Gieger:

Non- and semiparametric marginal regression models for ordinal response

Sonderforschungsbereich 386, Paper 71 (1997)

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

Non- and semiparametric marginal regression models for ordinal response

Christian Gieger

Institut für Statistik, Ludwig-Maximilians-Universität München
Ludwigstr. 33, 80539 München, Germany
email: gieger@stat.uni-muenchen.de

Summary

We present a class of multivariate regression models for ordinal response variables in which the coefficients of the explanatory variables are allowed to vary as smooth functions of other variables. In the first part of the paper we consider a semiparametric cumulative regression model for a single ordinal outcome variable. A penalized maximum likelihood approach for estimating functions and parameters of interest is described. In the second part we explore a semiparametric marginal modeling framework appropriate for correlated ordinal responses. We model the marginal response probabilities and pairwise association structure by two semiparametric regressions. To estimate the model we derive an algorithm which is based on penalized generalized estimating equations. This nonparametric approach allows to estimate the marginal model without specifying the entire distribution of the correlated response. The methods are illustrated by two applications concerning the attitude toward smoking restrictions in the workplace and the state of damage in a Bavarian forest district.

Keywords: Ordinal response, marginal cumulative models, varying coefficients models, penalized likelihood, penalized generalized estimating equations, smoothing splines

1 Introduction

Recently several authors have proposed modeling approaches for multivariate correlated ordinal outcomes. The methods are multivariate extensions of models for univariate multicategorical responses. In this paper, we focus attention on cumulative regression models for ordinal responses (McCullagh, 1980) and multivariate extensions. Such models exploit, in a parsimonious way, the ordered scale of the outcomes. An important example for a cumulative model is the well–known proportional odds model. This and other ordinal response models have been discussed in detail by Fahrmeir and Tutz (1994, ch. 3).

First, we give a short review of models for correlated ordinal outcomes; these models are related to the approach used in this paper. The methods are similar in that they all use odds ratios to describe the association between responses. Dale (1986) proposed a model for bivariate ordinal data. She used marginal means and global odds ratios to specify the bivariate joint distribution. Molenberghs and Lesaffre (1994) extended her model for the case of three and more correlated outcomes using a multivariate Plackett–distribution. In this likelihood–based model computation of the joint distribution is very computer–intensive. Heagerty and Zeger (1996) considered a mixed parameter model. Starting with a loglinear representation of the likelihood they transformed the first and second order canonical parameters to marginal mean parameters and marginal global odds ratio parameters. They assumed a regression model for these marginal parameters and used very simple models for the higher order canonical parameters, e.g., setting them to zero. Heumann (1996) extended the full likelihood approach of Fitzmaurice and Laird (1993) based on conditional log odds ratios to the case of multicategorical response. Finally Fahrmeir and Pritscher (1996) developed multicategorical generalized estimating equations (GEE) by extending the generalized estimation equations approach (Liang and Zeger, 1986) to ordinal response data. They formulated two distinct parametric regressions for the marginal mean and the global odds ratios as a measure for the pairwise association structure. For binary data their model reduces to a GEE1 with odds ratio parameterization of the association structure (Lipsitz, Laird and Harrington, 1991).

In this paper, we extend the parametric marginal model of Fahrmeir and Pritscher (1996) to a more flexible semiparametric model. Semiparametric means that the effects of some or all covariates in both regressions are allowed to vary smoothly with other covariates. For example, we allow the effect of gender of a person to depend on the level of the covariate age of the person. What we get is a special kind of multiplicative interaction between these two covariates. For longitudinal data we can assume, for example, a time–stationary model for the pairwise association structure by modeling the global odds ratios as an unspecified function of a time–lag variable (see Example 2). This yields a very flexible modeling framework for both, the mean structure and the association structure of the marginal model.

Our model belongs to the general class of varying coefficients models

(Hastie and Tibshirani, 1993). For joint estimation of the marginal mean and the pairwise association structure we derive penalized generalized estimation equations (PGEE), which can be motivated as an extension of penalized estimation equations derived from a penalized likelihood approach to correlated data. By this PGEE we are enabled to estimate the functions and parameters of interest without specifying the entire distribution of the correlated response.

Wild and Yee (1996) presented an additive extension of generalized estimating equation methods for correlated binary data. To describe the mean and association structure they used a logit model and a model for the log odds ratios. In both regressions the influence of covariates is modeled nonparametrically by unspecified functions. Their approach for univariate outcomes is included as a special case of our multivariate semiparametric model.

Section 2 of this paper describes a semiparametric extension of the cumulative regression models. A roughness penalty approach together with a system of orthonormal spline base functions (Demmler and Reinsch, 1975) is used to estimate the unknown functions. In Section 3 the methodology is illustrated by an application concerning the attitude toward smoking restrictions in the workplace. The marginal semiparametric model for correlated ordinal outcomes is described in Section 4. In Section 5 the model is applied to forest damage data; in this study the ordered response "damage state" is measured repeatedly over time.

2 Univariate cumulative models for ordinal response

We consider the common situation of a cross–sectional regression analysis. In this setting, the response variable Y is ordinal with $q+1$ ordered categories. In addition we have a vector $x = (x_1, \ldots, x_p)'$ of p covariates. The observations (Y_i, x_i), $i = 1, \ldots, N$ are assumed to be independent. A univariate cumulative regression model relates the cumulative response probabilities $\mathrm{pr}(Y_i \leq r|x_i)$, $r = 1, \ldots, q$ to the covariates x_i by a smooth link function g of the form

$$g(\mathrm{pr}(Y_i \leq r|x_i)) = \beta_1 z_{i1} + \ldots + \beta_q z_{iq} + \beta_{q+1} z_{i,q+1} + \ldots + \beta_m z_{im}, \qquad (1)$$

where z_{ik}, $k = 1, \ldots, q$ are indicator functions with $z_{ik} = 1$ if $k = r$ and $z_{ik} = 0$ if $k \neq r$ (McCullagh, 1980). The remaining design variables, $z_{i,q+1}, \ldots, z_{im}$, are functions of the covariates, e.g., 0/1 variables for a categorical covariate. Instead of a single probability, as for example in logistic regression, we have q linearly independent probabilities. Note that the highest response category is excluded from considerations to yield a non–redundant set of parameters.

The cumulative model provides a parsimonious way of describing how the category probabilities change as the covariates vary. A very intuitive and useful way of motivating this model is to think of an underlying continuous response, say U, which is unobserved. What we observe, Y, is a categorization of U into $q + 1$ intervals. By this mechanism the density of U is divided into slices determined by the thresholds $\beta_1, \ldots, \beta_q$. The explanatory part,

$\beta_{q+1}z_{q+1} + \ldots + \beta_m z_m$, shifts the location of the underlying response U (Fahrmeir and Tutz, 1994, ch.3).

To further explain the model, we have to give some definitions. Letting $z_i = (z_{i1}, \ldots, z_{im})'$ and $\beta = (\beta_1, \ldots, \beta_m)'$, we can define a predictor $\eta_i^{(r)} = z_i'\beta$. As usual, the response Y_i is represented as a vector $y_i = (y_i^{(1)}, \ldots, y_i^{(r)}, \ldots, y_i^{(q)})'$ of q indicator variables with $y_i^{(r)} = 1$ if $Y_i = r$ and $y_i^{(r)} = 0$ if $Y_i \neq r$, $r = 1, \ldots, q$. The associated response probabilities, cumulative response probabilities, and predictors are $\pi_i = (\pi_i^{(1)}, \ldots, \pi_i^{(r)}, \ldots, \pi_i^{(q)})'$ with $\pi_i^{(r)} = \mathrm{pr}(Y_i = r|x_i) = \mathrm{pr}(y_i^{(r)} = 1|x_i)$, $\xi_i = (\xi_i^{(1)}, \ldots, \xi_i^{(r)}, \ldots, \xi_i^{(q)})'$ with $\xi_i^{(r)} = \mathrm{pr}(Y_i \leq r|x_i)$, and $\eta_i = (\eta_i^{(1)}, \ldots, \eta_i^{(r)}, \ldots, \eta_i^{(q)})'$. Finally we can write (1) for each subject i in the compact form

$$\mathrm{g}(\xi_i) = \eta_i = Z_i\beta,$$

where g is the q-dimensional version of the univariate link function g and Z_i is the design matrix defined by

$$Z_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & z_{i,q+1} \cdots z_{im} \\ 0 & 1 & & 0 & z_{i,q+1} \cdots z_{im} \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & 1 & z_{i,q+1} \cdots z_{im} \end{pmatrix}.$$

In the examples, we use a cumulative logit link, i.e.

$$\mathrm{logit}(\xi_i^{(r)}) = \log\left\{\frac{\xi_i^{(r)}}{1 - \xi_i^{(r)}}\right\} = \eta_i^{(r)}, \qquad r = 1, \ldots, q. \tag{2}$$

We obtain the response probabilities as

$$\pi_i^{(1)}(\eta_i) = \frac{\exp(\eta_i^{(1)})}{1 + \exp(\eta_i^{(1)})}$$

and

$$\pi_i^{(r)}(\eta_i) = \frac{\exp(\eta_i^{(r)})}{1 + \exp(\eta_i^{(r)})} - \frac{\exp(\eta_i^{(r-1)})}{1 + \exp(\eta_i^{(r-1)})}, \qquad r = 2, \ldots, q,$$

from (2). The model parameters β are estimated by maximizing the log–likelihood.

