Czado, Heyn, Müller:

Modeling migraine severity with autoregressive ordered probit models

Sonderforschungsbereich 386, Paper 413 (2005)

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

Modeling migraine severity with autoregressive ordered probit models

Claudia CZADO, Anette HEYN, Gernot MÜLLER

Technische Universität München, Boltzmannstr. 3, D-85747 Garching, Germany

March 16, 2005

SUMMARY. This paper considers the problem of modeling migraine severity assessments and their dependence on weather and time characteristics. Since ordinal severity measurements arise from a single patient dependencies among the measurements have to be accounted for. For this the autoregressive ordinal probit (AOP) model of Müller and Czado (2004) is utilized and fitted by a grouped move multigrid Monte Carlo (GM-MGMC) Gibbs sampler. Initially, covariates are selected using proportional odds models ignoring this dependency. Model fit and model comparison are discussed. The analysis shows that humidity, windchill, sunshine length and pressure differences have an effect in addition to a high dependence on previous measurements. A comparison with proportional odds specifications shows that the AOP models are preferred.

KEY WORDS: Proportional odds; autoregressive component, ordinal valued time series, regression, Markov Chain Monte Carlo (MCMC), deviance;

---

email: cczado@ma.tum.de

1. INTRODUCTION

According to Prince, Rapoport, Sheftell, Tepper and Bigal (2004) forty-five million Americans seek medical attention for head pain yearly causing an estimated labor cost of $ 13 billion. They found in their study that about half of their migraine patients are sensitive to weather. However some studies investigating the relationship between weather conditions and headache have been negative or inclusive (see Prince et al. (2004) and Cooke, Rose and Becker (2000) for specific references). In these studies the frequency of headache occurrences and the daily maximum or total score of an ordinal severity assessment have been the focus.

Here we focus directly on studying and modeling the observed severity categories collected using a headache calendar. In particular we want to investigate the 4 daily ratings of the headache intensity obtained from a patient in study conducted by psychologist T. Kostecki-Dillon, Toronto, Canada, resulting in an ordinal valued time series.

Most studies ignore correlation among measurements on the same patient. We will show that this correlation can be very high and should not be ignored. For example Prince et al. (2004) use daily maximum and total scores as response variable relating to factors obtained from a factor analysis of the weather data alone in a regression setup ignoring this correlation.

For studying headache occurrences Piorecky, Becker and Rose (1996) used a generalized estimating approach (GEE) introduced by Zeger and Liang (1986) to adjust for the dependency between multiple measurements. While GEE could also be used for ordinal valued time series (see for example Liang, Zeger and Qaqish (1992), Heagerty and Zeger (1996) and Fahrmeir and

Pritscher (1996)), we prefer a likelihood based regression time series approach to investigate the influence of weather conditions on migraine severity.

Kauermann (2000) also considered the problem of modeling ordinal valued time series with covariates. He used a nonparametric smoothing approach by allowing for time varying coefficients in a proportional odds model. While Kauermann (2000) uses local estimation, Gieger (1997) and Fahrmeir, Gieger and Hermann (1999) consider spline fitting within the GEE framework. Wild and Yee (1996) focus on smooth additive components. While these approaches are useful for fitting the data, a hierarchical time series approach which we propose here is easier to interpret and has the potential for forecasting. In particular, we will use an autoregressive ordered probit (AOP) model recently introduced by Müller and Czado (2004). It is based on a threshold approach using a latent real valued time series. It is fitted and validated in a Bayesian setting using Markov Chain Monte Carlo (MCMC) methods.

We will restrict our analysis to data obtained from a single patient, since we believe that the complex relationship between headache severity and weather conditions is best captured by an individual analysis. Such an approach was also followed by Schmitz and Otto (1984). However they ignored the ordinal nature of the considered response time series. We investigate data collected by a 35 year old woman with chronic migraine who recorded her migraine severity four times a day on a scale from 0 to 5. To determine which weather conditions have an important effect on the migraine severity we used a proportional odds model commonly used for regression models with independent ordinal responses as a starting model for our AOP analysis. We will show that for this data the first order autocorrelation in the latent time series is high within the AOP model ($\approx .8$), demonstrating considerable dependence among the measurements.

The paper is organized as follows. In Section 2 we review the proportional odds model to motivate our AOP formulation. We address the problem of variable selection and model comparison. In Section 3 we describe the data in more detail and present some results from an exploratory analysis yielding three mean specifications for the proportional odds model and two for the AOP model. In Section 4 we give the results of the model fitting and model comparison, demonstrating the superiority of the AOP model. Finally Section 5 gives a summary and draws conclusions.

2. MODELS, PREDICTIONS, AND MODEL SELECTION

2.1 *Models*

In the migraine data we model an ordinal valued time series $\{Y_t, t = 1, \ldots, T\}$, where $Y_t \in \{0, \ldots, K\}$ denotes the pain severity at time t with ordinal levels given by $\{0, \cdots, K\}$. Together with the response Y_t we observe further a vector $\boldsymbol{x}_t$ of real-valued covariates for each $t \in \{1, \ldots, T\}$ representing metrological and time measurement information.

2.1.1 Proportional Odds Model A common ordinal regression model for independent responses is the ordinal logistic model first described by Walker and Duncan (1967) and later named proportional odds model by McCullagh (1980). We will use the proportional odds model as a starting model to identify important covariates for in the context of the migraine headache data more appropriate autoregressive ordinal probit (AOP) model. Since we will base our AOP model formulation on a threshold approach, we also present the proportional odds model in this way.

