t's and θ_t's. As initial condition we assume $\theta_0 \sim N(0, \sigma_0^2)\cdot$ Finally the latent threshold variables Z_t are related to the observed binary responses Y_t through the following latent variable representation:

$$Y_t = 1 \Longleftrightarrow Z_t \leq 0, t = 1, \cdots, T\cdot \tag{2.3}$$

Representation (2.3) ensures that the marginal distribution of Y_t given both state variable θ_t and covariate vector $\mathbf{X}_t$ follows a probit model, i.e., $p_t = P(Y_t = 1|\theta_t, \mathbf{X}_t) = \Phi(\mathbf{X}_t'\alpha + \theta_t)$ where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0, 1)$. The corresponding history vectors will be denoted by $\mathbf{Y}_t^* = (Y_1, \cdots, Y_t)'$, $\mathbf{Z}_t^* = (Z_1, \cdots, Z_t)'$, $\theta_t^* = (\theta_0, \cdots, \theta_t)'$ and $\sigma_t^{2*} = (\sigma_1^2, \cdots \sigma_t^2)'$. For the Bayesian approach, independent prior distributions for the parameters $(\alpha, \theta_T^*, \sigma_T^{2*}, \gamma)$, indicated in a joint density of the form $\pi(\alpha, \theta_T^*, \sigma_T^{2*}, \gamma) = \pi(\alpha) \times \pi(\theta_T^*) \times \pi(\sigma_T^{2*}) \times \pi(\gamma)$ are assumed.

Czado and Song (2001) followed Tanner and Wong's (1987) Gibbs Sampling approach with data augmentation. They showed 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 $[\gamma|\mathbf{Y}_T^*, \mathbf{Z}_T^*, \theta_T^*, \sigma_T^{2*}, \alpha]$ are tractable when appropriate prior distributions are chosen. We give now these conditional distributions for the binary response case, derivation and details can be found in Czado and Song (2001). The binomial response case is also considered there.

Latent Variable Update:

The conditional distribution of the latent variables given the remaining parameters is given by

$$[\mathbf{Z}_T^*|\mathbf{Y}_T^*, \alpha, \theta_T^*, \sigma_T^*, \gamma] = \prod_{t=1}^{T}[Z_t|Y_t, \alpha, \theta_t],$$

where $[Z_t|Y_t, \alpha, \theta_t]$ is independent univariate $N(-\mathbf{X}_t^T\alpha - \theta_t, 1)$ distributed truncated to $[-\infty, 0]([0, \infty])$ for $Y_t = 1(Y_t = 0)$

Regressions Parameter Update:

Let $\theta_{T1}^* = (\theta_1, \cdots, \theta_T)^{'}$ and assume a multivariate $N_p(\alpha_p, \Sigma_p)$ prior for α, the conditional distribution of $[\alpha|\mathbf{Y}_T^*\mathbf{Z}_T^*, \theta_{T1}^*]$ is multivariate normal with expectation vector

$$\alpha_m = -(\mathbf{X}^{'}\mathbf{X} + \Sigma_p^{-1})^{-1}(\Sigma_m^{-1}\alpha_p + \mathbf{X}^{'}(\mathbf{Z}_T^* + \theta_{T1}^{'}))$$

and covariance matrix

$$\Sigma_m = (\Sigma_p^{-1} + \mathbf{X}^{'}\mathbf{X})^{-1}\cdot$$

State Variable Update:
Czado und Song (2001) showed that $[\theta_T^*|\mathbf{Y}_T^*, \mathbf{Z}_T^*, \alpha, \sigma_T^{2*}, \gamma]$ is (T+1) dimensional normally distributed with expectation vector

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

and covariance matrix

$$\Sigma_{\gamma,\sigma} - \Sigma_{\gamma,\sigma}A^{'}(I_T + A\Sigma_{\gamma,\sigma}A^{'})^{-1}A\Sigma_{\gamma,\sigma},$$

where $\Sigma_{\gamma,\sigma} = P_\gamma^{-'}D_\sigma^{-1}P_\gamma^{-1}$ with

$$P_\gamma = \begin{pmatrix} 1 & -\gamma & 0 & \cdots & 0 \\ 0 & 1 & -\gamma & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}, \; D_\sigma = \begin{pmatrix} \sigma_0^{-2} & 0 & \cdots & 0 \\ 0 & \sigma_1^{-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_T^{-2} \end{pmatrix}$$

and

$$A = \begin{pmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix} \in \mathcal{R}^{T\times(T+1)}.$$

Alternatively to the joint update of the state variables, individual updates are also possible, however Czado and Song (2001) showed that this leads to poor mixing. Further, $\Sigma_{\gamma,\sigma}$ can be computed recursively.

State Variance Update:
Assuming an inverse gamma prior $IG(a_t, b_t)$ for $[\sigma_t^2]$ given by

$$\pi(\sigma_t^2) = \frac{1}{b_t^{a_t}\Gamma(a_t)(\sigma_t^2)^{a_t+1}}\exp(-\frac{1}{b_t\sigma_t^2}) \text{ with } a_t, b_t > 0,$$

it follows that $[\sigma_t^2|\theta_t, \theta_{t-1}, \gamma] \sim \text{IG}(a_t^*, b_t^*)$ with

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

In the case where prior information about σ_t is sparse, we assume $\sigma_t = \sigma$. It follows that for a flat prior for σ^2, the conditional $[\sigma^2|\theta_T^*, \gamma] \sim \text{IG}(\frac{T}{2} - 1, \left\{\frac{1}{b} + \frac{1}{2}\sum_{t=1}^T(\theta_t - \gamma\theta_{t-1})^2\right\}^{-1})$. In the case of a uniform(l,u) prior for σ^2,$[\sigma^2|\theta_T^*, \gamma]$ is inverse gamma distributed with the same parameters as above but truncated to [l,u]. A truncation might be considered to avoid domination of the dynamic term.

State Correlation Update:
A uniform prior on $[-1, 1]$ is used for γ. This implies that $[\gamma|\theta_T^*, \sigma_T^{2*}]$ is univariate normally distributed with mean μ_γ and variance σ_γ^2 truncated to [-1,1], where

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

3. Bayesian Goodness of Fit and Model Selection

After one has obtained posterior estimates of parameters or quantities of interest through MCMC, one is interested in assessing first the goodness of fit and secondly comparing several models with regard to model fit and model complexity.

