Kauermann, Tutz:

Semiparametric Modeling of Ordinal Data

Sonderforschungsbereich 386, Paper 193 (2000)

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

Semiparametric Modeling of Ordinal Data

Göran Kauermann and Gerhard Tutz

Ludwig Maximilian University Munich

Akademiestraße 1, D-80799 Munich

4th May 2000

Abstract

Parametric models for categorical ordinal response variables, like the proportional odds model or the continuation ratio model, assume that the predictor is given as a linear form of covariates. In this paper the parametric models are extended to a semiparametric or partially parametric form where parts of the covariates are modeled linearly and parts are modeled as unspecified but smooth functions. Estimation is based on a combination of local likelihood and profile likelihood and asymptotic properties of the estimates are derived. In a simulation study it is demonstrated that the profile likelihood approach is to be preferred over a backfitting procedure. A data example shows the applicability of the models.

KEYWORDS: Continuation Ratio Model, Cumulative Model, Local Likelihood, Ordinal Data, Profile Likelihood, Proportional Odds Model, Semiparametric Model, Sequential Model, Smoothing, Varying Coefficient Model

1 Introduction

We consider the ordinal response variable Y which takes values $Y \in \{1, \dots, k\}$ from a set of ordered categories. A widely used model for ordinal regression is the cumulative model which has been proposed by McCullagh (1980). This has the form

$$P(Y \leq r|x) = F\{\beta_{0r} + x^T\beta\} \quad \text{for} \quad r = 1, \dots, q = k-1 \tag{1}$$

where x is a set of explanatory quantities and the parameters β_{0r} fulfill the restrictions $\beta_{01} \leq \beta_{02} \leq \dots \leq \beta_{0q}$. In (1), the probability $P(Y \leq r|x)$ is linked to the linear predictor $\eta_r = \beta_{0r} + x^T\beta$ via the distribution function $F(\cdot)$. If $F(\cdot)$ is chosen as the logistic distribution function $F(\eta) = \exp(\eta)/\{1 + \exp(\eta)\}$ one obtains the widely used proportional odds model $\log\{P(Y \leq r|x)/P(Y > r|x)\} = \beta_{0r} + x^T\beta$.

The cumulative model (1) can be motivated by considering Y as a coarser version of a latent continuous variable $\mathcal{U}$, where $\mathcal{U}$ follows the "classical" linear regression model

$$\mathcal{U} = -x^T\beta + \varepsilon \tag{2}$$

with ε distributed according to the distribution function $F(\cdot)$ and $E(\varepsilon|x) = 0$, assuming that $F(\cdot)$ has zero mean. The connection between the observed response Y and the latent variable $\mathcal{U}$ is given by the threshold concept $Y = 1 \Leftrightarrow \mathcal{U} \leq \beta_{01}$ and $Y = r \Leftrightarrow \beta_{0r-1} \leq \mathcal{U} \leq \beta_{0r}$ for $r = 2, \dots, q$. Hence, the parameters β_{0r} can be seen as fixed unknown thresholds on a latent continuum, while the linear term $\eta = x^T\beta$ serves as predictor which parametrically shifts the mean of the latent variable on the latent continuum.

The structural assumption $E(\mathcal{U}|x) = -x^T\beta$ provides a model with linear predictor. If x consists of metrically scaled variables, however, linearity can be a too strong assumption for modeling data. In the same way as in smooth regression models, the parametric restriction can be circumvented by replacing the predictor $x^T\beta$ by a nonparametric function $\gamma(x)$, i.e. we set $E(\mathcal{U}|x) = -\gamma(x)$, where $\gamma(x)$ is unknown but assumed to be smooth. Due to the smooth modeling of the latent regression for $\mathcal{U}$ one obtains with fixed threshold parameters β_{0r} the smooth cumulative model

$$P(Y \leq r|x) = F\{\beta_{0r} + \gamma_0(x)\} \quad \text{for} \quad r = 1, \dots q. \tag{3}$$

It should be noted that (3) has a semiparametric structure since it contains the nonparametric specification of the covariate effect, $\gamma(x)$, and the thresholds β_{0r} as fixed parameters. Obviously, the parameters in model (3) are not uniquely defined and additional identifiability restrictions are needed. We set $\beta_{01} = 0$, i.e. we fix the first threshold, which in turn provides a simple numerical solution.