For this we assume that the covariate vector $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tp})'$ is p-dimensional. To model the $K+1$ different categories, an underlying unobserved real-valued time series $\{Y_t^*, t = 1, \ldots, T\}$ is used which produces the discrete valued Y_t by thresholding. In particular,

$$Y_t = k \iff Y_t^* \in (\alpha_{k-1}, \alpha_k], \qquad k = 0, \ldots, K, \tag{2.1}$$

$$Y_t^* = -\boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t^*, \qquad t = 1, \ldots, T, \tag{2.2}$$

where $-\infty =: \alpha_{-1} < \alpha_0 < \alpha_1 < \cdots < \alpha_K := \infty$ are unknown cutpoints, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown regression coefficients. The errors ε_t^* are assumed to be i.i.d. and follow a logistic distribution with distribution function given by

$$F(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

It is easy to see that (2.1)-(2.2) imply the more familiar representation given by

$$P(Y_t \leq k | \boldsymbol{x}_t) = F(\alpha_k + \boldsymbol{x}_t'\boldsymbol{\beta}) = \frac{\exp(\alpha_k + \boldsymbol{x}_t'\boldsymbol{\beta})}{1 + \exp(\alpha_k + \boldsymbol{x}_t'\boldsymbol{\beta})} \tag{2.3}$$

for $k = 0, 1, \cdots, K-1$. The properties of the proportional odds model are for example discussed in Harrell (2001) and Agresti (2002). Let $\{y_t, t = 1, \cdots, T\}$ be the observed responses and $\boldsymbol{\alpha} := (\alpha_0, \ldots, \alpha_{K-1})'$. Since the responses are assumed to be independent the joint likelihood is given by

$$L(\boldsymbol{\beta}, \boldsymbol{\alpha}) := L(\boldsymbol{\beta}, \boldsymbol{\alpha} | y_1, \cdots, y_T) = \prod_{t=1}^{T} \pi_{t, y_t}, \tag{2.4}$$

where $\pi_{tk} := P(Y_t = k | \boldsymbol{x}_t) = F(\alpha_k + \boldsymbol{x}_t'\boldsymbol{\beta}) - F(\alpha_{k-1} + \boldsymbol{x}_t'\boldsymbol{\beta})$ for $k = 0, \cdots, K-1$ and $\pi_{tK} := 1 - \sum_{k=0}^{K-1} \pi_{tk}$. The unknown $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ together with the ordering constraint $-\infty =: \alpha_{-1} < \alpha_0 < \alpha_1 < \cdots < \alpha_K := \infty$ can be estimated by maximum likelihood (ML) using the S-Plus Design Library by Frank Harrell.

2.1.2 Autoregressive Ordered Probit (AOP) Model Since the migraine severity at time t may depend not only on the covariates at time t, but also on the migraine severity at time $t-1$, it may be adequate to use the autoregressive ordered probit (AOP) model introduced by Müller and Czado (2004). Here, the latent process of the common ordered probit model is extended by an autoregressive component:

$$Y_t = k \iff Y_t^* \in (\alpha_{k-1}, \alpha_k], \qquad k = 0, \ldots, K, \tag{2.5}$$

$$Y_t^* = \boldsymbol{x}_t' \boldsymbol{\beta} + \phi Y_{t-1}^* + \varepsilon_t^*, \qquad t = 1, \ldots, T, \tag{2.6}$$

where $-\infty =: \alpha_{-1} < \alpha_0 < \alpha_1 < \cdots < \alpha_K := \infty$, $\varepsilon_t^* \sim N(0, \delta^2)$ i.i.d., and $\boldsymbol{x}_t = (1, x_{t1}, \ldots, x_{tp})'$ is a $p+1$-dimensional vector of real-valued covariates. Accordingly, β_0 is the intercept for the latent process. For reasons of identifiability the cutpoint α_0 is fixed to 0, and the variance δ^2 to 1. For notational convenience we use $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_{K-1})'$ as for the proportional odds model, however, since α_0 is fixed here, the vector $\boldsymbol{\alpha}$ has only $K-1$ components in the AOP case. More details on this model and a Markov chain Monte Carlo (MCMC) estimation procedure for the latent variables and parameters can be found in Müller and Czado (2004).

In particular, it is shown there that a standard Gibbs sampling approach is extremely inefficient and cannot be recommended in practice. This inefficiency of the Gibbs sampler was already noted by Albert and Chib (1993) for polychotomous regression models and Chen and Dey (2000) for correlated ordinal regression data using lagged covariates to account for correlation. Nandram and Chen (1996) proposed a scale reparametrization for ordinal regression models with three categories, which accelerated the Gibbs sampler in this situation sufficiently. The reason for the inefficiency in ordinal response models is that the updating scheme for the cutpoints $\boldsymbol{\alpha}$ allows only small movements from one iteration to the next in larger data sets. To overcome this inefficiency Müller and Czado (2004) developed a specific grouped move multigrid Monte Carlo (GM-MGMC) Gibbs sampler for the AOP model with arbitrary number of categories. GM-MGMC Gibbs samplers have been suggested by Liu and Sabatti (1996) as a general approach to accelerate Gibbs sampling schemes.

We emphasize that the right-hand side of Equation (2.2) includes the term $-\boldsymbol{x}_t'\boldsymbol{\beta}$ whereas the right-hand side of Equation (2.6) uses the term $\boldsymbol{x}_t'\boldsymbol{\beta}$. To make the parameters β_j in model specifications (2.2) and (2.6) comparable we decided to compute the posterior mean estimates in the AOP model for the response $Y_t^\circ := 5 - Y_t$. Therefore the worst migraine severity is associated with category 0, and no migraine is associated with category 5 when we fit the AOP model. Hence now in both the proportional odds and in the AOP model a negative value for β_j means that an increasing value of the covariate x_j leads to a more severe migraine.

2.2 *Model Selection with the Deviance Criteria*

2.2.1 Residual Deviance Test for the Proportional Odds Model Here we use the deviance statistic D defined as

$$D := 2 \log \frac{\sup_{\boldsymbol{\beta},\boldsymbol{\alpha}} L(\boldsymbol{\beta},\boldsymbol{\alpha})}{\sup_{\boldsymbol{p}_1,\cdots,\boldsymbol{p}_T} L(\boldsymbol{p}_1,\cdots,\boldsymbol{p}_T)},$$

where $L(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is defined in (2.4) and the supremum is taken over all $\boldsymbol{\alpha}$ which satisfy the ordering constraint. Further we denote by $L(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_T)$ for $\boldsymbol{p}_t := (p_{t0}, \cdots, p_{tk})'$ the joint likelihood of T independent discrete random variables Z_t taking on values $0, \cdots, K$ with probabilities $p_{t0}, \cdots, p_{tK}$, respectively. We call $L(\boldsymbol{p}_1, \cdots, \boldsymbol{p}_T)$ the likelihood of the corresponding unstructured model. It is straight forward to show that

$$D := \sum_{t=1}^{T} \log(\bar{\pi}_{tk}),$$

where $\bar{\pi}_{tk} := F(\bar{\alpha}_k + \boldsymbol{x}_t'\bar{\boldsymbol{\beta}}) - F(\bar{\alpha}_{k-1} + \boldsymbol{x}_t'\bar{\boldsymbol{\beta}})$ and $\bar{\boldsymbol{\beta}}$ and $\bar{\boldsymbol{\alpha}}$ the joint MLE of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ under the ordering constraint for $\boldsymbol{\alpha}$. Note that the proportional odds model can be considered as a special case of multi categorical models considered in Tutz (2000). Here he shows that the null hypothesis of model adequacy can be rejected at level α if

$$D > \chi^2_{T \cdot K - p, 1-\alpha},$$

where T is the number of observations, K the number of categories minus 1 and p the number of regression parameters to be estimated. The χ^2 approximation is most accurate when covariates are categorical and the expected cell counts formed by the cross classification of the responses and covariates are greater than 5. Alternative goodness-of-fit tests in ordinal regression models have been suggested in Lipsitz, Fitzmaurice and Molenberghs (1996). We restrict our attention to the residual deviance, since we want to use the deviance information criterion for the AOP model, which is closely related to the deviance.