In many situations such a model with fixed parameters seems to be too restrictive. One possibility to increase the flexibility of the fixed parameter model is to allow the coefficients of (1) to vary with the values of other covariates, say $v_1, \ldots, v_m$. We obtain a cumulative model with varying coefficients of the form

$$\begin{aligned} g(\mathrm{pr}(Y_i \leq r|x_i)) &= \gamma_1(v_{i1})z_{i1} + \ldots + \gamma_q(v_{iq})z_{iq} \\ &\quad + \gamma_{q+1}(v_{i,q+1})z_{i,q+1} + \ldots + \gamma_m(v_{im})z_{im}, \end{aligned} \tag{3}$$

where $\gamma_1(\cdot), \ldots, \gamma_m(\cdot)$ are sufficiently smooth real functions and the covariates $v_1, \ldots, v_m$ are assumed to be continuous. To keep the model interpretable we require the threshold functions $\gamma_1(\cdot), \ldots, \gamma_q(\cdot)$ to vary with the same variable, i.e. $v_{i1} = v_{i2} = \ldots = v_{iq}$. This yields a very flexible model, which allows us to modify the effects of covariates with the values of other covariates.

To estimate the unknown functions we choose a penalized likelihood approach and maximize the log–likelihood with integrated quadratic roughness penalty terms

$$lp(\gamma_1, \ldots, \gamma_m) = \sum_{i=1}^{N} l_i(\gamma_1(v_{i1}), \ldots, \gamma_m(v_{im})) - \frac{1}{2}\sum_{j=1}^{m} \lambda_j \int (\gamma_j''(v))^2 dv \quad (4)$$

over all twice continuously differentiable functions. In the penalized likelihood criterion, Equation (4), $l_i(\cdot)$ denotes the multinomial log–likelihood of subject i in terms of the functions $\gamma_1(v_{i1}), \ldots, \gamma_m(v_{im})$. The smoothing parameters $\lambda = (\lambda_1, \ldots, \lambda_m)'$ control the trade off between goodness of fit and roughness of the estimated functions. In this paper, we assume that all smoothing parameters are known and fixed.

Using the well–known fact that the maximizing functions of (4) are natural cubic splines with knots at the unique values of v_{ij}, we represent (3) in terms of basis functions for these spaces. Choosing a system of orthonormal basis functions $\{\phi_{j1}(\cdot), \ldots, \phi_{jn_j}(\cdot)\}$ introduced by Demmler and Reinsch (1975), we write $\gamma_j(v) = \sum_{k=1}^{n_j} \beta_{jk}\phi_{jk}(v)$, where n_j is the dimension of the finite dimensional spline space.

This yields a representation of (3) in terms of evaluated basis functions

$$\begin{aligned} g(\mathrm{pr}(Y_i \leq k|x_i)) &= \beta_{11}\phi_{11}(v_{i1})z_{i1} + \ldots + \beta_{1n_1}\phi_{1n_1}(v_{i1})z_{i1} \\ + \ldots &+ \beta_{q1}\phi_{q1}(v_{iq})z_{iq} + \ldots + \beta_{qn_q}\phi_{qn_q}(v_{iq})z_{iq} \\ + \ldots &+ \beta_{m1}\phi_{m1}(v_{im})z_{im} + \ldots + \beta_{mn_m}\phi_{mn_m}(v_{im})z_{im}. \quad (5) \end{aligned}$$

For simplification we define the $\sum_{j=1}^{m} n_j$–dimensional vector of basis coefficients

$$\beta = (\beta_{11}, \ldots, \beta_{1n_1}, \ldots, \beta_{q1}, \ldots, \beta_{qn_q}, \beta_{q+1,1}, \ldots, \beta_{mn_m})'$$

and the $q \times \sum_{k=1}^{m} n_k$ design matrix

$$Z_i = \begin{pmatrix} \phi_{i11} & \cdots & \phi_{i1n_1} & 0 & \ldots\ldots\ldots & 0 & \phi_{iq+1,1}z_{iq+1} & \cdots & \phi_{imn_m}z_{im} \\ & & & \ddots & & & \vdots & & \vdots \\ 0 & \ldots\ldots\ldots & 0 & \phi_{iq1} & \cdots & \phi_{iqn_q} & \phi_{iq+1,1}z_{iq+1} & \cdots & \phi_{imn_m}z_{im} \end{pmatrix}$$

with $\phi_{ijk} = \phi_{jk}(v_{ik})$. With this notation, we write

$$\mathrm{g}(\xi_i) = \eta_i = Z_i\beta,$$

exactly as in the case of fixed parameters.

The Demmler–Reinsch basis $\{\phi_{j1}(\cdot), \ldots, \phi_{jn_j}(\cdot)\}$ satisfies

$$\int \phi_{jk}''(v)\phi_{jl}''(v)dv = \delta_{kl}\rho_{jk},$$

where $k, l = 1, \ldots, n_j$ and $\delta_{kl} = I(k = l)$. The ρ_{jk}'s can be computed by solving an eigenvalue problem (see Eubank, 1988, ch.5).

Using (2), the infinite dimensional function estimation problem (4) reduces to the estimation of the basis coefficients β. We obtain the finite dimensional criterion

$$\hat{\beta} = \arg\max_{\beta}(lp(\beta)) = \arg\max_{\beta}\left(\sum_{i=1}^{N} l_i(\beta) - \frac{1}{2}\beta'\Lambda\mathrm{P}\beta\right), \tag{6}$$

where $\mathrm{P} = \mathrm{diag}(\rho_{jk})_{j=1,\ldots,m,k=1,\ldots,n_j}$ is a diagonal penalty matrix and Λ is a diagonal matrix of smoothing parameters.

Setting the derivative of (6) with respect to β to zero gives the penalized estimation equations

$$u(\beta) = \sum_{i=1}^{N} Z_i' D_i(\beta)\Sigma_i^{-1}(\beta)(y_i - \pi_i(\beta)) - \Lambda\mathrm{P}\beta = 0, \tag{7}$$

where $D_i(\beta) = \partial \mathrm{h}(\eta)/\partial\eta$ is the derivative of the response function, $\mathrm{h}(\eta) = \mathrm{g}^{-1}(\eta)$ at $\eta_i = Z_i\beta$, and $\Sigma_i(\beta) = \mathrm{diag}(\pi_i(\beta)) - \pi_i(\beta)\pi_i'(\beta)$ denotes the covariance matrix of observation y_i given β. The expected negative second derivative $H(\beta)$ of the penalized log–likelihood $lp(\beta)$ is given by

$$H(\beta) = E\left(-\frac{\partial^2 lp(\beta)}{\partial\beta\partial\beta'}\right) = \sum_{i=1}^{N} Z_i' W_i(\beta^{(k)})Z_i + \Lambda\mathrm{P}$$

with a weight matrix $W_i(\beta) = D_i(\beta)\Sigma_i^{-1}(\beta)D_i'(\beta)$.

The penalized estimation equations (7) are solved iteratively. In a (quasi-) Fisher scoring step the update is determined by

$$H(\beta^{(k)})(\beta^{(k+1)} - \beta^{(k)}) = u(\beta^{(k)}),$$

where $\beta^{(k)}$ is the result of the current step and $\beta^{(k+1)}$ denotes the next parameter vector. If we define a working observation vector $\tilde{y}_i(\beta)$ by

$$\tilde{y}_i(\beta) = Z_i\beta + (D_i^{-1}(\beta))'(y_i - \pi_i(\beta)),$$

we may equivalently express the (quasi-) Fisher scoring iterations in the form

$$\left(\sum_{i=1}^{N} Z_i' W_i(\beta^{(k)})Z_i + \Lambda\mathrm{P}\right)\beta^{(k+1)} = \sum_{i=1}^{N} Z_i' W_i(\beta^{(k)})\tilde{y}_i(\beta^{(k)}).$$

Iterations are stopped according to a termination criterion.