An alternative ordinal regression model is the sequential model or continuation ratio model

$$P(Y = r|Y \geq r, x) = F\{\beta_{0r} + x^T\beta\} \quad \text{for} \quad r = 1, \dots, q = k - 1. \tag{4}$$

The underlying idea behind (4) is that categories are reached successively. This means we model the binary transitions from category r to category $r + 1$, given category r is reached, as binary regression model with response function $F(\cdot)$. The linear predictor thereby has the form $\eta_r = \beta_{0r} + x^T\beta$ and describes the non-transition from r to $r+1$, where now no additional order restrictions are required for β_{0r}. This modeling approach essentially yields model (4) and in the same fashion as for the cumulative model, the sequential model (4) can be extended to the smooth version

$$P(Y = r|Y \geq r, x) = F\{\beta_{0r} + \gamma(x)\} \quad \text{for} \quad r = 1, \dots, q = k - 1, \tag{5}$$

simply by substituting $\gamma(x)$ for $x^T\beta$ in (4).

In this paper, estimation of the semiparametric structure in (3) and (5) will be based on local and profile likelihood estimation, as generally discussed in Severini & Wong (1992) or Severini & Staniswalis (1994). While (3) is treated as multivariate response model, (5) can be fitted as a univariate binary response. This property was emphasized for the parametric sequential model (4) e.g. by Cox (1988). We derive asymptotic variance formulae for both cases and discuss bias properties in the presence of ordinal data. In addition, models (3) and (5) will be considered in a more general form by including factorial effects and interactions which leads to varying coefficient models as introduced by Hastie & Tibshirani (1993). This allows to model smooth interaction between factorial and continuous regressors.

Background material to cumulative and sequential models is found for instance in Agresti (1990), Fahrmeir & Tutz (1994), Greenland (1994) or Barnhart & Sampson (1994). Simonoff (1996) discusses the smoothing of sparse ordinal data, which applies if the number of categories are large and correspondingly the cell frequencies are small, a topic not focussed in this paper (see also Hall & Titterington, 1987). Extensions of the cumulative model to nonparametric regression models of the generalized additive type are found in Hastie & Tibshirani (1990), Yee & Wild (1996) and Wild & Yee (1996). Estimation there has been based on the penalizing concept which yields spline fitting functions. We pursue a local and profile likelihood approach which extends smooth estimation as considered in semiparametric models (see also Severini & Wong, 1992). Cumulative models for dependent ordinal data are treated in Fahrmeir & Pritscher (1996) or Heagerty & Zeger (1996) in a parametric fashion and in Kauermann (1999) using nonparametric components. The latter paper does not make use of the semiparametric approaches, which are investigated here.

2 The Cumulative Model

2.1 Varying Coefficient Model

Although model (3) is already a semiparametric model, we consider a more general form of (3). This is done by including covariates $z = (z_1, z_2)$, where the effect of z_1 is modeled parametrically and the effect of z_2 is allowed to interact smoothly with x. This means the model for the latent variable $\mathcal{U}$ is specified by

$$E(\mathcal{U}|x, z) = -\{z_1^T\beta_1 + \gamma_0(x) + z_2^T\gamma_z(x)\}. \tag{6}$$

In (6), the regressors z_1 act additively in the usual parametric fashion $z_1^T\beta_1$. The main effect of x is given by the smooth function $\gamma_0(x)$, whereas $\gamma_z(x)$ represents the effect of z_2 which is smoothly modified by x. Thus the model contains varying coefficients in the sense of Hastie & Tibshirani (1993). If the variables in z are factorial covariates, parametric modeling like (6) is natural. If z contains also metrically scaled variables, the interaction of $z_1^T\beta$ means that one has a parametric form in mind, which may for instance be based on prior knowledge. The model for the ordinal response resulting from (6) equals

$$P(Y \leq r|z, x) = F\{\beta_{0r} + z_1^T\beta_z + \gamma_0(x) + z_2^T\gamma_z(x)\}\ ,\ r = 1, \dots, q, \tag{7}$$