2.2.2 Deviance Information Criterion for the AOP model As model selection criterion for the AOP models we use the Deviance Information Criterion (DIC) of Spiegelhalter, Best, Carlin and van der Linde (2002). Here model fit is measured by a deviance statistic and model complexity is taken into account.

First, the Bayesian deviance is defined as

$$D(\boldsymbol{\theta}) := -2\log\{f(y|\boldsymbol{\theta})\} + 2\log\{f(y)\},$$

where $2\log\{f(y)\}$ can be considered as a standardizing constant depending on the data alone. In the following we set this constant to zero, which is consistent with the unstructured model occurring in the deviance statistics and use

$$D(\boldsymbol{\theta}) = -2\log\{f(y|\boldsymbol{\theta})\}$$

as Bayesian deviance. The effective number of model parameters is denoted by p_D and can be evaluated as the difference of the posterior mean of the deviance, $\overline{D(\boldsymbol{\theta})} := (D(\boldsymbol{\theta})|y)$, and the deviance at the posterior mean for the parameters of interest, $D(\bar{\boldsymbol{\theta}}) := D(\mathbb{E}(\boldsymbol{\theta}|y))$ (cf. Spiegelhalter et al. (2002), p.587), so that

$$p_D := \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}),$$

where $\bar{\boldsymbol{\theta}} = \mathbb{E}(\boldsymbol{\theta}|y)$. As model selection criterion Spiegelhalter et al. (2002) now suggest the DIC defined as

$$\mathrm{DIC} := D(\bar{\boldsymbol{\theta}}) + 2p_D = \overline{D(\boldsymbol{\theta})} + p_D = 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}).$$

A model with smaller DIC fits the data better. Moreover we see that the effective number of parameters acts as penalizing term.

For the AOP model the parameter $\boldsymbol{\theta}$ includes the cutpoint vector $\boldsymbol{\alpha}$, the regression parameter vector $\boldsymbol{\beta}$, the autoregressive parameter ϕ, and all the latent variables Y_t^*. The Bayesian deviance for the model is

$$\begin{aligned} D(\theta) &= -2\log f(y \mid \boldsymbol{\theta}) \\ &= -2\sum_{t=1}^{T} \log\left[\Phi(\alpha_k - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*) - \Phi(\alpha_{k-1} - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*)\right] \end{aligned} \tag{2.7}$$

To compute the DIC, the expression $\overline{D(\boldsymbol{\theta})}$ can be estimated by averaging the terms $D(\boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ denotes the random sample for $\boldsymbol{\theta}$ drawn in iteration i of the MCMC sampler. The value of $D(\bar{\boldsymbol{\theta}})$ is given by inserting the corresponding posterior mean estimates in Equation (2.7).

2.3 *Pseudo-predictions*

One intuitive and quite simple way to investigate the quality of a model fit is to compute pseudo-predictions. In the proportional odds model this means that one predicts the response at time t using ML estimates for the regression parameters and cutpoints which are plugged into the model equations. This results in a forecast probability for each category. One can use the category with highest forecast probability as prediction for the response at time t. However, when the ML estimates are based on the whole data set we call these predictions more precisely pseudo-predictions. For the AOP model one uses posterior mean estimates instead of the ML estimates. Here, of course, one also needs a posterior mean estimate of Y_{t-1}^*.

2.3.1 Pseudo-predictions for the Proportional Odds Model The fitted probabilities for the proportional odds model for each category at time t

are defined by

$$\begin{aligned}
\bar{\pi}_{t0} &:= \hat{P}(Y_t = 0 \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}) = \frac{\exp(\overline{\alpha}_0 + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}{1+\exp(\overline{\alpha}_0 + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}, \\
\bar{\pi}_{tk} &:= \hat{P}(Y_t = k \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}) = \frac{\exp(\overline{\alpha}_k + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}{1+\exp(\overline{\alpha}_k + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})} - \frac{\exp(\overline{\alpha}_{k-1} + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}{1+\exp(\overline{\alpha}_{k-1} + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 1, \ldots, K-1, \\
\bar{\pi}_{tK} &:= \hat{P}(Y_t = K \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}) = 1 - \frac{\exp(\overline{\alpha}_{K-1} + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}{1+\exp(\overline{\alpha}_{K-1} + \boldsymbol{x}_t'\overline{\boldsymbol{\beta}})}
\end{aligned}$$

where $\overline{\boldsymbol{\alpha}}$ and $\overline{\boldsymbol{\beta}}$ denote maximum likelihood estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. The corresponding pseudo-prediction of Y_t is therefore given by the category k, which has the highest value among $\bar{\pi}_{t0}, \ldots, \bar{\pi}_{tK}$.