Up to now we have assumed that each coefficient in (3) must vary with a covariate. Of course, in general, this will not be the case. In many applications we want to allow only some of the coefficients in (1) to vary with covariates, while others stay fixed. To do so we have to modify our design matrix Z_i and our penalty matrix P as follows. In Z_i we replace the columns of multiplications with the basis functions by the design vector z_{ij} itself. In the penalty matrix P the diagonal element, which now corresponds to the fixed effect β_j, is set to zero. We get a semiparametric predictor and refer to our model as a semiparametric cumulative model. In the special case of no varying coefficients we reduce to a parametric model and our estimation procedure is ordinary Fisher scoring or iteratively weighted least squares.

It follows from the results of Hastie and Tibshirani (1993) and Wahba (1990) that a unique solution of (7) exists if the corresponding embedded parametric model, where we have restricted each function $\gamma_j(\cdot)$ to be linear, has an unique solution. This means that the design matrix $Z^{(par)} = (Z_1^{(par)}, \dots, Z_i^{(par)}, \dots, Z_N^{(par)})'$ with

$$Z_i^{(par)} = \begin{pmatrix} 1 & v_{i1} & 0 & \dots & & 0 & z_{iq+1} & v_{iq+1}z_{iq+1} & \dots & z_{im} & v_{im}z_{im} \\ 0 & & & & & & & & & & \\ & \ddots & \ddots & \ddots & \ddots & & \vdots & \vdots & & \vdots & \vdots \\ & & & & & 0 & & & & & \\ 0 & & \dots & 0 & 1 & v_{iq} & z_{iq+1} & v_{iq+1}z_{iq+1} & \dots & z_{im} & v_{im}z_{im} \end{pmatrix}$$

must be of full rank. Often this will not be the case. For example, if $z_{ij} = 1$, then the jth term in (3) is simply an unspecified smooth function $\gamma(v_{ij})$ in v_{ij} and clearly the corresponding parametric design matrix $Z^{(par)}$ does not have full rank. But we can solve this problem in an elegant way by reducing the system of basis functions. Here we use the fact that the system of basis functions can be divided in two spline function spaces which span up the linear part and the nonlinear part of the spline. Due to this split–up, we can now reduce the linear parts of the splines until we get a unique solution. This means in our example that we have to delete the redundant intercept term.

A rigorous asymptotic theory for models using cubic splines as smoothing method is still not available. We apply in a heuristic way the asymptotic theory of maximum likelihood estimation in misspecified generalized linear models to the penalized likelihood case (Fahrmeir, 1990). In analogy to Fahrmeir and Klinger (1996) we use

$$V(\hat{\beta}) = H(\hat{\beta})^{-1} F(\hat{\beta}) H(\hat{\beta})^{-1}, \qquad (8)$$

with $F(\hat{\beta}) = \sum_{i=1}^{N} Z_i' D_i(\hat{\beta}) \Sigma_i^{-1}(\hat{\beta}) D_i'(\hat{\beta}) Z_i$ as an approximation to the covariance matrix of the estimate $\hat{\beta}$. Among others the quality of the approximation depends on the ratio of sample size to parameters involved in criterion (6). In practice this seems to be a critical point, because in general the number of basis functions and with it the number of parameters grows with the number of distinct observations of a continuous covariate. But nevertheless we use (8) as a useful approximation. The estimated variances of $\hat{\beta}$ are on the diagonal of $V(\hat{\beta})$, we can use them to reduce the dimension of the problem considerably by performing a (heuristic) test procedure and select the "significant" basis functions. Pointwise confidence bands for the estimated functions themselves can be computed from the diagonal elements of the matrix $Z'V(\hat{\beta})Z$ with $Z = (Z_1, \dots, Z_i, \dots, Z_N)'$.

3 Example 1: Attitudes toward smoking restrictions in the workplace

To illustrate the methodology we use data from a study which was examined with the impact of a bylaw regulating smoking in the workplace. This bylaw was implemented in the city of Toronto in March 1988. A detailed description of the study together with a comprehensive analysis using a multicategorical logit model is given by Bull (1994). Using the ordered structure of the response variable "attitude toward smoking restrictions" we re–analyze the data set with a cumulative logit model. As data for the study two surveys were conducted; one immediately before implementation of the bylaw and a second, eight to nine months later. Sampling was carried out independently for each survey. Due to this design, it is possible to determine the impact of the bylaw on the attitudes of the residents.

The city of Toronto is geographically surrounded by several jurisdictions; these jurisdictions were not subject to the bylaw. However, many residents of this area were affected because their workplaces were in the city. For this reason, persons from the whole aera were included in the survey. So the survey covers persons who worked in the city of Toronto, as well as persons who worked outside the city, and finally persons who do not work outside the home. A total of 2855 residents participated in the study.

The outcome variable , say Y, "attitude toward smoking restrictions" has the following three categories: smoking in the workplace should not be permitted at all ($Y = 3 =$ reference), smoking should be permitted in restricted areas ($Y = 2$), and smoking should not be restricted at all ($Y = 1$). We use the cumulative logit model

$$\operatorname{logit}(\operatorname{pr}(Y \leq r)) = \eta_r \qquad r = 1, 2$$

to clarify how the attitude depends on the following covariates:

TS Time of the survey (post–implementation = 1, pre–implementation = reference)

PW Place of work (outside the city of Toronto and outside the home = 1, at home = 2, and in the city and outside the home = reference)

S Smoking status (current smoker = 1, former smoker = 2, and never smoked = reference)

K Knowledge of health effects of environmental tobacco smoke (score with a range from -6 to 6)

G Gender of the person (male = 1, female = reference)

A Age of the person in years (with a range from 18 to 85 years)

First, we estimate a model with fixed parameters of the form

$$\eta_r = \beta_r + \beta_3 TS + \beta_4 PW^{(1)} + \beta_5 PW^{(2)} + \beta_6 S^{(1)} + \beta_7 S^{(2)} + \beta_8 K + \beta_9 G + \beta_{10} A^c,$$

where $PW^{(1)}$, $PW^{(2)}$, $S^{(1)}$, $S^{(2)}$ represent the dummy variables for the categorical covariates PW and S. The variable A^c is the centered age with $A^c = (A - 50)/10$.

We compute the following maximum likelihood estimates:

Covariate	Estimate	SE	p–value
$TH^{(1)}$	-3.5558	0.1436	0.0000
$TH^{(2)}$	0.8451	0.1123	0.0000
TS	-0.0803	0.0826	0.3308
$PW^{(1)}$	0.2172	0.0975	0.0260
$PW^{(2)}$	0.0181	0.1137	0.8734
$S^{(1)}$	1.2706	0.1134	0.0000
$S^{(2)}$	0.2675	0.1049	0.0108
K	-0.1492	0.0177	0.0000
G	0.1524	0.0859	0.0759
A^c	-0.0918	0.0286	0.0013

We summarize the results as follows. The individual predicted probabilities for a fifty years old non–smoking woman, having a medium knowledge of health effects, and working in the city before the bylaw implementation, i.e. all design variables have the value zero, are 0.03 for "unrestricted", 0.67 for "restricted", and 0.30 for "prohibited smoking". The changes associated with the bylaw implementation (TS) are very small and not significant. Workers outside the city ($PW^{(1)}$) tend to prefer lower response categories, i.e. the predicted probabilities for "unrestricted" and "restricted" are higher compared to city workers. In contrast, the not–outside–home workers ($PW^{(2)}$) behave like the city workers. As expected smoking status is a relevant determinant of the attitude. Especially current smokers ($S^{(1)}$) prefer no restrictions. There is a strong, positive association of having good knowledge of health effects with support for prohibition, i.e. those with higher scores are more likely to prefer prohibition. Men are somewhat more likely than women to prefer unrestricted and restricted smoking. The negative value of the parameter belonging to age indicates that increasing age yields lower probabilities for the categories "unrestricted" and "restricted", i.e. preference for prohibition increases with age.