where $0 = \beta_{01} \leq \beta_{02} \dots \leq \beta_{0q}$. The additive term $\beta_{0r} + z_1^T\beta_1$ thereby represents the parametric part of the model, while the nonparametric part is given by $\gamma_0(x)$ as main effect and $z_2^T\gamma_z(x)$ as varying-coefficient term. Of course, (3) is a special case of (7) following from the omission of the components (z_1, z_2). Moreover, if

z_1 is neglected, (7) simplifies to the varying coefficient model $P(Y \leq r|z,x) = F\{\beta_{0r} + \gamma_0(x) + z_2^T\gamma_1(x)\}$. Hence, (7) represents a general class of smooth models for ordinal data, including parametric, nonparametric and varying coefficient terms.

2.2 Local and Profile Likelihood Estimation

It is convenient to write model (7) in matrix form. Let $\pi^T = (\pi_1, \ldots \pi_q)$ denote the vector of cell probabilities $\pi_r = P(Y = r|z,x)$. Then the multivariate model has the form

$$\begin{aligned} g(\pi) &= \left(\underbrace{\begin{matrix} 0 & \cdots & 0 & z_1^T \\ 1 & \cdots & 0 & z_1^T \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & z_1^T \end{matrix}}_{=:\, Z_1} \; \underbrace{\begin{matrix} 1 & z_2^T \\ 1 & z_2^T \\ \vdots & \vdots \\ 1 & z_2^T \end{matrix}}_{=:\, Z_2} \right) \begin{pmatrix} \beta_{02} \\ \vdots \\ \beta_{0q} \\ \beta_1 \\ \gamma_0(x) \\ \gamma_z(x) \end{pmatrix} \\ &= Z_1\beta + Z_2\gamma(x) \end{aligned}$$

with $\beta^T = (\beta_{02}, \ldots, \beta_{0q}, \beta_1^T)$ and $\gamma(x)^T = \{\gamma_0(x), \gamma_z(x)^T\}$. The link function $g = (g_1, ..., g_q)$ has components $g_r(\pi) = F^{-1}(\pi_1 + \ldots + \pi_r)$. Let now $(Y_i, z_i, x_i), i = 1, \ldots, n$ denote a random sample, and define $y_i = (y_{i1}, \ldots, y_{iq})^T$ as indicator vector with elements $y_{ir} = \delta(Y_i = r) = 1$ if $Y_i = r$ and 0 otherwise. Moreover, let $Z_{1,i}$ and $Z_{2,i}$ denote the matrices Z_1 and Z_2 evaluated at z_i.

Local and Profile Likelihood Estimation

With $w_{h,ij} = K\{(x_i - x_j)/h\}$ we define kernel weights, where $K(\cdot)$ denotes a symmetric, unimodal kernel and h is the bandwidth or smoothing parameter. We first treat the parametric part of the linear predictor as fixed, i.e. we consider $Z_1\beta$ as given offset. We then fit $\gamma_i = \gamma(x_i)$ for $i = 1, \ldots, n$ by maximizing the local likelihood $\sum_j w_{h,ij} l_j(\beta, \gamma_i)$ where $l_j(\beta, \gamma_i) = \log P(Y = Y_j | \eta_{i,j})$ is the log likelihood contribution with $\eta_{i,j} = Z_{1,j}\beta + Z_{2,j}\gamma_i$. Differentiation yields

$$0 = \sum_{j=1}^{n} w_{h,ij} Z_{2,j}^T \, l_{\eta,j}(\beta, \widehat{\gamma}_i) \tag{8}$$

where $l_{\eta,j}(\beta, \gamma_i) = \{\partial g^{-1}(\eta_{i,j})/(\partial\eta)\} \operatorname{cov}(y_j)^{-1} \{y_j - \pi_j(\eta_{i,j})\}$ is the standard score contribution with $\pi_j(\eta_{i,j})^T = \{P(Y_j = 1|\eta_{i,j}), ..., P(Y_j = q|\eta_{i,j})\}$. The solution of (8) may be found by iterative weighted Fisher scoring. Since (8) is solved for a particular choice of β, the resulting estimate $\widehat{\gamma}_i$ depends on β, which is however suppressed in the notation.