2.3.2 Pseudo-predictions for the AOP model The corresponding posterior probability estimates in the AOP model for each category at time t are defined by

$$\begin{aligned}
\bar{\pi}_{t0} &:= \hat{P}(Y_t = 0 \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \overline{Y}^*_{t-1}) = \Phi(\overline{\alpha}_0 - \boldsymbol{x}_t'\overline{\boldsymbol{\beta}} - \overline{\phi}\,\overline{Y}^*_{t-1}), \\
\bar{\pi}_{tk} &:= \hat{P}(Y_t = k \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \overline{Y}^*_{t-1}) = \Phi(\overline{\alpha}_k - \boldsymbol{x}_t'\overline{\boldsymbol{\beta}} - \overline{\phi}\,\overline{Y}^*_{t-1}) \\
&- \Phi(\overline{\alpha}_{k-1} - \boldsymbol{x}_t'\overline{\boldsymbol{\beta}\phi}\,\overline{Y}^*_{t-1}), \qquad k = 1, \ldots, K-1, \\
\bar{\pi}_{tK} &:= \hat{P}(Y_t = K \mid \boldsymbol{x}_t, \overline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\beta}}, \overline{\phi}, \overline{Y}^*_{t-1}) = 1 - \Phi(\overline{\alpha}_{K-1} - \boldsymbol{x}_t'\overline{\boldsymbol{\beta}} - \overline{\phi}\,\overline{Y}^*_{t-1}).
\end{aligned}$$

where $\overline{\boldsymbol{\alpha}}$, $\overline{\boldsymbol{\beta}}$, $\overline{\phi}$, and $\overline{Y}^*_{t-1}$ denote posterior mean estimates of the corresponding parameters and latent variables. The corresponding pseudo-prediction of Y_t are therefore given by the category k which has the highest value among $\bar{\pi}_{t0}, \cdots, \bar{\pi}_{tK}$.

2.3.3 Assessing Model fit based on Pseudo-predictions Now we suggest to use the pseudo-predictions for model assessment. For this we define the variables P_{tk}^{obs} which correspond to the 'observed' probabilities for category k at time t in contrast to the 'predicted' probabilities $\bar{\pi}_{tk}$ defined in the previous subsections:

$$P_{tk}^{obs} := \begin{cases} 1 & \text{if } Y_t = k, \\ 0 & \text{else.} \end{cases}$$

When category k is observed at time t, it is clear that a good model fit leads to a high probability $\bar{\pi}_{tk}$, and to small probabilities $\bar{\pi}_{tj}$ for the other categories $j \neq k$. A large difference should be punished more than a small difference. Therefore we compute the verification score introduced by Brier (1950) defined by

$$\mathrm{S} := \frac{1}{T} \sum_{k=0}^{K} \sum_{t=1}^{T} (P_{tk}^{obs} - \bar{\pi}_{tk})^2$$

to get an idea of the model fit. Of course, the smaller the value of S, the better the model. The Brier score has been heavily used to evaluate forecasts in the metrological sciences and has the attractive property of being a strictly proper scoring rule (see for example Gneiting and Raftery, 2004).

3. ANALYSIS OF MIGRAINE SEVERITY DATA

3.1 *Data description and exploratory analysis*

We investigate the migraine headache dairy of a 35 year old female, who is working full-time as a manager. She suffers from migraine without aura for 22 years. In this study she recorded her headache four times a day on an ordinal scale from 0 to 5, where 0 means that she did not feel any migraine headache, and 5 the worst migraine headache she can feel. For a precise definition of the migraine intensity categories see Table 1. The data is part of a larger study on determinants of migraine headaches collected by the psychologist T. Kostecki-Dillon, York University, Toronto, Canada. The migraine headache dairy was completed between January 6, 1995, and

September 30, 1995, which is a period of 268 subsequent days. Therefore the length of the data set is $4 \cdot 268 = 1072$. Since she believes that her migraine headache is triggered by weather conditions, also weather related information on a daily basis was collected. This includes information on humidity, windchill, temperature and pressure changes, wind direction, and length of sun shine on the previous day.

Table 1 contains also the frequencies for the six possible response categories in the data set. As can be seen from this table 150 observations are unequal to zero which corresponds to suffering from migraine headaches in about 14% of the time. On the one hand we use covariates which reflect weather conditions, on the other hand covariates which contain information about the measurement time points. A description of the covariates in our analysis is also provided in Table 1. We point out that the humidity index is measured only in the period from May to October and the windchill index only in the period from November to April. This means that always only one of these covariates is contained in the data set.

In the following we conduct a short exploratory analysis. As described in Müller and Czado (2004), the idea is to compute the average response for each category of a categorical covariate and for intervals, when a continuous covariate is considered. Depending on the shape of the graph one can then decide to use an appropriate transformation of the covariate or to use indicator variables, which is, of course, the most flexible way of modeling.

PMND1P (mean pressure change from previous day, cf. Figure 1, top panel): We group the observed PMND1P values into six intervals with equal number of observations and compute the average response for each interval.

Table 1

Description of response scales with observed frequencies and weather and time measurements related covariates

Response categories		
intensity	frequency	condition
0	922	No headache
1	27	Mild headache: Aware of it only when attending to it
2	46	Moderate headache: Could be ignored at times
3	47	Painful headache: Continuously aware of it, but able to start or continue daily activities as usual
4	24	Severe headache: Continuously aware of it. Difficult to concentrate and able to perform only undemanding tasks
5	6	Intense headache: Continuously aware of it, incapacitating. Unable to start or contiunue activity.

weather conditions	
PMND1P	mean pressure change since previous day in 0.01 kilopascal
S1P	length of sunshine on previous day in hours
HDXDD	humidity index based on maximal temperature and humidity, only in period May to October, 0 otherwise
WCD	windchill index based on minimal temperature and wind speed, only in period November to April, 0 otherwise
WC.IND	indicator for windchill: 1 if WCD unequal 0, 0 otherwise
time of measurement	
WDAY	weekday, also coded by 1 (Monday) to 7 (Sunday)
MESS	time of measurement: HAAM = morning (also coded by 1), HANOON = noon (2), HAPM = afternoon (3), HABED = late evening (4)
HAPM.IND	indicator for afternoon: 1 if MESS=HAPM, 0 otherwise

A linear relationship seems to be sufficient, since a possibly present quadratic part is obviously small.

S1P (sunshine on previous day): This covariate has not been collected 120 times in the considered period. The remaining 952 observations are grouped in intervals. The relationship is quite linear (not shown), and a sunny day seems to increase the probability for headache on the following day, since the average response increases with the length of the sunshine. The range of the average response is 0.31.

HDXDD (humidity index): We computed the average response for each interval. and decided to use a quadratic transformation. The relative high range among these average responses of 0.83 is a first hint at the importance of this covariate.

WCD (windchill): We use an indicator for windchill. If windchill is present, the patient suffered from more intense migraine headaches.

WDAY (weekday): Because of the periodicity a polynomial or logarithmic transformation does not make sense. Perhaps a sine transformation could be used. We use indicator variables since this choice provides the most flexible way for modeling the influence of the weekdays. Indicator variables are abbreviated in a natural way. For example, the variable TUEWED is 1 if the measurement was done on a Tuesday or Wednesday, otherwise 0.