We consider the problem of assessing the goodness of fit in a model first. For this we utilize posterior predictive distributions, which were introduced by Guttman (1967) and Rubin (1981, 1984). In contrast to the classical approach these measures can depend on unknown parameters. Meng (1994) uses posterior p-values for testing hypotheses of parameters within a given model, while Gelman et. al. (1996) concentrate on discrepancy measures which are not traditional test statistics. Gelman and Meng (1996) showed how these can be facilitated in an MCMC framework by posterior predictive simulation. For this, one chooses appropriate discrepancy measures, which measure the fit of the model. For our data example, we chose classical measures such as the Pearson χ^2 statistics or the deviance given by

$$D_{\chi^2}(\alpha, \theta^*_{T1}, Y^*_T) = \sum_{t=1}^{T} \frac{(Y_t - p_t)^2}{p_t(1-p_t)} \tag{3.1}$$

$$D_{dev}(\alpha, \theta^*_{T1}, Y^*_T) = -2 \sum_{t=1}^{T} \left[Y_t \log(p_t) + (1 - Y_t) \log(1 - p_t) \right], \tag{3.2}$$

where $p_t = \Phi(\mathbf{X}_t'\alpha + \theta_t)$. Note that given the state variables independence is assumed. To conduct the posterior predictive simulation, we draw regression parameter estimates α^r and state variable estimates $\theta^r_t, t = 0, \cdots, T$ for $r = m+1, \cdots, R$ as an approximate sample from the posterior. Here, m is an appropriately chosen burnin. Given these parameter estimates α^r and $\theta^r_t, t = 0, \cdots, T$, we generate hypothetical data Y^{*r}_T from the corresponding binary state space mixed model. If the model fits adequately, we expect the distributions of Y^{*r}_T and the observed data Y^*_T to be similar. In particular we expect the corresponding discrepancy measures (3.1) and (3.2) to be close. Therefore we determine the estimated posterior predictive p-values

$$p_{\chi^2} = \frac{1}{R-m} \sum_{r=m+1}^{R} I_{\{D_{\chi^2}(\alpha, \theta^*_{T1}, Y^*_T) \le D_{\chi^2}(\alpha, \theta^*_{T1}, Y^{*r}_T)\}} \tag{3.3}$$

$$p_{dev} = \frac{1}{R-m} \sum_{r=m+1}^{R} I_{\{D_{dev}(\alpha, \theta^*_{T1}, Y^*_T) \le D_{dev}(\alpha, \theta^*_{T1}, Y^{*r}_T)\}}. \tag{3.4}$$

We expect an estimated posterior predictive p-value of around .5, if the model fits adequately.

We now turn to the problem of model selection. Model comparison in a classical framework usually assumes a measure of model fit together with a measure of model complexity. Since increasing complexity of model results in a better model fit, models are compared by trading these two quantities off. Likelihood ratio tests and Akaike's information criterion (Akaike 1973) are such measures. Spiegelhalter et al. (1998) propose the deviance information criterion (DIC) to use for model selection within the MCMC framework. For this, they use the deviance for measuring the goodness of fit. As in Dempster (1974), Spiegelhalter et. al. consider the posterior distribution of the saturated deviance given by (3.2) in binary state space mixed models. They suggest to use the posterior mean of the saturated deviance as summary of the model fit. For the binary state mixed models this is given by:

$$\overline{D} = E_{\alpha, \theta^*_{T1}, \gamma, \sigma^2 | Y^*_T}(D_{dev}(\alpha, \theta^*_{T1}, Y^*_T)).$$

Table 1. *Potentially important covariates identified by ordinary probit analyses*

WCD	Windchill index when present, 0 otherwise
HDXD	Humidity index when present, 0 otherwise
WC.IND	1 if WCD present, 0 otherwise
S.SE	South-Southeast Wind Indicator
TMND1P	Mean temperature change from previous day
WDAY	Weekday

As measure of model complexity Spiegelhalter et. al (1998) propose to use the effective number of parameters p_D, defined as difference between the posterior mean of the deviance and the deviance evaluated at the posterior expectations of the model parameters. For the binary state space mixed model p_D is given by

$$p_D = E_{\alpha,\theta^*_{T1},\gamma,\sigma^2|Y^*_T}(D_{dev}(\alpha,\theta^*_{T1},Y^*_T)) - D_{dev}(E_{\alpha|Y^*_T}(\alpha), E_{\theta^*_{T1}|Y^*_T}(\theta^*_{T1}), Y^*_T)).$$

Finally, the deviance information criterion is defined as

$$DIC = \overline{D} + p_D.$$

Spiegelhalter et al. showed that DIC can be considered a natural generalization of Akaike's information criterion.

4. Bayesian Analysis of the Headache Data

For the headache data we define

$$Y_t = \begin{cases} 1 & \text{headache on day t} \\ 0 & \text{otherwise} \end{cases}$$

for $t =$ Feb. 25, $\cdots$, Nov. 30, 1995. This gives a total of 279 binary observations. An initial exploratory analysis using ordinary probit models, thus ignoring the dependency identified potentially important covariates given in Table 1. In particular, the effect of pressure and pressure changes was not identified as significant in this explanatory analysis.

For our analysis we assume the following binary state space mixed model

$$\begin{aligned} Y_t &= 1_{\{Z_t \leq 0\}}, t = 1, \cdots, 279 \\ Z_t &= -\eta_t - \theta_t + u_t, u_t \sim N(0,1) \text{ independent} \\ \theta_t &= \gamma\theta_{t-1} + \epsilon_t, \epsilon_t \sim N(0,\sigma^2) \text{ independent} \\ & u_t \text{ and } \epsilon_t \text{ are independent for all t} \end{aligned}$$

together with the following two mean specifications:

Model 1:

$$\begin{aligned} \eta_t = {} & \alpha_0 + \alpha_{WCD} WCD_t + \alpha_{WC\cdot IND} WC{\cdot}IND_t \\ & + \alpha_{HDXD} HDXD_t + \alpha_{TUES} TUES_t + \alpha_{WED} WED_t \\ & + \alpha_{THUR} THUR_t + \alpha_{FRI} FRI_t + \alpha_{SAT} SAT_t + \alpha_{S\cdot SE} S{\cdot}SE_t \end{aligned} \tag{4.1}$$

Model 2:

$$\begin{aligned}\eta_t = {} & \alpha_0 + \alpha_{SAT} SAT_t + \alpha_{SUN} SUN_t + \alpha_{MON} MON_t \\ & + \alpha_{TUES} TUES_t + \alpha_{WED} WED_t + \alpha_{THUR} THUR_t \\ & + \alpha_{TMND1P} TMND1P_t + \alpha_{S \cdot SE} S{\cdot}SE_t \\ & + \alpha_{SAT*TMND1P} SAT_t * TMND1P_t + \alpha_{SUN*TMND1P} SUN_t * TMND1P_t \\ & + \alpha_{MON*TMND1P} MON_t * TMND1P_t + \alpha_{TUES*TMND1P} TUES_t * TMND1P_t \\ & + \alpha_{WED*TMND1P} WED_t * TMND1P_t + \alpha_{THUR*TMND1P} THUR_t * TMND1P_t \end{aligned} \tag{4.2}$$

The deviance in the standard probit analysis of Model 1 (Model 2) results in 326.36 (316.89) with 268 (264) degrees of freedom.

Following the experience gained from comparing different MCMC algorithms in Czado and Song (2001), we utilize the MCMC algorithm with a flat noninformative prior for the regression parameters and a uniform prior on the state variance $\sigma^2 = \sigma_t^2$. As truncation interval for σ^2 we chose $[0{\cdot}1, 1]$ and $[0{\cdot}1, 10]$. 10000 iterations were run, with every 10th iteration recorded. The time sequence plots of the parameters show that a burnin of 50 recorded iterations is sufficient.