Profile Likelihood Estimation

Estimation of the parameter β is done by profile likelihood estimation. We insert $\widehat{\gamma}_i$ from above in the likelihood for β and maximize the profile likelihood function $\sum_i l_i(\beta, \widehat{\gamma}_i)$. Differentiation with respect to β yields the estimating equation

$$0 = \sum_{i=1}^{n} \left(Z_{1,i}^T + \frac{\partial \widehat{\gamma}_i^T}{\partial \beta} Z_{2,i}^T \right) l_{\eta,i}(\widehat{\beta}, \widehat{\gamma}_i) \tag{9}$$

where $\partial \widehat{\gamma}_i^T/(\partial\beta)$ is found by differentiating (8) with respect to β. This yields

$$\frac{\partial \widehat{\gamma}_i^T}{\partial \beta} = -\left\{ \sum_{j=1}^{n} w_{h,ij} Z_{1,j}^T l_{\eta\eta,j}(\beta, \widehat{\gamma}_i) Z_{2,j} \right\} \left\{ \sum_{j=1}^{n} w_{h,ij} Z_{2,j}^T l_{\eta\eta,j}(\beta, \widehat{\gamma}_i) Z_{2,j} \right\}^{-1} \tag{10}$$

where $l_{\eta\eta,j}(\beta, \widehat{\gamma}_i) = \partial l_{\eta,j}(\beta, \widehat{\gamma}_i)/\partial \widehat{\eta}_{i,j}$ with $\widehat{\eta}_{i,j} = (Z_{1,j}\beta + Z_{2,j}\widehat{\gamma}_i)$. For simplicity one can replace $\widehat{\gamma}_i$ by γ_j and $l_{\eta\eta,j}(\beta, \widehat{\gamma}_i)$ by $-F_j \;:=\; E\{l_{\eta\eta,j}(\beta, \gamma_j)\} \;=$

$\{\partial g^{-1}(\eta_j)/(\partial\eta)\}\mathrm{cov}(y_j)^{-1}\{\partial g^{-1}(\eta_j)/(\partial\eta^T)\}$ to reduce the computational effort. Defining $\tilde{Z}_{1,i} = Z_{1,i} + \partial\widehat{\gamma}_i^T/\partial\beta \times Z_{2,i}^T$, the profile score equation (9) becomes $0 = \sum_i \tilde{Z}_{1,i}^T l_{\eta,i}(\widehat{\beta},\widehat{\gamma}_i)$. One should note that $\tilde{Z}_{1,i}$ depends on β so that iteration between (8) and (9) is required to provide the final estimates for β and $\gamma(u)$.

Asymptotic Consideration of the Estimates

Severini & Wong (1992) show that the profile likelihood approach provides efficient estimation for β. In particular, the fixed parameter β is estimated by the usual parametric $\sqrt{n}$ asymptotic rate. Bias consideration and variance estimation is directly available from the estimating equations, as shown in the appendix. One finds

$$\begin{aligned} E(\widehat{\beta}-\beta) &= O(h^4) && (11)\\ \mathrm{cov}(\widehat{\gamma}_i) &= \left\{\sum_{j=1}^n w_{ij} Z_{2,j}^T F_j Z_{2,j}\right\}^{-1}\left\{\sum_{j=1}^n w_{ij}^2 Z_{2,j}^T F_j Z_{2j}\right\}\left\{\sum_{j=1}^n w_{ij} Z_{2,j}^T F_j Z_{2,j}\right\}^{-1} && (12)\\ \mathrm{cov}(\widehat{\beta}) &= \left\{\sum_{j=1}^n \tilde{Z}_{1,j}^T F_j \tilde{Z}_{1,j}\right\}^{-1}. && (13) \end{aligned}$$

Profile Likelihood Versus Backfitting

An alternative to profile likelihood estimation is backfitting, which has become a popular estimation method in recent years for models with smooth components. The principle of backfitting is that in each step the components are fitted separately without consideration of mutual dependence. Applied to semiparametric models of the present setting, this means that the profile likelihood estimating equation (9) is replaced by

$$0 = \sum_{i=1}^n Z_{1,i}^T l_{\eta,i}(\widehat{\beta},\widehat{\gamma}), \qquad (14)$$

i.e. $\tilde{Z}_{1,i}$ in (9) is replaced by $Z_{1,i}$. Though this seems to be numerically simpler, no direct variance formula like (13) is available for the backfitting estimate. Moreover, the backfitting estimate is generally more biased, as has been shown by Severini & Wong (1992) (see also Speckman, 1988). In a simulation study in the next section we demonstrate that this point holds also in the ordinal case.