MESS (time of measurement, cf. Figure 1, bottom panel): In the afternoon the average response is the highest with 0.51. The difference between the range of the average response is $0.51 - 0.26 = 0.25$. The afternoon indicator HAMP.IND is used.

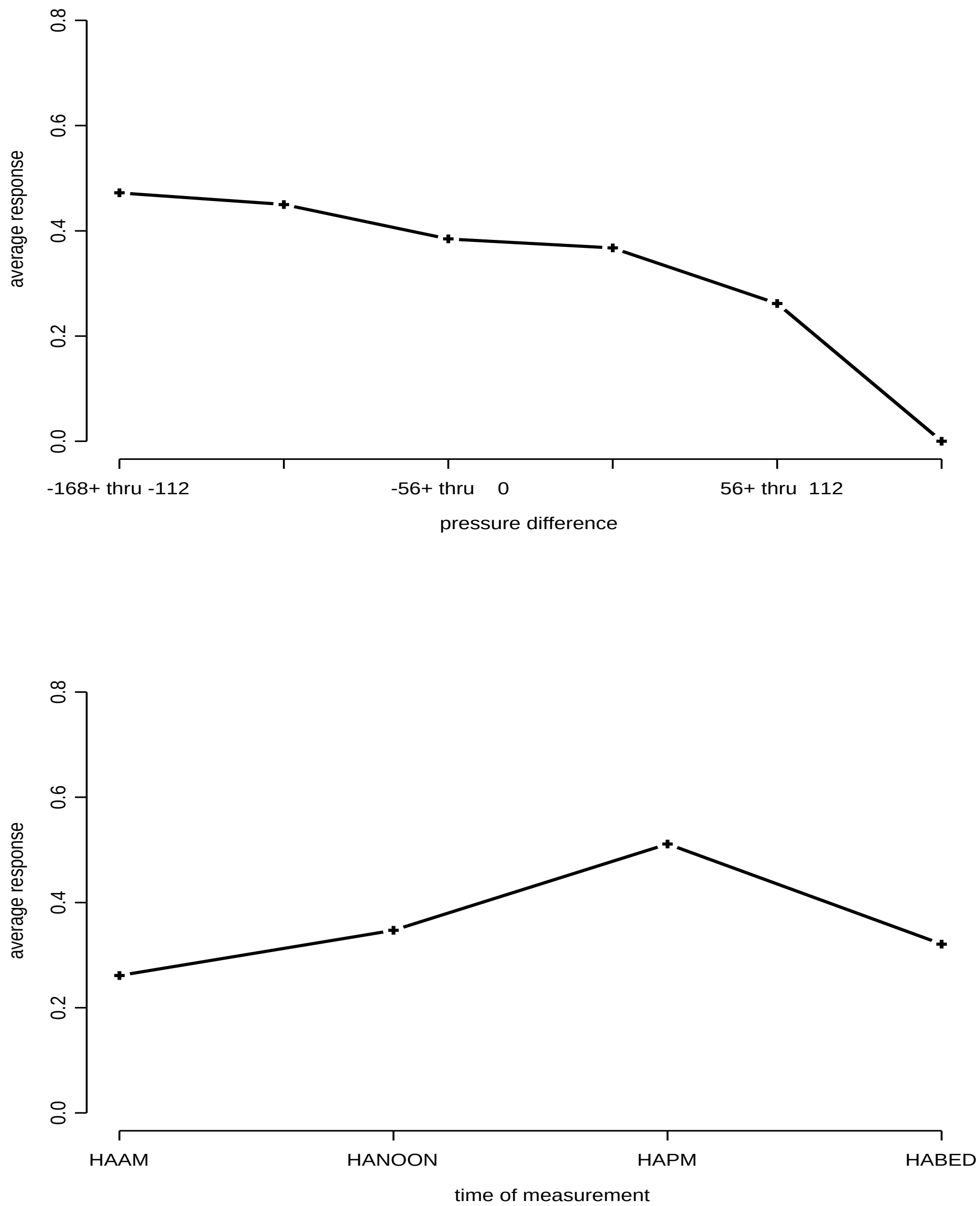


Figure 1. Relationship between average response and pressure difference intervals (top panel) and average response and time of measurement (bottom panel).

3.2 *Proportional Odds Model Specifications*

To determine reasonable mean specifications for the AOP model we ignore in an initial analysis the dependency among the responses und utilize the proportional odds model. For the proportional odds model we analyzed models with different sets of covariates. As mentioned the covariate 'sunshine on previous day' has not been collected 120 times in the period. We remove these measurements and reduce our data set to the length $1072 - 120 = 952$. The 3 models A, B, and C considered in the following are found by a forward selection procedure. In each step the p-values for each covariate were determined by a Wald test. The covariate with the best p-value was included in the model until the 5% level was reached. This means that the covariates of Model A, B, and C are all significant on the 5% level.

Model A contains only main effects. For time of measurement we use only an indicator for the afternoon measurement and an indicator for Tuesday or Wednesday. In Model B and C we consider 3 weekday indicators following our exploratory analysis. Furthermore, in Model B we also allow for 3 interaction effects, whereas Model C contains even 9 interaction components. The covariates which are used are seen in Table 2. This table also gives the ML estimators for the regression coefficients and the cutpoints.

3.3 *AOP Model Specifications*

For the AOP model with latent variables given by

$$Y_t^* = x_t' \boldsymbol{\beta} + \phi Y_{t-1}^* + \varepsilon_t^*$$

we investigate two models. For numerical stability we use covariates which have been standardized such that they have empirical mean 0 and empirical

Table 2

Maximum likelihood estimates of regression parameters and cutpoint parameters, residual deviances and Brier scores using the proportional odds model ignoring dependency