In a second analysis, we examine the data with a more flexible model. We try to check if the influence of the covariate age on the attitude is really linear as assumed in the first model. We also investigate if the effect of gender is constant over age of the person. For this reason, we fit a semiparametric cumulative logit model excluding the covariate time of survey. The predictor of the model is now:

$$\eta_r = \gamma_r(A) + \gamma_3 PW^{(1)} + \gamma_4 PW^{(2)} + \gamma_5 S^{(1)} + \gamma_6 S^{(2)} + \gamma_7 K + \gamma_8(A)G,$$

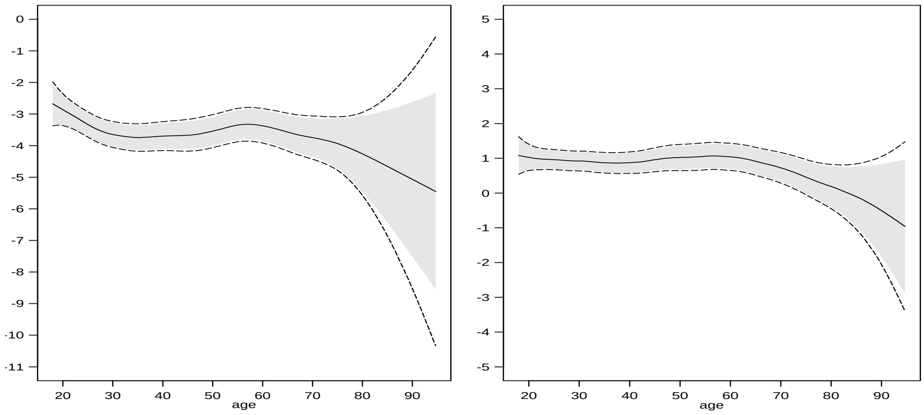

Figure 1: Estimated threshold functions $\hat{\gamma}_1$ (left) and $\hat{\gamma}_2$ (right) with pointwise 2×standard error bands (model based - dashed line, robust - shaded region)

i.e. we allow the threshold functions and the effect of gender to change with age of the person.

Figure 1 shows the estimated threshold functions $\hat{\gamma}_1(A)$ and $\hat{\gamma}_2(A)$. The first threshold is decreasing across age up to 35 years showing that increasing age yields a lower probability for the category "unrestricted". The second threshold is almost constant in this period this means that the probability is shifted to the middle category "restricted". Afterwards the probabilities do not change too much up to about 65 years. The shift to the category "prohibited" for older people must be interpreted with some caution because the data are very sparse. This interpretation applies to women; for men we have to consider the effect of gender. Figure 2 shows that especially younger men prefer the lower categories "restricted" and "unrestricted". The table gives the remaining fixed effects

Covariate	Estimate	SE	SE	p–value	p–value
		(model)	(robust)	(model)	(robust)
$PW^{(1)}$	0.2100	0.0982	0.0981	0.0322	0.0322
$PW^{(2)}$	0.0087	0.1195	0.1192	0.9418	0.9417
$S^{(1)}$	1.2709	0.1136	0.1135	0.0000	0.0000
$S^{(2)}$	0.2722	0.1069	0.1069	0.0109	0.0108
K	-0.1557	0.0180	0.0180	0.0000	0.0000

The values are almost unchanged compared to the first model and thus the interpretation remains the same.

With the semiparametric model it is possible to get a deeper insight into the structure of the dependence on the covariates. By allowing the effect of

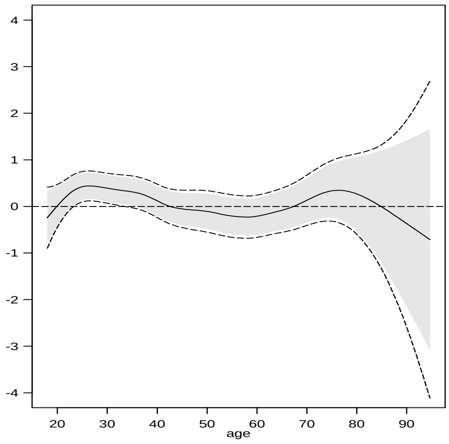

Figure 2: Estimated effect of the male category and pointwise 2× standard error bands (model based - dashed line, robust - shaded region)

gender to vary smoothly with age of the person we consider the interaction between the continuous covariate age and the 0/1 covariate gender in a natural way. By allowing age to modify each threshold function separately we drop the sometimes too restrictive proportional odds assumption.

4 Marginal cumulative models for correlated ordinal response

Suppose that a study has been conducted with N subjects as primary units. For the ith subject T_i ordinal responses Y_{it} with $q+1$ categories together with covariates x_{it} are observed. For simplicity, we assume a longitudinal data situation where we have $T_i = T$ repeated measurements at times $t = 1, \ldots, T$. The discrete or continuous covariates are either time-constant, for example, individual characteristics, or time-varying, such as time. To avoid complications we assume that the time-varying covariates are either non-stochastic or stochastic but external, i.e. their values are not influenced by outcomes of the responses Y_{it}. Thus the data are given by (Y_{it}, x_{it}), $i = 1, \ldots, N$, $t = 1, \ldots, T$.

As in the application of this paper, the influence of covariates on the marginal probabilities of the response categories is often of prime interest, whereas association is regarded as nuisance. Marginal regression models permit separate modeling of the marginal means and the association among repeated observations of the response for each subject. To estimate the marginal mean and association parameters without specifying the entire likelihood of the multivariate response we will derive a penalized estimation equations approach that allows the estimation of the mean structure even under misspecification of the association structure.

In the first step we specify a model for the influence of the covariates on the marginal probabilities of the response categories. To utilize the ordinal scale of Y_{it} we again adopt a cumulative model. It relates the marginal cumulative probabilities of Y_{it} to the predictor $\eta_{it} = (\eta_{it}^{(1)}, \ldots, \eta_{it}^{(r)}, \ldots, \eta_{it}^{(q)})'$ in the form

$$g\left\{\mathrm{pr}(Y_{it} \leq r|x_{it})\right\} = \eta_{it}^{(r)}, \qquad r = 1, \ldots, q,$$

with a known link function g, e.g, the logit link function. For a complete determination of the marginal mean model we have to specify the functional form of the predictor $\eta_{it} = (\eta_{it}^{(1)}, \ldots, \eta_{it}^{(r)}, \ldots, \eta_{it}^{(q)})'$. As in the cross–sectional case we allow the effects of the design variables to vary with the values of other variables and obtain the predictor

$$\eta_{it}^{(r)} = \gamma_r(v_{itr}) + \gamma_{q+1}(v_{itq+1})z_{itq+1} + \ldots \quad + \quad \gamma_m(v_{itm})z_{itm}, \qquad r = 1, \ldots, q, \qquad (9)$$

with real–valued smooth functions $\gamma_1(\cdot), \ldots, \gamma_m(\cdot)$.

It is worthwhile to look at special cases of the model. If we put the restriction on a function to be the constant function, i.e. $\gamma_k(v_{itk}) = \gamma_k$, then that term is linear in z_{itk} and we get a marginal semiparametric model. If all terms are linear, then (9) reduces to the predictor of a marginal parametric cumulative model (Fahrmeir and Pritscher, 1996). For $v_{it1} = \ldots = v_{itm} = t =$ time we can regard the model as a dynamic marginal regression model with parameters changing smoothly with time. With appropriate specifications of the predictor (9), we get several useful models. For example, we can model the effect of a time–constant covariate separately for each time by introducing T unspecified functions $\gamma_t(\cdot)$, $t = 1, \ldots, T$ of this covariate and setting the corresponding design variables either to 1 or to 0 depending on the actual observed time. In a similar way we can form predictors with category–specific covariate effects.

To simplify the notation, we define again some vectors. As before the response Y_{it} is represented as a vector $y_{it} = (y_{it}^{(1)}, \ldots, y_{it}^{(r)}, \ldots, y_{it}^{(q)})'$ of q dummy variables. For the vectors of marginal response probabilities and marginal cumulative response probabilities we get $\pi_{it} = (\pi_{it}^{(1)}, \ldots, \pi_{it}^{(r)}, \ldots, \pi_{it}^{(q)})'$ with $\pi_{it}^{(r)} = \mathrm{pr}(Y_{it} = r|x_{it}) = \mathrm{pr}(y_{it}^{(r)} = 1|x_{it})$, and $\xi_{it} = (\xi_{it}^{(1)}, \ldots, \xi_{it}^{(r)}, \ldots, \xi_{it}^{(q)})'$ with $\xi_{it}^{(r)} = \mathrm{pr}(Y_{it} \leq r|x_{it})$. The time–specific vectors are combined to vectors of responses $y_i = (y_{i1}', \ldots, y_{iT}')'$, response probabilities $\pi_i = (\pi_{i1}', \ldots, \pi_{iT}')'$, predictors $\eta_i = (\eta_{i1}', \ldots, \eta_{iT}')'$ and covariates $x_i = (x_{i1}', \ldots, x_{iT}')'$ of subject i.

As in section 2, we use cubic smoothing splines to estimate the unknown smooth functions. This means that we can construct a parameter vector β and a design matrix $Z_i = (Z_{i1}, \ldots, Z_{iT})'$, where the time–specific matrices Z_{it} are formed from the design variables and the spline basis functions. This yields again a multivariate predictor of the form

$$\eta_i = Z_i\beta.$$

In addition to the model for the effects of covariates on the marginal probabilities, we also have to specify a model for the association among observations from each subject. Here we only consider the association between two outcomes and ignore the higher order associations.