Choice of the Bandwidth

As generally in the smoothing context, the bandwidth h has to be chosen. For simplicity we assume that x is univariate and h is a scalar. A common data driven procedure is available by using cross validation. However, in a semiparametric model as considered here, cross validation goes along with the burden of heavy computation, because iterated refitting is required. We therefore suggest to make use of Generalized Cross Validation (see Hastie & Tibshirani, 1990), which applied to this setting means we minimize

$$GCV(h) \quad = \quad \frac{\sum_i (y_i - \pi)^T \mathrm{var}(y_i)^{-1} (y_i - \pi)}{nq\{1 - df/(nq)\}^2} \tag{15}$$

where df is the degree of the model which we calculate from $df = p + \sum_i \mathrm{tr}(S_{h,i})$. Here p is the dimension of β and $tr(S_{h,i})$ denotes the trace of the matrix $S_{h,i} = (\sum_j^n w_{h,ij} Z_{2,j}^T F_j Z_{2,j})^{-1} Z_{2,i}^T F_i Z_{2,i}$ One should note that for $h \to \infty$, i.e. if $\gamma = constant$ is fitted, the resulting degree df equals the number of parameters fitted. The elements in $S_{h,i}$ can thereby be seen as the diagonal block in a smoothing type matrix for fitting $\gamma(x_i)$.

2.3 Example and Simulation

Simulation

We simulate data from the sequential model $\mathcal{U} = -\{z_1\beta_1 + \gamma_0(x)\} + \varepsilon$ with $\gamma_0(x) = 1 - \cos(3.14x)$ and $\beta_1 = 1$, where ε is distributed according to the logistic distribution function. As threshold parameters we choose $\beta_{01} = 0$, $\beta_{02} = 1$ and $\beta_{03} = 2$. We take z_1 as binary dummy coded factor and choose x as 20 equidistant points on $[0, 1]$. For each design point of x we draw 4 replicates of $\mathcal{U}$ from two different settings for z_1. First, we choose z_1 as balanced binary factor, i.e. $z_1 = 1$ for half of the data, and secondly, we draw z_1 in each simulation from the Bernoulli distribution $P(z_1 = 1|x) = \text{logit}^{-1}(1.5 - 3x)$. We fit the cumulative logit model $P(Y \leq r|z_1, x) = \text{logit}^{-1}\{\beta_{0r} + z_1\beta + \gamma_0(x)\}$ by profile likelihood estimation. For comparison we also pursue backfitting estimation by using (8) and (14). We run 150 simulations in each setting and choose bandwidth h in each run using (15). Table 1 shows the results of the simulation. Both, backfitting and profile likelihood estimate behave comparable, though the backfitting estimate is slightly more biased. The profile likelihood estimate provides valid and direct variance estimates, as proven above. In contrast, taking $(\sum_i Z_{1,i}^T F_i Z_{1,i})^{-1}$ as variance estimate for the backfitting estimate, i.e. using $Z_{1,i}$ instead of $\widetilde{Z}_{1,i}$ in (13), one obtains variance estimates which are too small. This shows in Figure 1 where the normal quantiles are plotted against the quantiles of the standardized estimate $(\widehat{\beta}_1 - \beta_1)/\sqrt{\text{var}(\widehat{\beta}_1)}$, where standardization is done by the use of the corresponding variance estimate. Profile likelihood estimation yields the right variance estimates while backfitting shows the need for a correction of the variance estimates. Moreover, the backfitting estimate yields more extreme values which also signals the bias problem of the backfitting estimate.