Figures 1 and 2 present posterior mean estimates of the regression parameters together with 90% credible intervals for mean specifications (4.1) and (4.2) for different prior choices for σ^2. For comparison we also include probit estimates together with 90% confidence intervals. They show that the data shows evidence that strong weekday effects are present, while the effects of windchill and humidity indices are less pronounced. The presence of south-southeast winds reduces the headache probability. This is consistent with weather pattern for southwestern Ontario indicated in Nursall and Phillip (1980) and Nursall (1981), who reported that these wind directions are often related to nice weather. The effect of temperature changes from the previous day interacts with weekday effects, which might be explained by the part time work schedule of the patient. For both mean specifications we see that weather conditions have an influence on the occurrence or nonoccurrence of headache for this patient.

Since the variances of the latent variables Z_t vary, it might be appropriate to consider scaled regression parameters. In particular, Czado and Song showed that

$$\theta^*_{T1} \sim N_T(0, \Sigma_{1\gamma\sigma}),$$

where $\Sigma_{1\gamma\sigma} = (\Sigma_{\gamma\sigma})_{t,s=1\cdots,T}$. Using the recursion formulas given in Czado and Song (2001), it follows that

$$Var(Z_t) = 1 + \sigma^2 \frac{1-\gamma^{2(t+1)}}{1-\gamma^2} \tag{4.3}$$

$$Cov(Z_t, Z_s) = \sigma^2 \gamma^{t-s} \frac{1-\gamma^{2(s+1)}}{1-\gamma^2}, s \leq t\cdot \tag{4.4}$$

We can approximate $Var(Z_t)$ by $1 + \frac{\sigma^2}{1-\gamma^2}$, therefore it makes sense to consider scaled regression parameters α_s defined by

$$\alpha_s = \frac{\alpha}{\sqrt{1 + \frac{\sigma^2}{1-\gamma^2}}}\cdot$$

Posterior mean estimates of these scaled regression parameters are given in Figures 3 and 4. They show that the credible intervals are larger than the corresponding confidence intervals of the probit estimates, which is to be expected when dependency is not ignored. Further we see

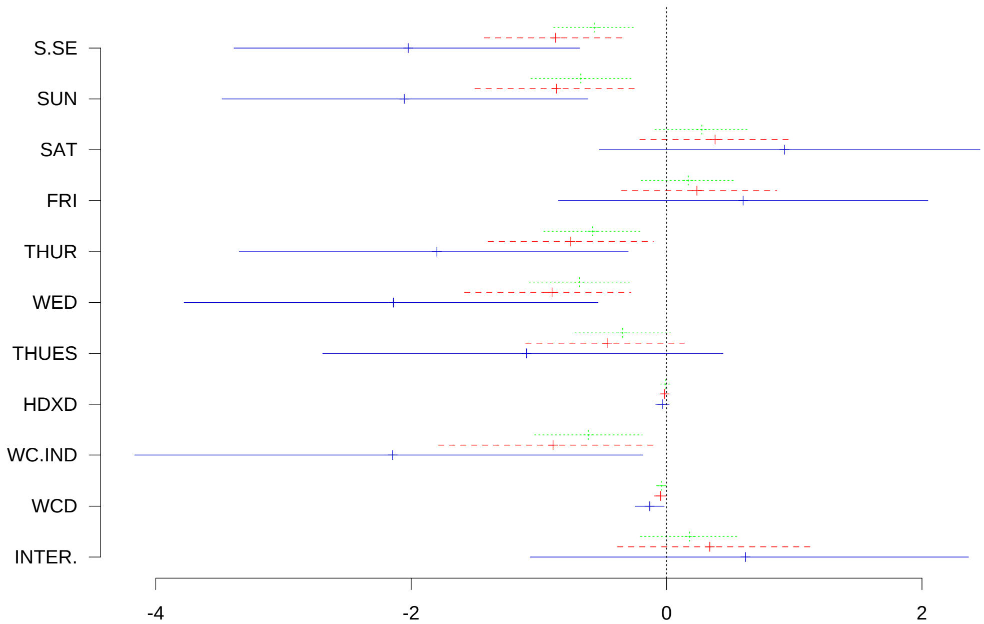

Fig. 1. Posterior mean estimates with 90% credible intervals for the regression parameters in Model 1 (. . . . . = probit estimate with 90% CI, - - - - - = posterior estimate with 90% credible interval using a uniform(.1,1) σ^2 prior, —— = posterior estimate with 90% credible interval using a uniform(.1,10) σ^2 prior)

that all estimates of the scaled regression parameters are close, thus indicating similar effects regardless of the σ^2 prior.

We now investigate the posterior estimates of the parameters determining the dynamic structure of the model. Posterior density estimates of γ and σ are presented in Figures 5 and 6, respectively. The chosen truncation interval influences the posterior estimate of γ. A larger upper bound for σ^2 reduces the correlation. We also see that the MCMC algorithm wants to fit larger state variances.

The results for the state variables θ_t are given in Figure 7. They indicate similar effects as for the state correlation. The range of the state variables increases as the upper truncation bound for σ^2 prior increases.

Finally we investigate the behavior of posterior mean estimates of the success probabilities given by

$$p_t = P(Y_t = 1) = \Phi(\mathbf{X}_t'\alpha + \theta_t)\cdot$$

Figure 8 gives posterior mean estimates of the success probabilities together with the observed successes and failures. From this we see that posterior mean estimates using a uniform(.1,10) σ^2 prior are closer to the observed successes and failures, which indicates a better model fit.