	Model A	Model B	Model C
weather conditions			
HDXDD_t	-0.4592	-0.4513	-0.4011
HDXDD_t^2	0.0109	0.0106	0.0097
$\mathrm{S1P}_t$	-0.1055	-0.1205	-0.0651
$\mathrm{WC.IND}_t$	-4.6821	-4.7190	-4.3610
$\mathrm{PMND1P}_t$	0.0035	-0.0149	-0.0147
time of measurement			
$\mathrm{HAPM.IND}_t$	-0.4719	-0.5051	-0.5433
TUEWED_t	0.5298		
TUESUN_t		-0.2180	-1.0196
WEDFRI_t		-0.2542	1.9105
THUSAT_t		-0.3935	-0.5628
interactions			
$\mathrm{PMND1P}_t \cdot \mathrm{TUESUN}_t$		0.0150	0.0174
$\mathrm{PMND1P}_t \cdot \mathrm{WEDFRI}_t$		0.0284	0.0297
$\mathrm{PMND1P}_t \cdot \mathrm{THUSAT}_t$		0.0185	0.0188
$\mathrm{S1P}_t \cdot \mathrm{TUESUN}_t$			0.0703
$\mathrm{S1P}_t \cdot \mathrm{WEDFRI}_t$			-0.2218
$\mathrm{S1P}_t \cdot \mathrm{THUSAT}_t$			-0.0413
$\mathrm{WC.IND}_t \cdot \mathrm{TUESUN}_t$			0.5248
$\mathrm{WC.IND}_t \cdot \mathrm{WEDFRI}_t$			-0.9426
$\mathrm{WC.IND}_t \cdot \mathrm{THUSAT}_t$			1.3245
cutpoints			
α_0	6.8128	7.3810	6.5040
α_1	7.0478	7.6272	6.7591
α_2	7.6310	8.2314	7.3874
α_3	8.6903	9.3101	8.5073
α_4	10.2024	10.8279	10.0509
residual deviance (df)	1106 (4753)	1083 (4748)	1056 (4742)
Brier score	.2545	.2467	.2405

variance 1. We call these standardized covariates $\boldsymbol{x}^s_{.i} = (x^s_{1i}, \ldots, x^s_{Ti})'$, where the components are given by

$$x^s_{ti} := \frac{x_{ti} - \bar{x}_{.i}}{\sqrt{\frac{1}{n}\sum_{t=1}^{T}(x_{ti} - \bar{x}_{.i})^2}} \tag{3.8}$$

with $\bar{x}_{.i} = \frac{1}{n}\sum_{t=1}^{T} x_{ti}$. Only indicator variables $\boldsymbol{x}_{.i}$ (where $x_{ti} \in \{0, 1\}$ for all $t \in \{1, \ldots, T\}$) are not standardized. The proportional odds model specifications from above were used as a starting point for the model specifications of the AOP models considered. If the 95% credible interval of a parameter contained zero, the corresponding covariate was removed from the model. In this way proportional odds model A and B lead to AOP model I and II, respectively. Table 3 shows the posterior mean estimates together with estimated 2.5% and 97.5% quantiles for all parameters based on 10000 iterations with a burnin of 1000 iterations. For Model I, the 95% credible interval for every main effect does not contain zero, so every covariate is significant. For Model II, the 95% credible intervals for PMND1P^s_t and WEDFRI_t contain the value 0. However, these two covariates must remain in the model since they appear in an interaction term which is itself significant.

4. RESULTS

Now we conduct a model comparison analysis for the five models investigated in Sections 3.2 and 3.3. First we consider the proportional odds models. To decide which of the proportional odds models fits the data best, we use the residual deviance test of Section 2.2.1. As mentioned there a model does not describe the data well, if

$$D > \chi^2_{T \cdot K - p, 1-\alpha}.$$

Table 3

Posterior mean and quantile estimates for standardized regression parameters and cutpoint parameters using the AOP model and their deviance information criterion and Brier score

		Model I			Model II	
	2.5%	mean	97.5%	2.5%	mean	97.5%
intercept	0.8817	1.2969	1.7624	1.0610	1.4764	1.9456
weather conditions						
HDXDD_t^s	-2.3685	-1.2880	-0.3530	-2.3874	-1.3548	-0.4171
$(\mathrm{HDXDD}_t^2)^s$	0.4096	1.1552	2.0173	0.4616	1.2054	2.0311
$\mathrm{S1P}_t^s$	-0.2368	-0.1322	-0.0314	-0.2688	-0.1619	-0.0569
$\mathrm{WC.IND}_t$	-1.6215	-0.8410	-0.1464	-1.6499	-0.8959	-0.2006
$\mathrm{PMND1P}_t^s$				-0.1331	-0.0172	0.0937
time of measurement						
$\mathrm{HAPM.IND}_t$	-0.9163	-0.5924	-0.2612	-0.9194	-0.5769	-0.2469
WEDFRI_t				-0.2672	-0.0079	0.2609
THUSAT_t				-0.5213	-0.2899	-0.0535
interactions						
$\mathrm{PMND1P}_t^s$						
$\times\mathrm{WEDFRI}_t$				0.0839	0.3077	0.5402
autoregressive parameter						
ϕ	0.7404	0.8077	0.8718	0.7250	0.7932	0.8541
cutpoints						
α_1	0.4706	0.7314	1.0221	0.4596	0.7383	1.1732
α_2	1.0821	1.3851	1.6962	1.1002	1.4021	1.8384
α_3	1.5870	1.8979	2.2151	1.6049	1.9250	2.3588
α_4	1.8548	2.1704	2.5127	1.8644	2.2013	2.6321
deviance information criterion						
	$\overline{D(\theta)}$	$D(\bar{\theta})$	DIC	$\overline{D(\theta)}$	$D(\bar{\theta})$	DIC
	799.6967	701.8899	897.5035	787.8536	695.2686	880.4387
Brier score						
		.1688			.1724	