Example

We analyze data collected at $n = 123$ patients in a clinical study on the healing of sports related injuries of the knee (for further information on the data see Tutz, 2000). By random design a therapy was chosen and the patients were treated under two different condition. The treatment group used an anti-inflammatory spray while the placebo group used a spray without active ingredients. After a ten days treatment with the spray, the mobility of the knee was investigated in a standardized experiment during which the knee was actively moved by the patient. The pain Y occurring during the movement investigation was assessed on a four point scale from 1 representing no pain to 4 giving severe pain. We model the data by the smooth cumulative logit model

$$P(Y \leq r|age, treatment) \quad = \quad F(\beta_{or} + treatment\, \beta_t + \gamma_0(age)\} \tag{16}$$

with *age* as the patient's age and *treatment* as indicator variable with value 0 for the placebo group and 1 for the treatment group. The estimates are shown in Table 2. For comparison a parametric model with quadratic effect of age is also fitted. It is seen that the treatment effect and its studentized value are about the same in both models. If a parametric model is fitted, a quadratic *age* effect seems necessary to model. For the semiparametric model it is however not necessary to decide a priori upon the modeling of age, since the influence of age is modeled smoothly. Hence, (16) is a more general model, but the parametric effect of *treatment* is still estimated with the usual parametric accuracy. Figure 2 shows the age effect which indicates that pain is decreasing for patients above 40 years of age. In Figure 3 the cumulative model is illustrated by plotting the mean $E(\mathcal{U}|x)$ of the latent score separately for the placebo and treatment group. The horizontal lines thereby show the thresholds β_{0r} on the latent scale. The larger the value of $\mathcal{U}$ the more pain is felt by the patient. Hence, $\mathcal{U}$ may be interpreted as an unobserved *pain score*. Although there is some variation in the distribution of $\mathcal{U}$, it is seen that the treatment group is shifted towards less pain.

3 The Sequential Model

3.1 Smooth Modeling

We now consider smooth extensions of the sequential model (4). As in the previous section, we let $z = (z_1, z_2)$ denote (factorial) covariates while x are continuous regressors. We model the conditional probabilities $P(Y = r|Y \geq r, z, x)$ for $r = 1, \ldots q = k - 1$ as varying coefficient model

$$P(Y = r|Y \geq r, x, z) \quad = \quad F\{\beta_{0r} + z_1\beta_1 + \gamma_0(x) + z_2\gamma_z(x)\} \tag{17}$$

where $\beta_{01} \equiv 0$ is set to ensure identifiability. The cell probabilities according to (17) are easily found by

$$P(Y = r|x, z) \quad = \quad P(Y = r|Y \geq r, x, z) \prod_{k=1}^{r-1} \{1 - P(Y = k|Y \geq k, x, z)\}.$$

Model (17) can again be derived from the consideration of latent variables. Let $\mathcal{U}_1, \ldots \mathcal{U}_q$ denote independent latent variables following the semiparametric regression model

$$\mathcal{U}_r = -\{z_1\beta_1 + \gamma_0(x) + z_2\gamma_z(x)\} + \varepsilon_r \tag{18}$$

where ϵ_r are independent, identically distributed residuals following the distribution function $F(\cdot)$. The response Y corresponds to the number of the first $\mathcal{U}_r$ which is less than the threshold β_{0r}. This means we get $Y = 1$ if $\mathcal{U}_1 \leq \beta_{01} = 0$. If $\mathcal{U}_1 > \beta_{01}$ we get $Y = 2$ if $\mathcal{U}_2 \leq \beta_{02}$ and so on, until the first category r for which $\mathcal{U}_r \leq \beta_{0r}$ occurs (see also Tutz, 1991).

3.2 Local and Profile Likelihood Estimation

The sequential model is composed as sequential stopping process and it shows similarities to discrete survival models. This mirrors in the likelihood as derived below. Let again (Y_i, z_i, x_i), $i = 1, \dots, n$, denote a random sample with $y_i^T = (y_{i1}, \dots, y_{iq})$ as indicator vector, where $y_{ir} = 1$ if $Y_i = r$ and zero otherwise. Moreover, we define the censor type variables $d_i^T = (d_{i1}, \dots, d_{iq})$ with $d_{ir} = 1$ if $Y_i \geq r$ and $d_{ir} = 0$ otherwise. This allows to write the log-likelihood function as

$$l\{\beta, \gamma(x)\} = \sum_{i=1}^{n} \sum_{r=1}^{q} d_{ir} l_{ir}(\beta, \gamma_i) \tag{19}$$