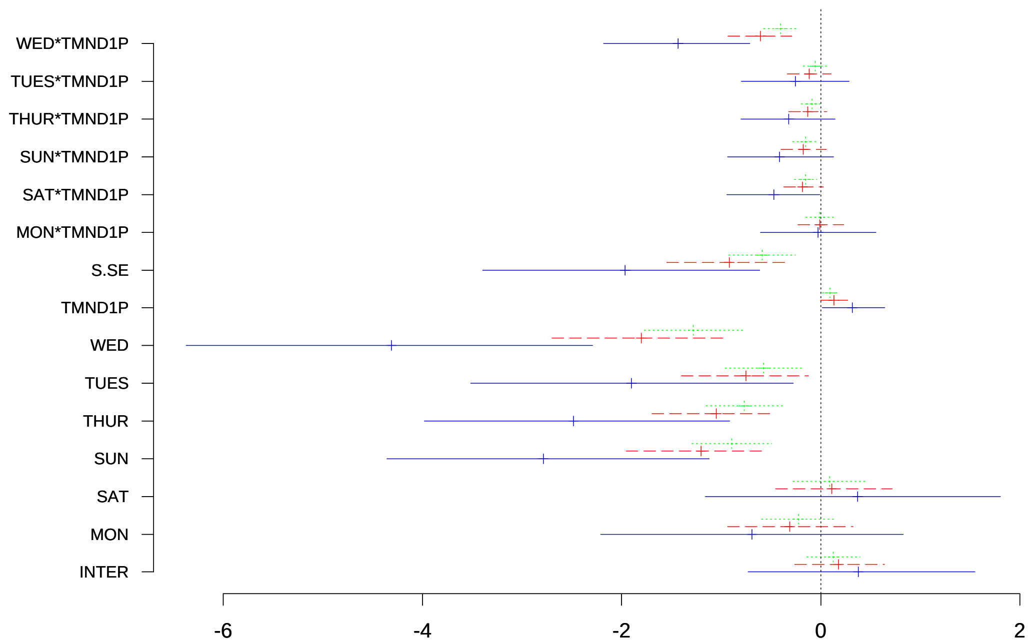

Fig. 2. Posterior mean estimates with 90% credible intervals for the regression parameters in Model 2 (. . . . . = probit estimate with 90% CI, - - - - - = posterior estimate with 90% credible interval using a uniform(.1,1) σ^2 prior, —— = posterior estimate with 90% credible interval using a uniform(.1,10) σ^2 prior)

To see if a higher state variance has influence on the covariate proportion explained by the model, we considered the following quantity:

$$prop_{covariate} = \frac{1}{T} \sum_{t=1}^{T} \frac{|\mathbf{X}_t' \alpha|}{|\mathbf{X}_t' \alpha| + |\theta_t|}.$$

Note that the in the probit model the mean of the latent variable is close to the log odds; for a marginal logit model equality holds. Therefore it makes sense to consider what proportion of the log odds is explained on the average by the parametric part and the dynamic part, respectively. Table 2 gives posterior mean estimates of $prop_{covariate}$, which shows that the dynamic part of the model does not dominate the parametric part, even though there is a moderate decline when a larger upper truncation point for the state variance is used.

We now investigate the model fit and model selection of the two mean specifications and the two different prior choices. First we present the results of the posterior predictive simulation using the last 500 recorded iterations of MCMC chain.

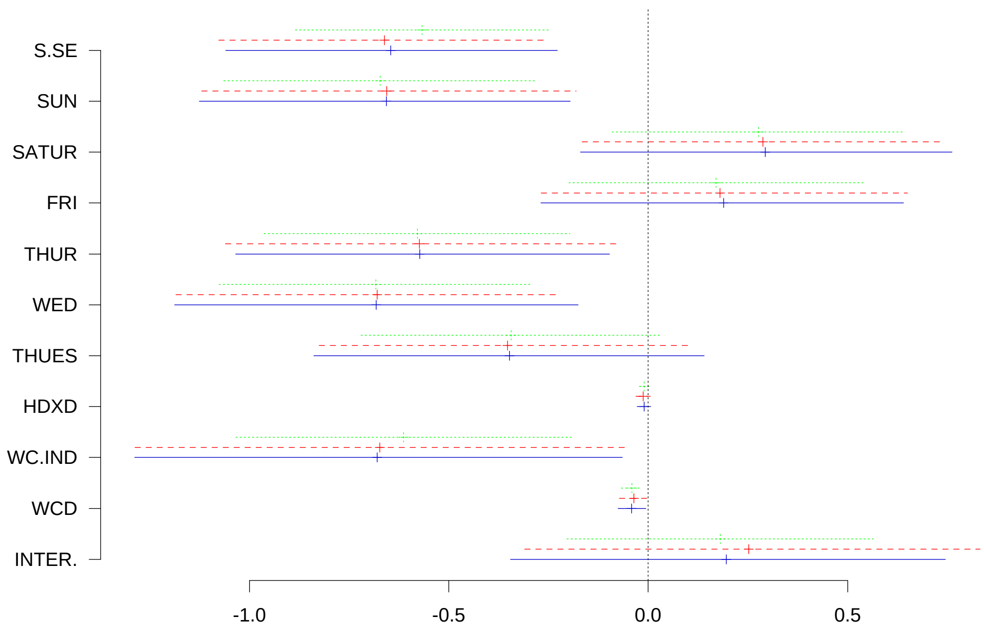

Fig. 3. Posterior mean estimates with 90% credible intervals for the scaled regression parameters in Model 1 (. . . . . = probit estimate with 90% CI, - - - - - = posterior estimate with 90% credible interval using a uniform(.1,1) σ^2 prior, —— = posterior estimate with 90% credible interval using a uniform(.1,10) σ^2 prior)

Table 2. *Posterior Mean Estimates of prop*$_{covariate}$

Model 1		Model 2	
uniform[.1,1] σ^2 prior	uniform[.1,10] σ^2 prior	uniform[.1,1] σ^2 prior	uniform[.1,10] σ^2 prior
.55	.45	.53	.46

Table 3. *Estimated p-values from the posterior predictive simulation*

Model	p_{χ^2}	p_{dev}
Model 1 (uniform(.1,1) σ^2 prior)	.32	.24
Model 1 (uniform(.1,10) σ^2 prior)	.44	.41
Model 2 (uniform(.1,1) σ^2 prior)	.23	.18
Model 2 (uniform(.1,10) σ^2 prior)	.45	.44

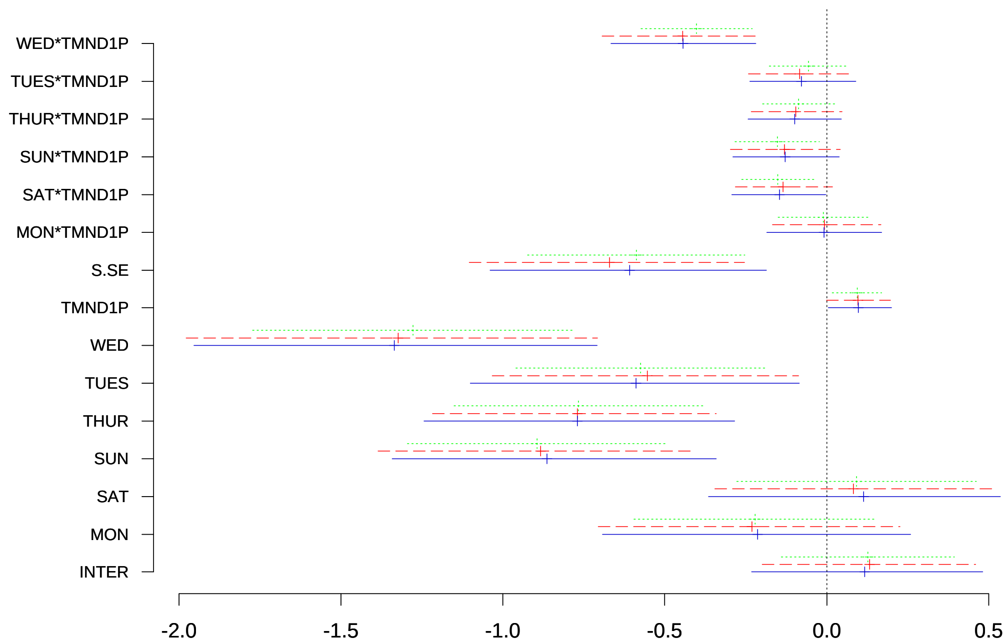

Fig. 4. Posterior mean estimates with 90% credible intervals for the scaled regression parameters in Model 2 (. . . . . . = probit estimate with 90% CI, - - - - - = posterior estimate with 90% credible interval using a uniform(.1,1) σ^2 prior, —— = posterior estimate with 90% credible interval using a uniform(.1,10) σ^2 prior)

The results in Table 3 show that using a larger upper truncation limit for the σ^2 prior yields a better fit. The difference between the two mean specifications is moderate with a very slight preference toward Model 2. The predictive simulations also show that the model fit is adequate when a larger upper truncation limit for the σ^2 prior is used.