Often the analysis is based on the working assumptions of independence. Letting $\Sigma_{it} = \text{diag}(\pi_{it}(\beta)) - \pi_{it}(\beta)\pi'_{it}(\beta)$ denote the covariance matrix of the response y_{it}, we we obtain by setting

$$V_i = \text{blockdiag}(\Sigma_{i1}, \ldots, \Sigma_{iT})$$

the simplest model for the covariance matrix of the marginal model.

With this assumption we get penalized generalized estimating equations for the estimation of the parameters β. These penalized generalized estimating equations are identical to estimating equations obtained from a penalized likelihood criterion for independent observations.

Instead of using the independence model, we can supplement the marginal mean model by a model for the pairwise association which determines the covariance of the marginal model. A common measure for the pairwise association of two ordinal responses Y_{it} and Y_{is} of the same subject is the global odds–ratios (Dale, 1986, Fahrmeir and Pritscher, 1996). For each pair of categories l and r of Y_{it} and Y_{is} the global odds–ratio at cutpoint (l, r) is given by

$$\psi^{(lr)}_{i,st} = \frac{\text{pr}(Y_{is} \le l, Y_{it} \le r|x_i)\text{pr}(Y_{is} > l, Y_{it} > r|x_i)}{\text{pr}(Y_{is} > l, Y_{it} \le r|x_i)\text{pr}(Y_{is} \le l, Y_{it} > r|x_i)}, \quad l, r = 1, \ldots, q. \quad (10)$$

This means that the $(q + 1) \times (q + 1)$ contingency table of probabilities $\pi^{(lr)}_{i,st} = \text{pr}(Y_{it} = l, Y_{is} = r)$ is collapsed at cutpoint (l, r) to a 2×2 table and a usual odds ratio is computed with this coarser table. Furthermore, by solving (10) we can express the bivariate cumulative probability function $\xi^{(lr)}_{i,st} = \text{pr}(Y_{is} \le l, Y_{it} \le r|x_i)$ of Y_{is} and Y_{it} in terms of the corresponding global odds–ratio $\psi^{(lr)}_{i,st}$ and the marginal cumulative probabilities $\xi^{(l)}_{is}$ and $\xi^{(r)}_{it}$, yielding

$$\xi^{(lr)}_{i,st} = \begin{cases} \xi^{(l)}_{is}\xi^{(r)}_{it} & \text{,if} \quad \psi^{(lr)}_{i,st} = 1, \\ \dfrac{\kappa - \sqrt{\kappa^2 + 4\psi^{(lr)}_{i,st}(1 - \psi^{(lr)}_{i,st})\xi^{(l)}_{is}\xi^{(r)}_{it}}}{2(\psi^{(lr)}_{i,st} - 1)} & \text{,if} \quad \psi^{(lr)}_{i,st} \ne 1, \end{cases} \quad (11)$$

where $\kappa = 1 + (\xi^{(l)}_{is} + \xi^{(r)}_{it})(\psi^{(lr)}_{i,st} - 1)$. With this formula we can calculate the second–order moments $\pi^{(lr)}_{i,st} = E(y^{(l)}_{is} y^{(r)}_{it}) = \text{pr}(Y_{is} = l, Y_{it} = r)$ in terms of the univariate marginal cumulative probabilities $\xi^{(l)}_{is}$, $\xi^{(r)}_{it}$ and also the global odds ratios $\psi^{(lr)}_{i,st}$ through the relation

$$\pi^{(lr)}_{i,st} = \begin{cases} \xi^{(lr)}_{i,st} & ,l = r = 1 \\ \xi^{(lr)}_{i,st} - \xi^{(l,r-1)}_{i,st} & ,l = 1, r > 1 \\ \xi^{(lr)}_{i,st} - \xi^{(l-1,r)}_{i,st} & ,l > 1, r = 1 \\ \xi^{(lr)}_{i,st} - \xi^{(l,r-1)}_{i,st} - \xi^{(l-1,r)}_{i,st} + \xi^{(l-1,r-1)}_{i,st} & ,l > 1, r > 1. \end{cases} \quad (12)$$

Now the off–diagonal elements of the model covariance matrix V_i are determined by $\mathrm{cov}(y_{is}^{(l)}, y_{it}^{(r)}) = E(y_{is}^{(l)} y_{it}^{(r)}) - \pi_{is}^{(l)}\pi_{it}^{(r)} = \pi_{i,st}^{(lr)} - \pi_{is}^{(l)}\pi_{it}^{(r)}$.

We use a logarithmic model for the vector of global odds ratios $\psi_i = (\ldots, \psi_{i,st}^{(lr)}, \ldots)'$, $l, m = 1, \ldots, q$, $s < t = 1, \ldots, t$ of subject i and get

$$\log(\psi_i) = \tilde{Z}_i \alpha. \qquad (13)$$

The design matrix $\tilde{Z}$ and the vector of association parameters α result either from a parametric predictor of the association structure, e.g.,

$$\log(\psi_{i,st}^{(lr)}) = \alpha_{lr} + \frac{\alpha}{|t-s|}$$

or from a semiparametric predictor, e.g.,

$$\log(\psi_{i,st}^{(lr)}) = \alpha_{lr} + \gamma(|t-s|),$$

where $\gamma(\cdot)$ is a smooth function of the time lag. Of course more complex predictors, possibly including covariates, $z_{i,st} = z(x_{is}, x_{it})$, are possible and in some situations useful.

To estimate the marginal model, we propose using penalized generalized estimating equations (PGEE1) by extending the estimating equations (7) derived from the penalized maximum likelihood criterion (4) to the case of correlated data. This transfer can be justified by the same arguments as in the parametric case (Laird, 1996). We estimate β and thereby the unknown spline functions by solving the multivariate penalized generalized estimating equation

$$\sum_{i=1}^{N} Z_i' D_i V_i^{-1}(y_i - \pi_i) - \Lambda \mathrm{P} \beta = 0, \qquad (14)$$

where D_i =blockdiag$(\partial \pi_{it}/\partial \eta_{it})$. The matrix Λ contains the smoothing parameters and P is the penalty matrix. Both are defined as in Section 2. Equation (14) is a multivariate version of equation (7), except that the 'true' covariance matrix of the multivariate response y_i is replaced by a 'working' covariance matrix V_i. This covariance matrix V_i is determined by the marginal response probabilities and the pairwise associations and thus by α and β. The other terms in (14) are not influenced by the association model.

To estimate the association parameter α jointly with β we augment (14) by a second PGEE

$$\sum_{i=1}^{N} \tilde{Z}_i' C_i U_i^{-1}(w_i - \nu_i) - \Delta \Omega \alpha = 0, \qquad (15)$$

where Δ is a second matrix of smoothing parameters and Ω is the penalty matrix corresponding to the association model. In (15) $w_i = (\ldots, w_{i,st}^{(lr)}, \ldots)'$ contains the centered products $w_{i,st}^{(lr)} = (y_{is}^{(l)} - \pi_{is}^{(l)})(y_{it}^{(r)} - \pi_{it}^{(r)})$, $l, m = 1, \ldots, q$, $s < t = 1, \ldots, t$ and $\nu_i = \nu_i(\alpha, \beta) = (\ldots, \nu_{i,st}^{(lr)}, \ldots)'$ is the vector of expectations

$$\nu_{i,st}^{(lr)} = E(w_{i,st}^{(lr)}) = E(y_{is}^{(l)} y_{it}^{(r)}) - \pi_{is}^{(l)}\pi_{it}^{(r)} \qquad (16)$$

for observations $(y_{is}^{(l)}, y_{it}^{(r)})$ of subject i. The matrix C_i is the Jacobian, obtained from inserting (11) and (12) in (16) and differentiating with respect to the predictor in (13). The matrix U_i is a further working covariance matrix, now for the 'observations' w_i. For the PGEE1 approach the following two simple diagonal specifications are useful. As in the binary case (Prentice, 1988) the simplest choice is the identity matrix specification

$$U_i = I.$$

Another choice is

$$U_i = \mathrm{diag}(\mathrm{var}(w_{i,st}^{(lr)}))_{l,r=1,\ldots,q,s<t=1,\ldots,T},$$

where

$$\mathrm{var}(w_{i,st}^{(lr)}) = \pi_{is}^{(l)}(1-\pi_{is}^{(l)})\pi_{it}^{(r)}(1-\pi_{it}^{(r)}) - (\nu_{i,st}^{(lr)})^2 + \nu_{i,st}^{(lr)}(1-2\pi_{is}^{(l)})(1-2\pi_{it}^{(r)}).$$

Note that the PGEE1 for α reduces to an ordinary GEE1 if we use a parametric model for the pairwise association structure and of course the second PGEE is not necessary if we use the independence assumption.