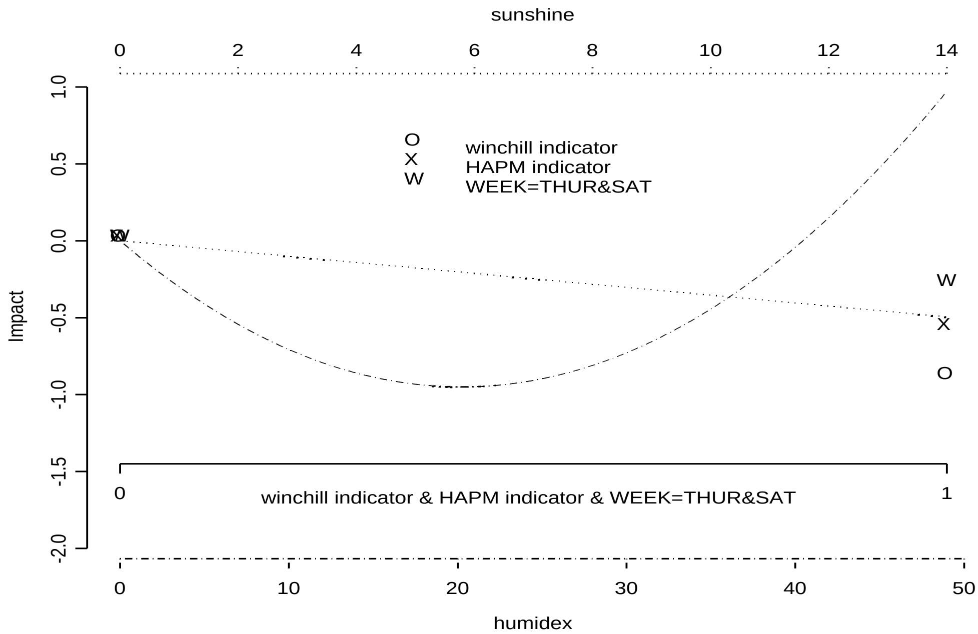
Here we have $T = 952$ and $K = 5$. We test on the 5% and 1% level and compute the p-value. Table 2 shows the results of the deviance analysis for the three models. For all three models the deviance D is not larger than the corresponding 99% quantiles of the χ^2-distribution, therefore all considered models fit the data quite well. Next we compare the AOP models using the DIC criterion. The values of the DIC for Model I and Model II are given in Table 3. Both the posterior mean of the deviance, $\overline{D(\boldsymbol{\theta})}$, and the deviance at the posterior mean, $D(\bar{\boldsymbol{\theta}})$, are smaller for Model II. From the values of the DIC we derive that, in spite of the penalizing term for the model complexity in the DIC, the more complex Model II fits better than the simpler Model I.

Finally we compare all proportional odds models and AOP models using the pseudo-predictions defined in Section 2.3. The corresponding Brier scores are given in Table 2 and 3, respectively. We conclude that the two AOP models describe the data better than all the proportional odds models. Although this model selection criterion prefers Model I to all other models, we are in favor of Model II, since the Brier scores on the one hand show clearly that the AOP models fit the data better than the proportional odds models, but on the other hand we prefer the DIC criterion for selection between the AOP models, and this criterion voted for Model II.

The signs of the regression parameters in Table 3 agree nearly everywhere with the signs in Table 2. This means that both the proportional odds models and the AOP models lead to the same conclusions, when asking which covariates have a high and which a low value to reduce the migraine severity. For example from the negative signs for S1P in all models we conclude that a sunny day increases the headache severity on the next day. This agrees with our conjecture from the exploratory analysis. The indicator for afternoon, HAPM.IND, also has a coefficient with negative sign. Again this approves our conjecture: The afternoon headache is usually worse than in the morning, at noon, and during the night. Considering the coefficients of the weekday indicators in Model II we see that the headache is worse between Wednesday and Saturday which might be a consequence of an (over)exertion on the job.

We now compute the impact of the covariate x_j on the conditional means of the latent variables in Model II defined as the product of the posterior mean estimate $\bar{\beta}_j$ and the value of x_j. Since we used the inverted order of categories for the AOP model, hence a smaller conditional mean increases the probability for a severe headache. The impact of the main effects is shown in the top panel of Figure 2, the impact of the interaction in the bottom panel of Figure 2. From the top panel of Figure 2 we see that the humidity index has the largest range and exhibits a quadratic influence. An humidity index around 20 has the lowest conditional mean for Y_t^* in the AOP model, whereas a higher or lower index increases the conditional mean. This means that a humidity index of 20 results in the most severe headaches, while deviations from 20 in both directions results in lower migraine severity. With regard to the interaction effect, we see that pressure differences have only on Wednesdays and Fridays an effect. On these days a high negative pressure difference gives the lowest conditional mean for Y_t^*.

We provide now a quantitative interpretation of the covariate impacts. For this we match the first two moments of the standard normal distribution



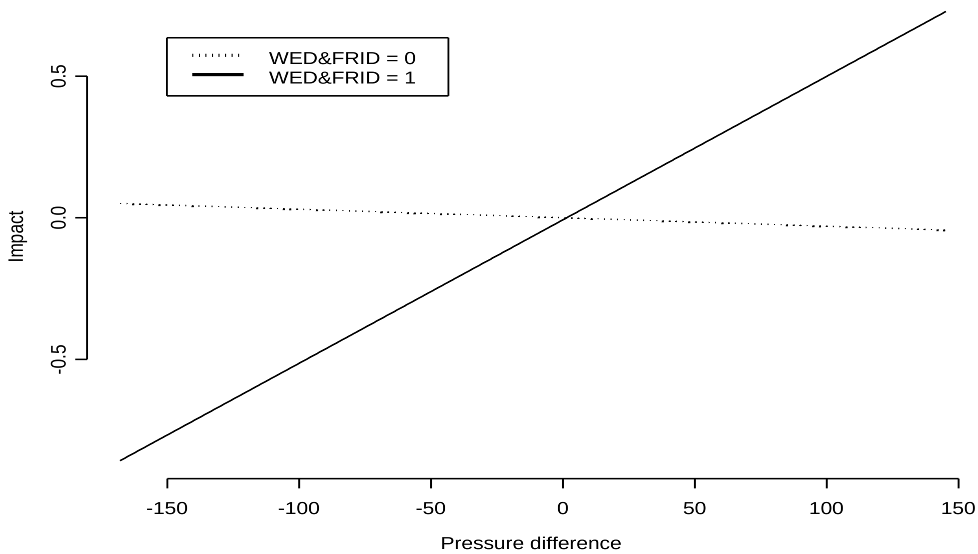


Figure 2. Impact of main effects (top panel, sunshine ...., humidex -.-.-.-) and interaction effect between pressure difference and weekday (bottom panel) of the AOP Model II.

to the logistic distribution to give the approximation

$$\Phi(z) \approx \frac{\exp\left(\frac{\pi}{\sqrt{3}}z\right)}{1+\exp\left(\frac{\pi}{\sqrt{3}}z\right)}.$$

For the AOP model it follows that the cumulative log odds ratio can approximated by

$$l_t(k) := \log\left(\frac{F_t(k)}{1-F_t(k)}\right) \approx \frac{\pi}{\sqrt{3}}(\alpha_k - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*),$$

where $F_t(k) := P(Y_t \leq k|\boldsymbol{x}_t, \boldsymbol{\alpha}, \boldsymbol{\beta}, \phi, Y_{t-1}^*) = \Phi(\alpha_k - \boldsymbol{x}_t'\boldsymbol{\beta} - \phi Y_{t-1}^*)$. Therefore the scaled impact

$$\frac{\pi}{\sqrt{3}}\beta_j x_j$$

approximates the effect on the cumulative log odds ratio. This allows us to quantify covariate effects. In particular, a change from 0 to 14 hrs of sunshine on the previous day changes the cumulative log odds ratio by $-.9$. The same change is seen if one changes humidity from 6 to 20 or from 34 to 20 index points. The presence of a Thurday or Saturday changes the cumulative log odds ratio by $-.45$, while the presence of windchill yields a change of -1.45. An afternoon measurement has a scaled impact change of -1.09. Finally, a pressure change from -1.5 kilopascal to 1.5 kilopascal on Wednesdays and Fridays changes the cumulative log odds ratio by 1.9.