We now consider density estimates of the posterior deviance which are given in Figure 9. Here we also can see a lower posterior deviance for the models with a larger upper truncation, while the difference between the two mean specifications is minimal. It should also be noted that the posterior mean deviances are considerably lower than the deviances obtained using ordinary probit, thus indicating an improvement when using binary state mixed models over ordinary probit analyses.

Finally we present in Table 4 the results using the deviance information criterion. DIC would select mean specification Model 2 with a uniform(.1,10) prior for σ^2.

In summary, the model adequacy is achieved when a uniform(.1,10) prior for σ^2 is used, while the difference in the mean specifications is minimal. Therefore we present now predictions using Model 1 specification, which can be used for the pain management of this patient. To ease

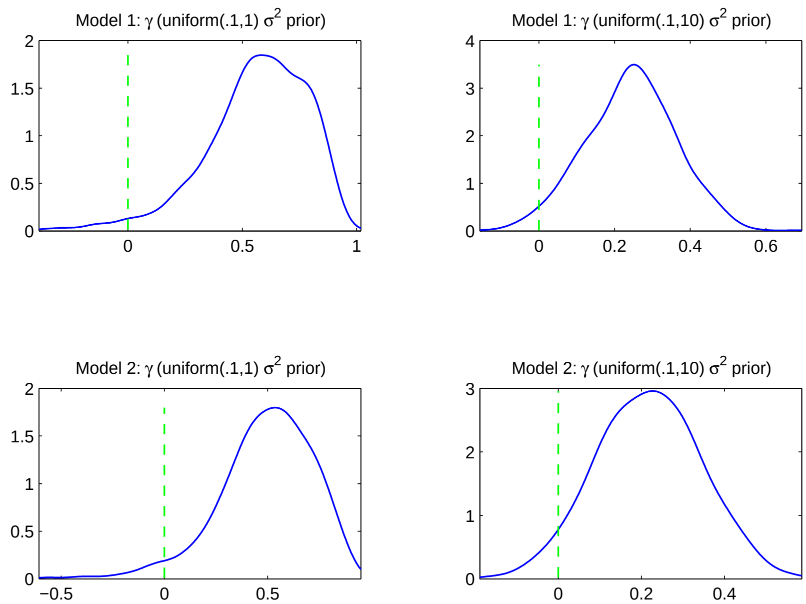

Fig. 5. Posterior mean estimates of the state correlation γ

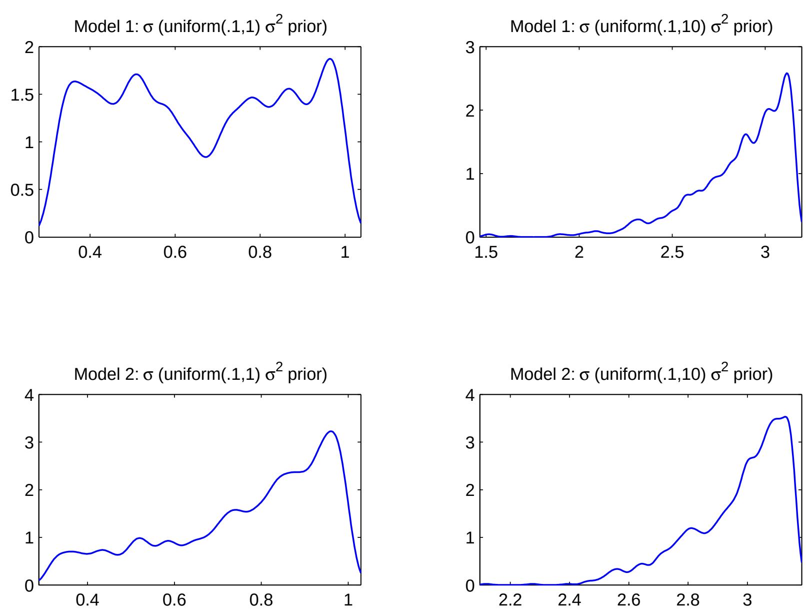

Fig. 6. Posterior mean estimates of σ

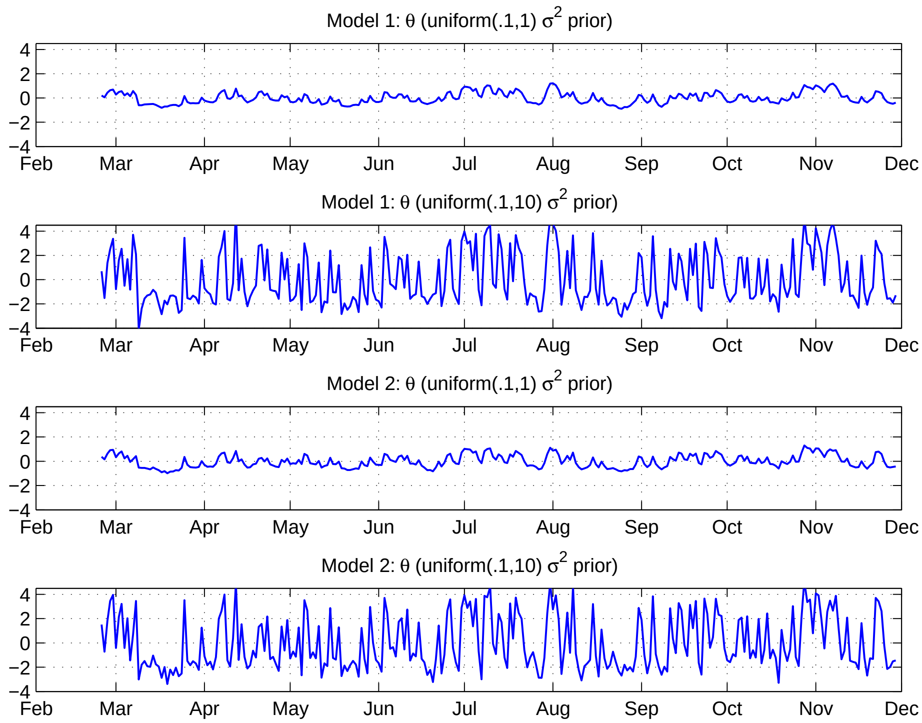

Fig. 7. Posterior mean estimates of state variables θ_t

Table 4. *Model fit* $\overline{D}$*, effective parameters* p_D *and DIC*

Model	$\overline{D}$	p_D	DIC
Model 1 (uniform(.1,1))	268.29	65.25	333.53
Model 1 (uniform(.1,10))	111.62	100.05	211.67
Model 2 (uniform(.1,1))	252.95	73.28	326.23
Model 2 (uniform(.1,10))	104.37	95.38	199.75