We can compute the estimates of the parameters β and α by switching between the iterations

$$\begin{aligned}\left(\sum_{i=1}^{N} Z_i' D_i V_i^{-1} D_i' Z_i + \Lambda \mathrm{P}\right)\beta^{(k+1)} &= \sum_{i=1}^{N} Z_i' D_i V_i^{-1} D_i' \tilde{y}_i \\ \left(\sum_{i=1}^{N} \tilde{Z}_i' C_i U_i^{-1} C_i' \tilde{Z}_i + \Delta\Omega\right)\alpha^{(k+1)} &= \sum_{i=1}^{N} \tilde{Z}_i' C_i U_i^{-1} C_i' \tilde{w}_i\end{aligned}$$

until convergence. Here we use again working observations $\tilde{y}_i = Z_i\beta + (D_i^{-1})'(y_i - \pi_i)$ and $\tilde{w}_i = \tilde{Z}_i\alpha + (C_i^{-1})'(w_i - \nu_i)$. Note that the solution of the penalized generalized estimating equations for β depends on α only through the working covariance V_i. Therefore like in parametric GEE approaches the estimator $\hat{\beta}$ should be robust against the misspecification of the association structure.

To get an approximation for the covariance of the final estimate $\hat{\beta}$ we use the robust sandwich matrix

$$V(\hat{\beta}) = H^{-1} G H^{-1}$$

with

$$H = \sum_{i=1}^{N} Z_i D_i' V_i^{-1} D_i Z_i' + \Lambda \mathrm{P}$$

and

$$G = \sum_{i=1}^{N} Z_i D_i' V_i^{-1}(y_i - \hat{\pi}_i)(y_i - \hat{\pi}_i)' V_i^{-1} D_i Z_i'.$$

Alternatively we can use the model–based or naive covariance matrix

$$V(\hat{\beta}) = H^{-1}.$$

5 Example 2: Forest damage data

Since 1983 a yearly visual forest damage inventory is carried out in the forest district of Rothenbuch in the northern part of Bavaria. There are 80 observation points with occurrence of beeches spread over the whole area. In this damage study we analyze the influence of covariates, e.g., age of the trees, pH value of the soil, and canopy density of the stand, on the defoliation of beeches at the stand. We use the degree of defoliation as an indicator for damage state of the trees. Due to the survey design, responses must be assumed to be serially correlated. The ordinal response variable, Y_t, "damage state" at time t is measured in 3 categories: none ($Y_t = 1$), light ($Y_t = 2$), and distinct/strong ($Y_t = 3 =$ reference) defoliation. Figure 3 shows the relative frequencies of the damage categories in the sample for the years 1983 to 1994. There is an apparent change for the worse up through the year 1988 followed by an improvement during the next 5 years. A detailed survey and data description can be found in Göttlein and Pruscha (1996).

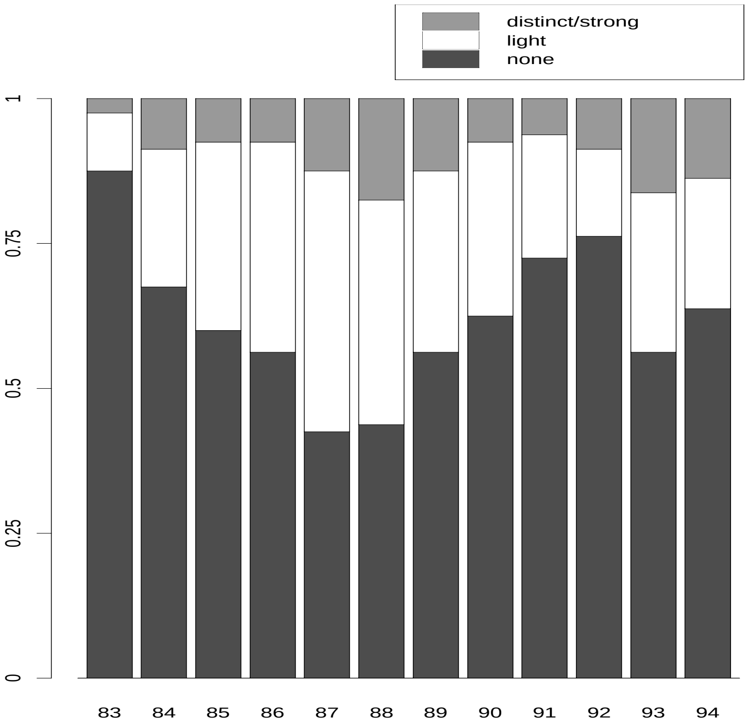

Figure 3: Damage class distribution by time

Due to the ordinal scale of the response, we use a cumulative logistic model to relate the marginal probabilities of "damage state" to the following covariates:

- A Age of the trees at the beginning of the study with categories: below 50 years (=1), between 50 and 120 years (=2), and above 120 years (=reference).
- PH PH value of the soil in 0-2 cm depth. The measures range from a minimum of 3.3 to a maximum of 6.1.
- CD Canopy density at the stand with categories: low (=1), medium (=2), and high (=reference).

The covariates pH value and canopy density vary for each stand over time, while the variable age is time constant by construction. In particular, we assume for the marginal cumulative probabilities of no damage ($r = 1$) and none or light damage ($r = 2$) the following model

$$\text{logit}(\text{pr}(Y_t \leq r)) = \gamma_r(t) + \gamma_3(t)A^{(1)} + \gamma_4(t)A^{(2)} + \gamma_5(PH_t) + \gamma_6 CD_t^{(1)} + \gamma_7 CD_t^{(2)}, \quad r = 1, 2,$$

where $A^{(1)}, A^{(2)}, CD_t^{(1)}, CD_t^{(2)}$ are dummy variables for the categorical covariates A and CD. To capture the time trend in the data we allow the threshold functions and the effects of the time constant variable age to vary smoothly with time t. Due to a lack of information about the form of the influence, it is reasonable to model the effect of pH value nonparametrically by an unspecified smooth function. The effects of canopy density are assumed to be fixed like in an ordinary parametric model.

First, we analyze the data with the working assumption of independent responses. This results in point estimates which are identical to the penalized maximum likelihood estimates of a model based on T independent responses. Figure 4 shows the estimated threshold functions $\hat{\gamma}_1(t)$ and $\hat{\gamma}_2(t)$. Both curves decrease up to the year 1988 with a more pronounced decrease of the first threshold $\hat{\gamma}_1(t)$. This indicates a shift to higher probabilities for the categories light and distinct/strong damage up to this year. After an improvement, i.e. a shift to the none damage category, up to 1992 there is another increase in damage up to 1994. This result is true for beeches above 120 years, i.e. for the reference category of age. For the other two categories of age we have in addition to consider the effects of the corresponding dummies. Both effects are positive over the 12 years (Figure 5, left plot). This indicates a positive influence on minor damage, i.e. younger beeches are less damaged. The positive effect of the category with below 50 years old trees (upper curve) is greater and the increase of this effect after 1988 corrects the change to the worse after 1992 indicated by the threshold functions. These interpretations are further illustrated by Figure 6, where the sums of the threshold functions and effects of the dummy variables of age are plotted against time. The estimated function for the influence of pH value is almost linear over the range of observed pH values (Figure 5, right plot). Stands

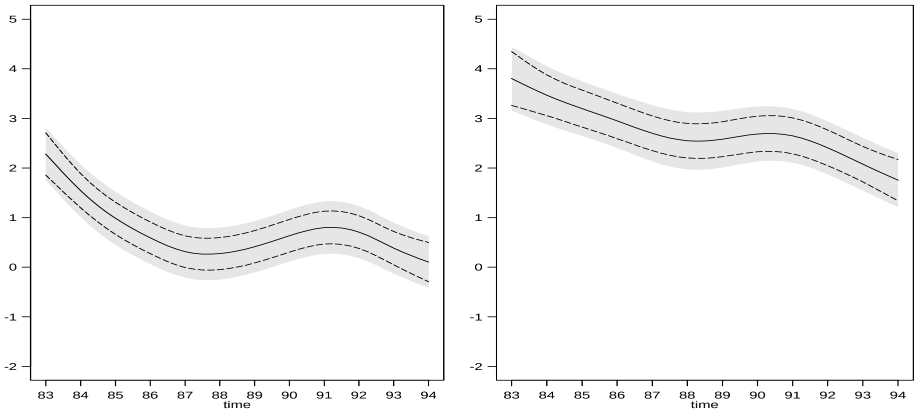

Figure 4: Independence model: Estimated thresholds $\hat{\gamma}_1$ (left) and $\hat{\gamma}_2$ (right) with pointwise standard error bands (model based - dashed lines, robust - shaded region)

with low pH values have a negative influence on damage state compared to stands with less acid soils, i.e. low pH values aggravate the condition of the trees. Finally we have the following parameter estimates for the effects of canopy density together with model based and robust standard errors:

Covariate	Estimate	SE	SE	p–value	p–value
		(model)	(robust)	(model)	(robust)
$CD_t^{(1)}$	-1.4868	0.2325	0.5285	0.0000	0.0049
$CD_t^{(2)}$	-0.3378	0.1687	0.3425	0.0452	0.3239

This means that stands with low ($CD_t^{(1)}$) or medium ($CD_t^{(2)}$) density have an increased probability for high damage compared to stands with a high canopy density.