Comparing different covariates, we see that a sunshine length change from 0 to 14 hrs has the same effect as a humidity index point change from 6 to 20 or 34 to 20. In comparison the effect of a Thursday or Saturday is only half the size, while the effect of an afternoon measurement is slightly higher and the windchill effect is 50 % higher than the sunshine change from zero to 14 hours. Finally a pressure change of -1.5 kilopascal to 1.5 kilopascal on Wednesdays and Fridays has a three times higher opposite effect.

Finally we note that the autoregressive component for the latent time series Y_t^* is around .8 indicating large positive dependency among the ordinal intensity measurements.

In summary we recommend to this patient to avoid long sunshine and windchill exposures. With regard to humidity, either a low or high humidity index has a more favorable influence than exposure to moderate humidity. Since about 70 % of observed humidity values are above 20 index points, the exposure to moderate humidity might not occur too often at her present residence location. Further negative pressure changes on the previous day cause problems especially on Wednesdays and Fridays. Changing work and life style arrangement on these days might help.

5. Conclusions

We applied the autoregressive ordererd probit (AOP) model suggested by Müller and Czado (2004) to an ordinal valued time series arising from headache intensity assessments. Here the ordered categories are produced by threshholding a latent real-valued time series with regression effects. To model the dependencies among the measurements the latent time series includes beside regression components also an autoregressive component. Parameter estimation is facilitated using a grouped move multigrid Monte Carlo (GM-MGMC) Gibbs sampler in a Bayesian setting. Models were compared using the DIC criterion suggested by Spiegelhalter et al. (2002) and the Brier score based on pseudo predictions.

For the migraine headache intensity data the latent time series shows a high first order autocorrelation of around .8 demonstrating considerable dependence among the ordinal measurements. For this patient we were able to demonstrate considerable impact of weather related variables such as humidity, windchill, sunshine length and pressure differences. In addition, time measurement effects were present. The final model specification includes nonlinear and interaction terms. Specific recommendations to this patient to lower the risk factors for severe migraine headaches have been provided. Even though an individual analysis offers the opportunity to develop more precise migraine control mechanisms, it is of interest to identify common risk factors in groups of patients. This problem is the subject of current research.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistical Analysis of Discrete Structures*. We would like to thank T. Kostecki-Dillon for providing the data.

References

Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley & Sons, 2 edition.

Albert, J. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **7**, 1–3.

Chen, M. and Dey, D. (2000). Bayesian analysis for correlated ordinal data models. In *Generalized linear Models - A Bayesian Perspective (eds Dey, D.K. and Ghosh, S.K. and Mallick, B.K.* New York: Marcel Dekker.

Cooke, L., Rose, M. and Becker, W. (2000). Chinook winds and migraine headache. *Neurology* **54(2)**, 302–307.

Fahrmeir, L., Gieger, C. and Hermann, C. (1999). An application of isotonic longitudinal marginal regression to monitoring the healing process. *Biometrics* **55**, 951–956.

Fahrmeir, L. and Pritscher, L. (1996). Regression analysis of forest damage by marginal models for correlated ordinal responses. *Journal of Environmental and Ecological Statistics* **3**, 257–268.

Gieger, C. (1997). Non- and semiparametric marginal regression models for ordinal response. *SFB Discussion Paper 71, Institut für Statistik, Ludwig-Maximilians-Universität, München, Germany* .

Gneiting, T. and Raftery, A. (2004). Strictly proper scoring rules, prediction and estimation. *Technical report No. 463, Department of Statistics, University of Washington, Seattle, U.S.A.* .

Harrell, F. (2001). *Regression Modeling Strategies.* New York: Springer Verlag.

Heagerty, P. and Zeger, S. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association* **91**, 809–822.

Kauermann, G. (2000). Modeling longitudinal data with ordinal response by varying coefficients. *Biometrics* **56**, 692–698.

Liang, K.-Y., Zeger, S. and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical*

Society B **54**, 2–24.

Lipsitz, S., Fitzmaurice, G. and Molenberghs, G. (1996). Goodness-of-fit test for ordinal response regression models. *Applied Statistics* **45**, 175–190.

Liu, J. and Sabatti, C. (1996). Generalised Gibbs sampler and multigrid monte carlo for bayesian computation. *Journal of Statistical computations and Simulation* **87**, 353–369.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society B* **42**, 109–142.

Müller, G. and Czado, C. (2004). An Autoregressive Ordered Probit Model with Application to High-Frequency Finance. *to appear in the Journal of Computational and Graphical Statistics (JCGS)* .

Nandram, B. and Chen, M. (1996). Accelerating Gibbs sampler convergence in generalized linear models via a reparametrization. *Journal of Statistical Computations and Simulation* **45**, 129–144.

Piorecky, J., Becker, W. and Rose, M. (1996). The effect of chinook winds on the probability of migraine headache occurence. *Headache* **37**, 153–158.

Prince, P., Rapoport, A., Sheftell, F., Tepper, S. and Bigal, M. (2004). The effect of weather on headache. *Headache* **44**, 153–158.

Schmitz, B. and Otto, J. (1984). Die Analyse von Migraine mit allgemeinpsychologischen Stresskonzepten. Zeitreihenanalysen für einen Einzelfall (in German). *Archiv der Psychologie* **136**, 211–234.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* **64**, 583–639.

Tutz, G. (2000). *Die Analyse Kategorialer Daten (in German).* München:

Oldenbourg.

Walker, S. and Duncan, D. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika* **54**, 167–178.

Wild, C. and Yee, T. (1996). Additive extensions to generalized estimating equation methods. *Journal of the Royal Statistical Society B* **58**, 711–725.

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