presentation we restrict this analysis to Wednesdays for the marginal analysis and to Tuesdays and Wednesdays for the joint analysis. To obtain prediction of the marginal headache probability under certain conditions, we utilize now that the $Var(Z_t)$ can be approximated by $1 + \frac{\sigma^2}{1-\gamma^2}$, therefore estimated marginal headache probabilities are given by

$$\hat{p}_t = \Phi\left[\frac{\mathbf{X}_t'\hat{\alpha}}{\sqrt{1 + \frac{\hat{\sigma}^2}{1-\hat{\gamma}^2}}}\right].$$

For example, in Figure 10 we plot estimated marginal headache probabilities for Wednesdays in dependency of windchill and humidity, respectively. The solid (dotted) lines indicate the presence

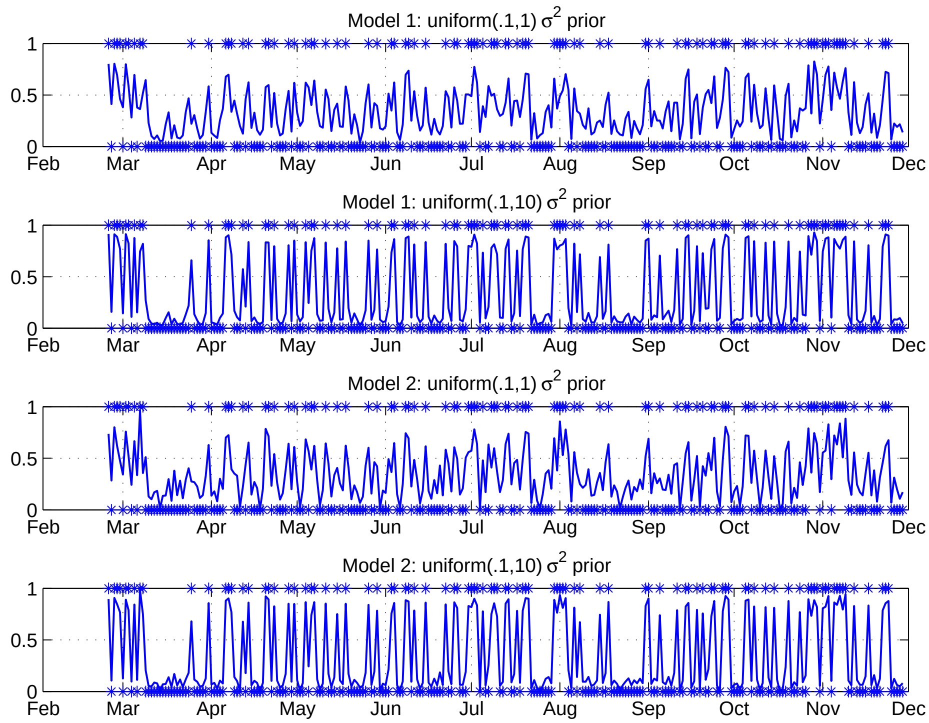

Fig. 8. Posterior mean estimates of probabilities p_t

(absence) of south-southeast winds. For this patient windchill influences the headache probability more severely than humidity. It is interesting to note that an increase in humidity decreases the headache probability somewhat for this patient.

Since dependency is present, a marginal analysis as above is insufficient. Therefore, we are interested in determining estimated joint headache probabilities. For this we utilize that according to (4.4) the correlation between Z_t and Z_{t+1} can be approximated by γ. Using the bivariate normal distribution function, Figures 11 and 12 give contour lines of estimated headache/no headache probabilities depending on different levels of windchill and humidity values for Tuesdays and Wednesdays when no south-southeast winds are present on both days, respectively.

Finally, we consider corresponding conditional probabilities, which are given in Figure 13. From this we see that the effect of a previous day headache influences the headache probability of the current day considerably. This shows that even a moderate correlation between the state variables induces a considerable dependency on the conditional probabilities. Thus dependency cannot be ignored.

5. Conclusions and Discussion

This paper investigates the usefulness and applicability of binary state space mixed models to pain risk management for migraine headache patients. Binary state space mixed models allow the joint assessment of a dynamic structure to model longitudinal dependency together with time

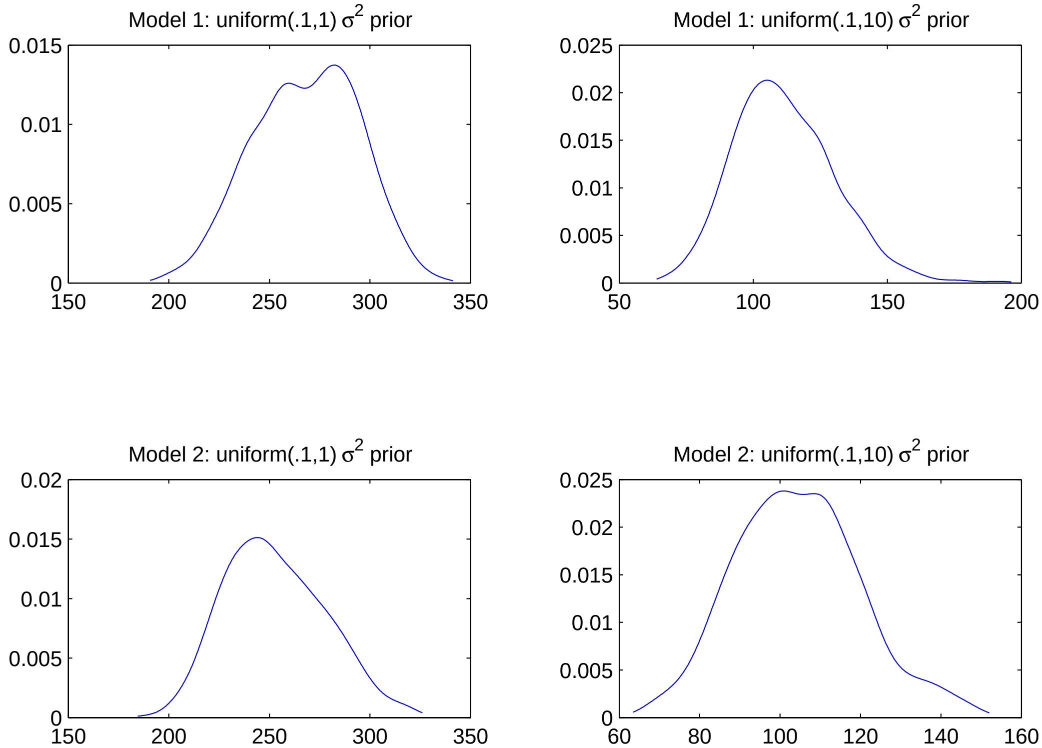

Fig. 9. Estimated density of the posterior deviance

dependent covariate effects such as weather conditions for migraine sufferers.