We see that the model assuming independence distinctly underestimates the standard errors for the thresholds and covariate effects. This indicates that the independence assumption is quite far away from the real association structure. Hence we try to find a more realistic working model.

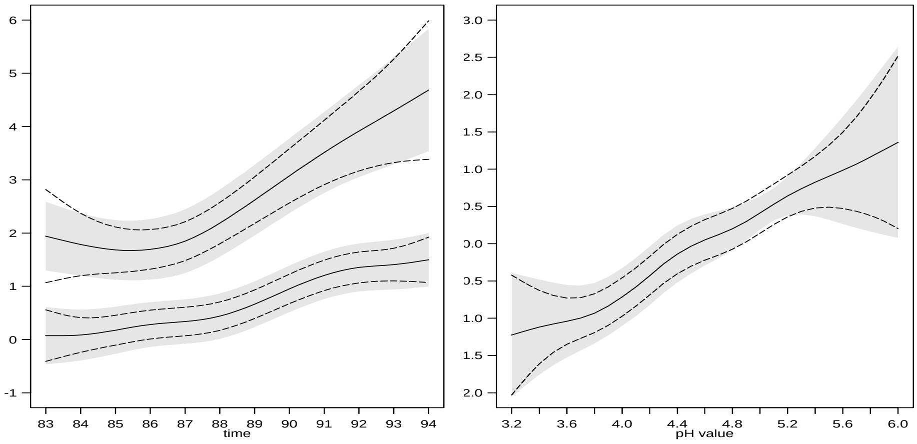

Figure 5: Independence model: Estimated effects of age (left) and pH value (right) with pointwise standard error bands (model based - dashed line, robust - shaded region)

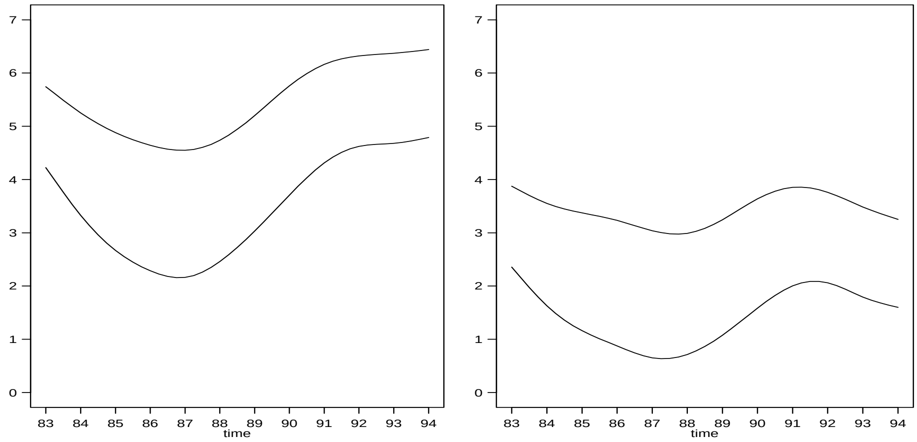

Figure 6: Independence model : $\hat{\gamma}_r(t) + \hat{\gamma}_3(t) AGE^{(1)}$ (left) and $\hat{\gamma}_r(t) + \hat{\gamma}_4(t) AGE^{(2)}$ (right)

In the second analysis, we combine the marginal mean model with a model for the pairwise association structure. With $T = 12$ measures per stand we have 66 time pairs at which we measure the pairwise association by global odds ratios. A preliminary descriptive analysis with empirically estimated global odds ratios indicates different values of the odds ratios for each cutpoint (l, r) and a decline in association with the time distance between the visits to the stand. Thus the association structure is parameterized by the logarithmic model of the form

$$\log(\psi_{i,st}^{(lr)}) = \alpha_{lr} + \gamma(|t - s|) \qquad l, r = 1, 2,$$

i.e. we use a time–stationary association for each cutpoint (l, r). Note that we do not force the dependence on the time lag into a special parametric

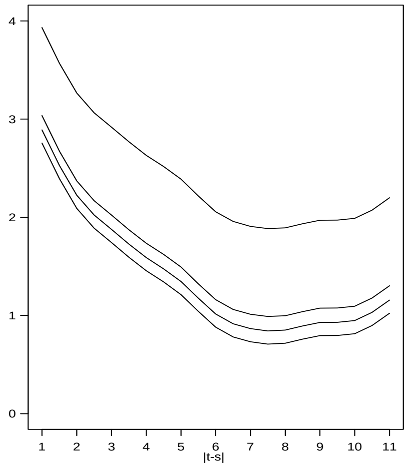

Figure 7: Estimated log global odds ratios $\log(\psi_{st}^{(lr)})$

form. We use again a unspecified smooth function to determine the influence of the time-lag $|t-s|$ on the global odds ratios.

Estimating simultaneously both models using two penalized estimating equations yields the following estimates for the association parameters

	$\hat{\alpha}_{11}$	$\hat{\alpha}_{12}$	$\hat{\alpha}_{21}$	$\hat{\alpha}_{22}$
Estimate	2.3053	2.4393	2.5855	3.4809

The estimates are quite similar for the association parameters α_{11},α_{12} and α_{21}; only the global odds ratio α_{22} for the cutpoint (2,2) is on a higher level. Figure 7 shows the logarithmic global odds ratio. There is a distinct decrease in the association between two responses as the time distance increases.

Looking at the estimated mean structure we observe that the predicted marginal probabilities for both models are very similar. Figure 8 shows these probabilities of the second model for the three categories of age assuming the stand has a medium canopy density and a medium PH value ($PH = 4.2$).

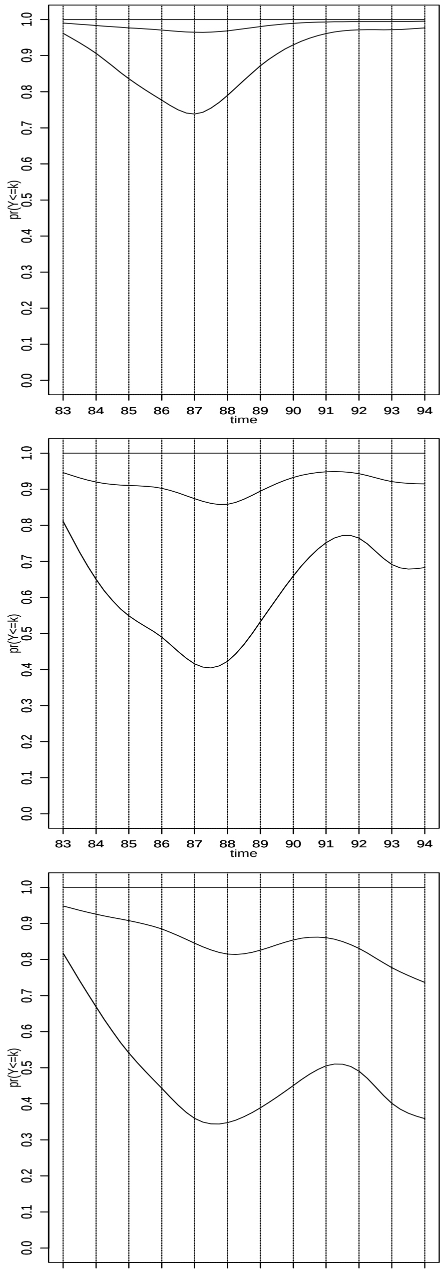

Figure 8: Estimated cumulative probabilities $\mathrm{pr}(Y_t \leq 1), \mathrm{pr}(Y_t \leq 2)$ and $\mathrm{pr}(Y_t \leq 3) = 1$ for the three age categories. From top to bottom: up to 50 years, between 50 and 120 years, above 120 years.