Parameter estimation is facilitated by MCMC. In addition to parameter estimation, the adaptation of several model fit and selection for a Bayesian inference using MCMC methods are presented. They allow for the assessment of mean and prior specifications. For the data set studied a reasonable model was found, which allows for a patient specific risk assessment.

For the final model selected, estimated risk profiles for suffering headaches are constructed under possible weather conditions. In addition to a marginal analysis, joint and conditional analyses are given. They show that even a moderate correlation among the state variables can induce considerable influence on estimated conditional headache probabilities. For example, the headache probability for Wednesdays is increased roughly by .10 when a headache was also present on Tuesday, while the fitted correlation among the state variable was only .25.

The original data set also measured headache severity on a five point scale. Therefore an extension of the methods presented to mixed state space models with ordinal responses is necessary and is the focus of current research efforts of the author. Another extension is to construct multivariate binary state space mixed models for the analysis of several patient dairies.

Acknowledgements:

The author's research was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungs-

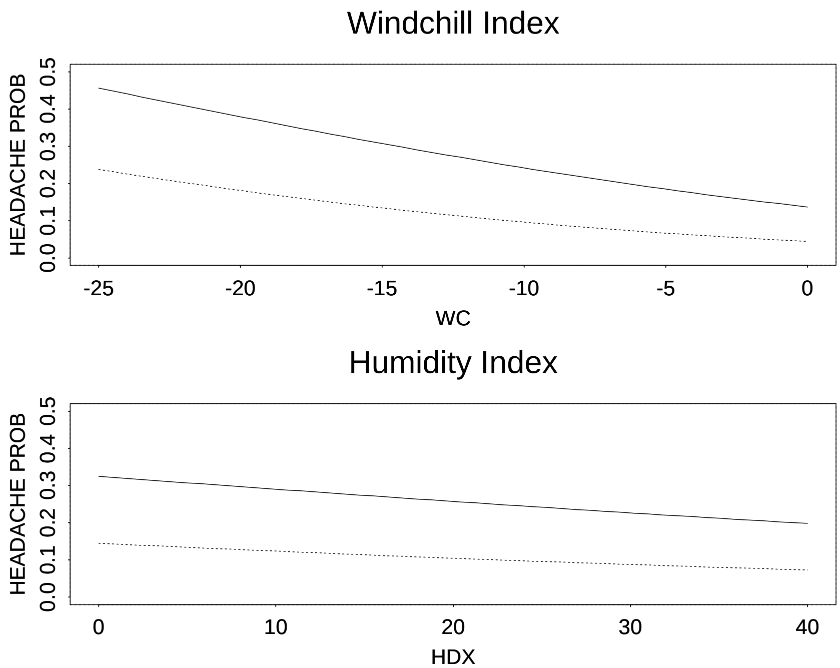

Fig. 10. Influence of windchill and humidity on the estimated marginal headache probabilities for Wednesdays in Model 1 with uniform(0,10) prior for σ^2 (—-= no south-southeast winds .....= south-southeast winds) winds)

bereich 386 *Statistische Analyse diskreter Strukturen.* The author is grateful to T. Kostecki-Dillon for permission to use her data on migraine headaches.

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.

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.

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

Cooke, L.J., Rose, M.S. and Becker, W.J. (2000). Chinook winds and migraine headache. *Neurology*, Jan 25;**54(2)** 302-7.

Cugini, P., Romit, A., Di Palma, L. and Giacovazzo, M. (1990). Common Migraine as a weekly and seasonal headache. *Chronobiology Int.*, **7**, 467-469.

Cull, R.E. (1981). Barometric pressure and other factors in migraine. *Headache*,**21**, 102-103.

Czado, C. and Song, P. X-K. (2001). State Space Mixed Models for longitudinal Observations with Binary and Binomial Responses, Discussion Paper 232, SFB 386, *Statistische Analyse Diskreter Strukturen* (http://www.stat.uni-muenchen.de/sfb386/).

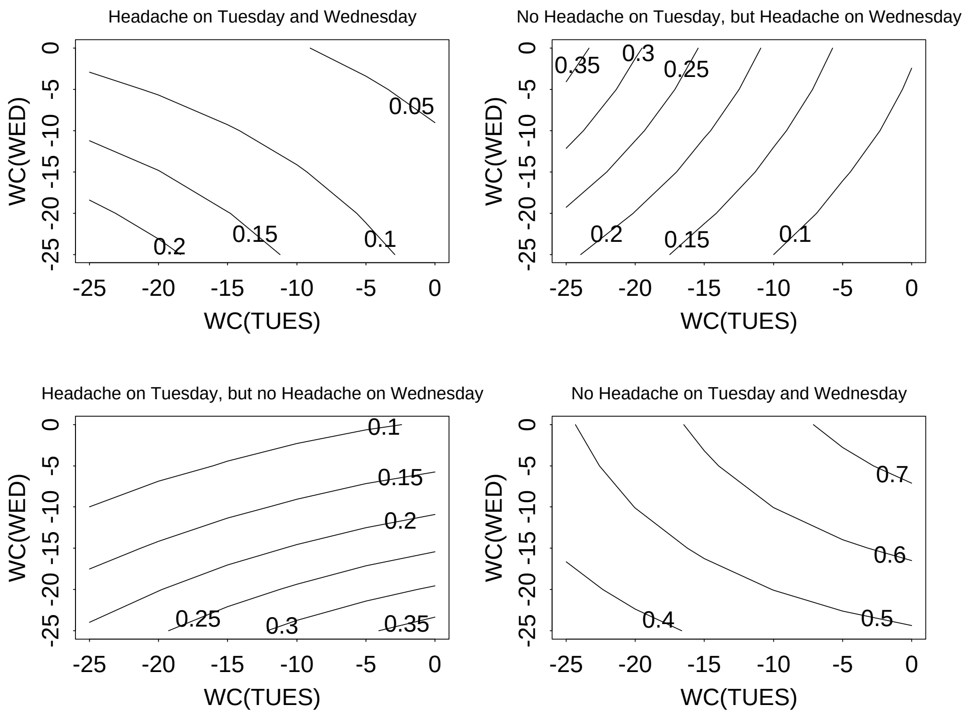

Fig. 11. Influence of windchill on joint probability estimates for Tuesdays and Wednesdays in Model 1 with uniform(0,10) prior for σ^2 when no south-southeast winds are present on both days

Dempster, A.P. (1974). The direct use of likelihood for significance testing. In *Proceedings of Conference on Foundational Questions in Statistical Inference*, (ed. O. Barndorff-Nielsen, P. Blaesild, and G. Schou), 335-52. Department of Theoretical Statistics: University of Aarhus.