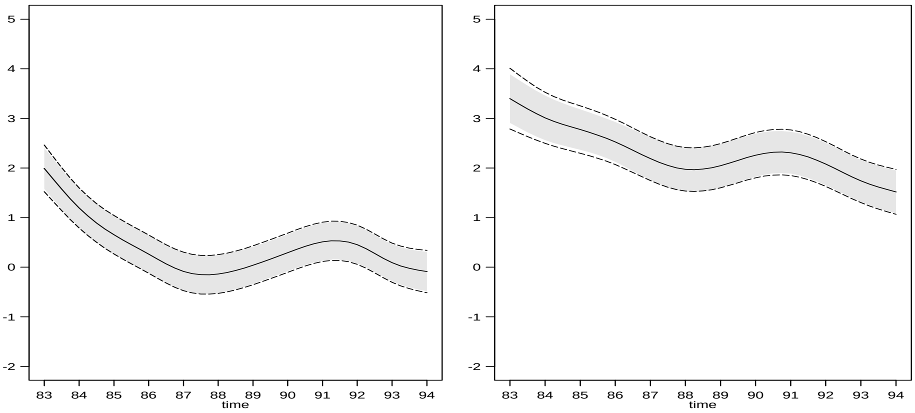

Figure 9: Refined model – Estimated thresholds $\hat{\gamma}_1$ (left plot) and $\hat{\gamma}_2$ (right plot) with pointwise standard error bands (model based - dashed line, robust - boundary of shaded region)

In contrast to this similarity, the estimated thresholds and effects partially change considerable. In particular, the effect of different PH–values is less distinct, i.e the course of the estimated curve is flatter (Figure 10, right plot). In fact there is some doubt that PH–value has an influence at all. Also the fixed effects of canopy density are quite different for both models .

Covariate	Estimate	SE	SE	p–value	p–value
		(model)	(robust)	(model)	(robust)
$CD_t^{(1)}$	-1.3007	0.3955	0.3741	0.0010	0.0005
$CD_t^{(2)}$	-0.4565	0.2628	0.2424	0.0825	0.0596

But note that the model–based and robust standard errors for the second model are closer together. We take this observation as an indication that the association structure is better considered in the second model. The effects of age are almost unchanged (Figure 10, left plot). To adjust these changes the threshold functions are on a lower level (Figure 9) compared to the independence model, but their overall shape remains the same.

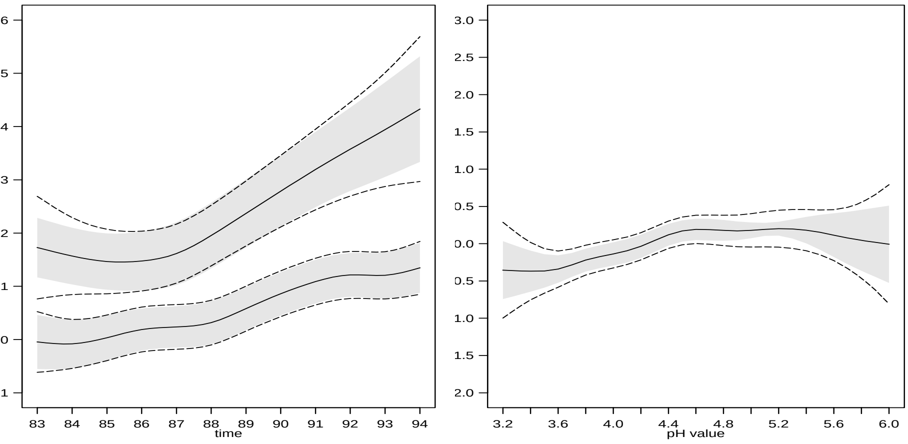

Figure 10: Refined model – Estimated effects of age (left) and pH value (right) with pointwise standard error bands (model based - dashed line, robust - boundary of shaded region)

6 Conclusions

In two applications, we have shown that semiparametric methods provide a flexible tool for an analysis of multicategorical and multivariate correlated data. Therefore they may supplement existing parametric standard methods.

Our estimation approach is either based on penalized likelihoods for univariate ordinal response or on penalized estimating equations for correlated response. Alternatively we can specify the entire likelihood of the correlated outcomes by one of the parametrizations mentioned in Section 1. Starting from this full likelihood we can derive penalized likelihood methods for marginal regression models. This will be the topic of forthcoming paper.

It should be mentioned that the methodology can be applied to other multivariate structures. For example, correlated nominal responses can be analysed by a semiparametric marginal model using local odd ratios to measure the association.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 “Statistische Analyse diskreter Strukturen, Modellierung und Anwendung in Biometrie und Ökonometrie”. Data from the Attitudes Toward Smoking Legislation Survey, which was funded by NHRDP (Health and Welfare Canada), were collected by the Institute for Social Research at York University for Dr. L. Pederson and Dr. S. Bull. Data from the Forest Damage Survey were kindly made availble by Dr. H. Pruscha and Dr. A. Göttlein

References

BULL, S. (1994). Analysis of Attitudes toward Workplace Smoking Restrictions. In *Case Studies in Biometry* edited by Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L., and Greenhouse, J. New York: Wiley, 249–271.

DALE, J.R. (1986). Global Cross–Ratio Models for Bivariate, Discrete, Ordered Responses. *Biometrics*, 42, 909–917.

DEMMLER, A., REINSCH, C. (1975). Oscillation matrices with spline smoothing. *Numerische Mathematik* 24, 375–382.

EUBANK, R. (1988). *Spline Smoothing and Nonparametric Regression.*New York: Marcel Dekker.

FAHRMEIR, L. (1990). Maximum Likelihood Estimation in Misspecified Generalized Linear Models. *Statistics*, 21, 487–502.

FAHRMEIR, L., KLINGER, A. (1996). A Nonparametric Multiplicative Hazard Model for Event History Analysis. *Discussion paper 12*, SFB 386, Ludwig–Maximilians Universität, München.

FAHRMEIR, L., PRITSCHER, L. (1996). Regression Analysis of Forest Damage by Marginal Models for Correlated Ordinal Responses. *Journal of Environmental and Ecological Statistics*, 3, 257–268.

FAHRMEIR, L., TUTZ, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer-Verlag.

FITZMAURICE, G.M., LAIRD, N.M. (1993). A Likelihood–based Method for Analysing Longitudinal Binary Responses. *Biometrika*, 80, 141–151.

GÖTTLEIN, A., PRUSCHA, H. (1996). Der Einfluß von Bestandskenngrößen, Topographie, Standort und Witterung auf die Entwicklung des Kronenzustandes im Bereich des Forstamtes Rothenbuch. *Forstwirtschaftliches Centralblatt*, 114, 146–162.

HASTIE, T., TIBSHIRANI, R. (1993) Varying–coefficient Models. *Journal of the Royal Statistical Society*, B 55, 757–796.

HEAGERTY, P., ZEGER, S. (1996). Marginal Regression Models for Clustered Ordinal Measurements. *Journal of the American Statistical Association*, 91, 1024–1036.

HEUMANN, C. (1996). Marginal Regression Modeling of Correlated Multicategorical Response: A Likelihood Approach. *Discussion paper 19*, SFB 386, Ludwig–Maximilians Universität, München.

LAIRD, N.M. (1996). Longitudinal Panel Data: An Overview of Current Methodology. In *Time Series Models in Econometrics, Finance and Other Fields* edited by Cox, D.R., Hinkley, D.V. and Barndorff–Nielsen, O. London: Chapman and Hall, 143–175.

LIANG, K.Y., ZEGER, S. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73, 13–22.

LIANG, K.Y., ZEGER, S.L., QAQISH, B. (1992). Multivariate Regression Analysis for Categorical Data. *Journal of the Royal Statistical Society*, B 54, 3–40.

LIPSITZ, S., LAIRD, N., HARRINGTON, D. (1991). Generalized Estimation Equations for Correlated Binary Data: Using The Odds Ratio as a measure of Association. *Biometrika*, 78, 153–160.

MC CULLAGH, P. (1980). Regression Model for Ordinal Data. *Journal of the Royal Statistical Society*, B 42, 109–127.

MOLENBERGHS, G., LESAFFRE, E. (1994). Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution. *Journal of the American Statistical Association*, 89, 633–644.

PRENTICE, R.L. (1988). Correlated Binary Regression with Covariates Specific to Each Binary Observation. *Biometrics*, 44, 1033–84.

WAHBA, G. (1990). *Spline Models for Observational Data.* Philadelphia: Society for Industrial and Applied Mathematics.

WILD, C.J., YEE, T.W. (1996). Additive Extensions to Generalized Estimating Equation Methods. *Journal of the Royal Statistical Society*, B 58, 711–725.