Diamond, S., Freitag, F.G. and Nursall, A. (1990) The effects of weather on Migraine frequency in Chicago, *Headache Quarterly*,**1:2**, 136-145.

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

Gelman, A., Meng, X-L. and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion) *Statistica Sinica*, **6**, 733-760.

Gelman, A. and Meng, X-L. (1996). Model checking and model improvement. In *Markov Chain Monte Carlo in Practice*, 189-201, (ed. W.R. Gilks, S. Richardson and D.J. Spiegelhalter), Chapman & Hall, London.

Guttman, I. (1967). The use of the concept of a future observation in goodness-of-fit-problems. *J. R. Statist. Soc. B*,**29**, 83-100.

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.

Meng, X.-L. (1994). Posterior predictive p-values. *Ann. Statist.*, **22**, 1142-1160.

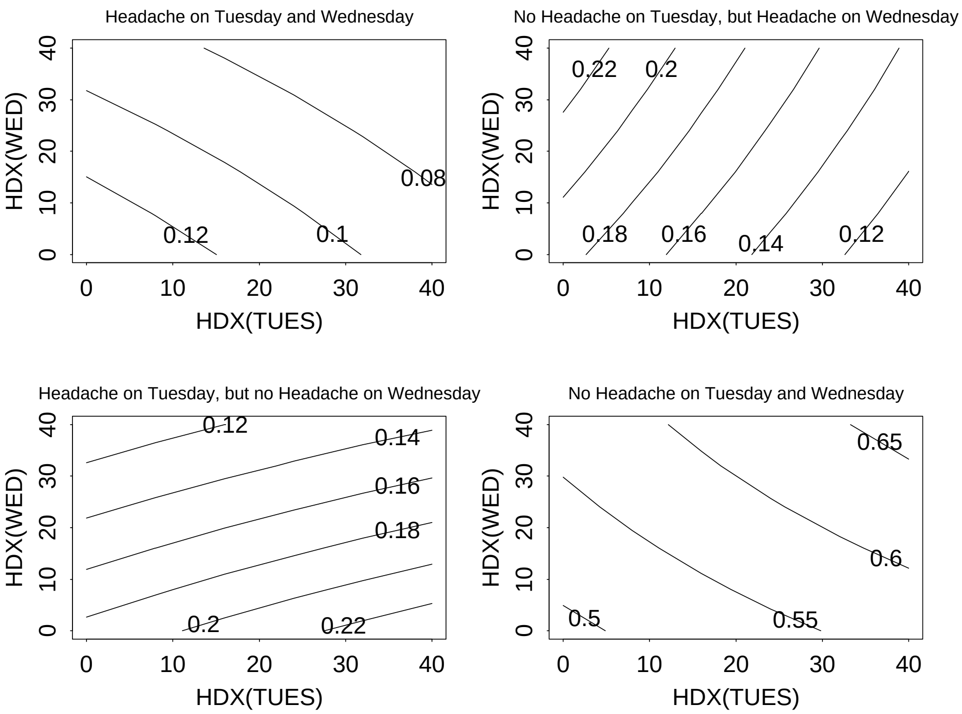

Fig. 12. Influence of humidity on joint probability estimates for Tuesdays and Wednesdays in Model 1 with uniform(0,10) prior for σ^2 when no south-southeast winds are present on both days

Nursall, A. (1981) *The effects of weather conditions on the frequency and severity of migraine headaches in southwestern Ontario*, unpublished M.Sc. thesis, University of Alberta.

Nursall, A., Phillips, D.W. (1980). The effects of weather on the frequency and severity of migraine headaches in southwestern Ontario, unpublished manuscript, *Canadian Climate Centre Report*,**80-87**, Atmospheric Environment Service, Toronto.

Raskin, N.H. (1988). *Headache*, 2nd ed. New York, Churchill Livingstone.

Rubin, D. B. (1981). Estimation in parallel randomized experiments. *J. Educ. Statist.*,**6**, 377-401.

Rubin, D. B. (1984). Bayesian justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **12**, 1151-1172.

Schulman, J., Leviton, A., Slack, W., Porter, D. and Graham, J.R. (1980). The relationship of headache occurrence to bariometric pressure. *Int. J. Biometeorol.*, **24**, 263-269.

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. and Carlin, B.P. (1998). Bayesian deviance, the effective number of parameters and the comparison of arbitrarily complex models, preprint.

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

West, M. and Harrison, P.J. (1989). *Bayesian Forecasting and Dynamic Models*, Springer, New York.

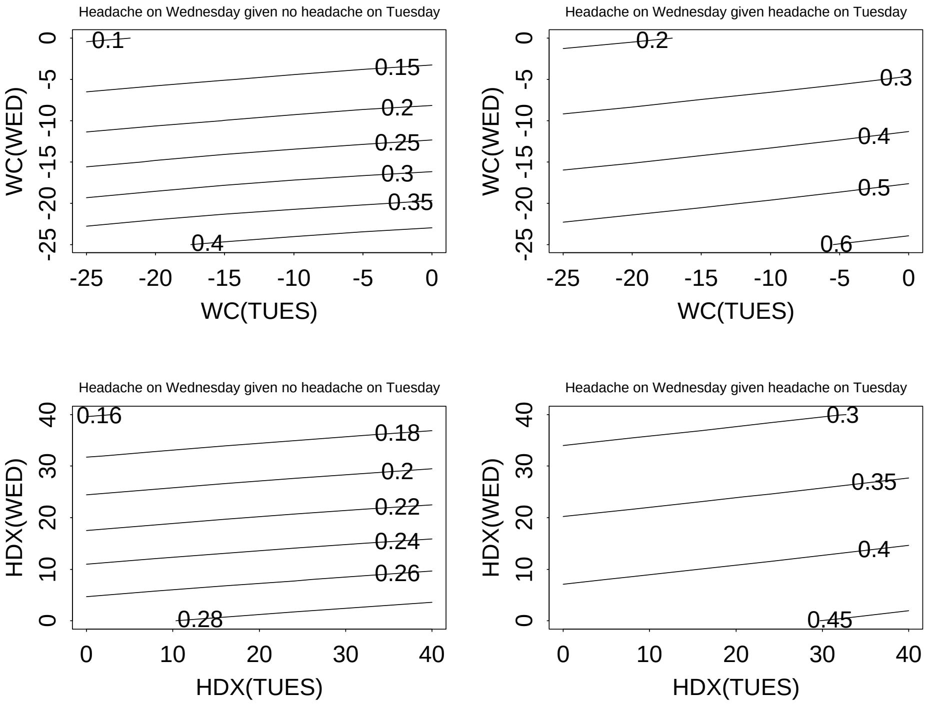

Fig. 13. Influence of windchill and humidity on conditional probability estimates for Tuesdays and Wednesdays in Model 1 with uniform(0,10) prior for σ^2 when no south-southeast winds are present on both days

Wilkinson, M. and Woodrow, J. (1979). Migraine and weather. *Headache*, **19** , 375-378.

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

Zeger, S.L. and Liang K-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121-130.

[*Received*]