with $l_{ir}(\beta, \gamma_i) = y_{ir} \log\{\tilde{\pi}_{ir}/(1-\tilde{\pi}_{ir})\} + \log(1-\tilde{\pi}_{ir})$ and $\tilde{\pi}_{ir} = P(Y_i = r | Y_i \geq r, x_i, z_i)$, as given in (17), and $\beta^T = (\beta_{02}, \dots \beta_{0q}, \beta_1^T)$, $\gamma(x) = \{\gamma_0(x), \gamma_1(x)^T\}^T$. Therefore the likelihood may be written nicely as sum of contributions of a univariate binary response model. With $F(\cdot)$ being the logistic distribution function, the likelihood (19) is equal to the likelihood of a logit model for the random sample y_{ir}, x_i, z_i for $i = 1, \dots n$, $r = 1, \dots, q$ and $d_{ir} = 1$. The corresponding model for y_{ir} equals $P(y_{ir} = 1 | x_i, z_i, d_{ir} = 1) = F\{\beta_{0r} + z_{1i}\beta_z + \gamma_0(x_i) + z_{2i}\gamma_z(x_i)\}$ and restriction $\beta_{01} \equiv 0$ (compare Fahrmeir & Tutz, 1994, ch 9). Estimation is carried out by local and profile likelihood as in the previous section. Let $Z_{1,i}$ and $Z_{2,i}$ be defined as in the previous section and let $Z_{1,ir}$ and $Z_{2,ir}$ denote the r-th row of $Z_{1,i}$ and $Z_{2,i}$, respectively. Considering β as fixed, the local likelihood for estimating $\gamma_i = \gamma(x_i) = \{\gamma_0(x_i), \gamma_z(x_i)\}^T$ becomes

$$0 = \sum_{j=1}^{n} w_{h,ij} \sum_{r=1}^{q} d_{ir} \{Z_{2,jr}^T l_{\eta,ir}(\beta, \hat{\gamma}_i)\} \tag{20}$$

where $l_{\eta,ir}(\cdot) = \partial F(\eta)/\partial\eta \ \mathrm{var}(y_{ir}|d_{ir} = 1)^{-1}(y_{ir} - \tilde{\pi}_{ir})$. Inserting $\hat{\gamma}_i$ in the profile likelihood leads to the estimating equation

$$0 = \sum_{i=1}^{n}\sum_{r=1}^{q} d_{ir} \{\tilde{Z}_{1,ir}^T l_{\eta,ir}(\hat{\beta}, \hat{\gamma}_i)\} \tag{21}$$

where $\tilde{Z}_{1,ir} = Z_{1,ir} - \partial\hat{\gamma}_i^T/(\partial\beta) Z_{2,jr}^T$ and

$$\frac{\partial\hat{\gamma}_i^T}{\partial\hat{\beta}} = \{\sum_{j=1}^{n} w_{h,ij} \sum_{r=1}^{q} d_{jr} Z_{1,jr}^T l_{\eta\eta,jr}(\beta, \hat{\gamma}_i) Z_{2,jr}\}\{\sum_{j=1}^{n} w_{h,ij} \sum_{r=1}^{q} d_{jr} Z_{2,jr}^T l_{\eta\eta,jr}(\beta, \hat{\gamma}_i) Z_{2,jr}\}^{-1}$$

is found by differentiating (20). The results derived in the previous section hold in the same way for the sequential models. In particular, due to the factorization of the likelihood, (20) and (21) give equations from a univariate response model. One finds as estimates for the variance $\mathrm{var}(\hat{\gamma}_i)$ and $\mathrm{var}(\hat{\beta})$

$$\begin{aligned}\widehat{\mathrm{var}}(\hat{\gamma}_i) &= \{\sum_{j=1}^{n} w_{h,ij} \sum_{r=1}^{q} d_{jr} Z_{2,jr}^T F_{jr} Z_{2,jr}\}^{-1} \\ &\quad \times\{\sum_{j} w_{h,ij}^2 \sum_{r=1}^{q} d_{jr} Z_{2,jr}^T F_{jr} Z_{2,jr}\}\{\sum_{j=1}^{n} w_{h,ij} \sum_{r=1}^{q} d_{jr} Z_{2,jr}^T F_{jr} Z_{2,jr}\}^{-1} \qquad (22) \\ \widehat{\mathrm{var}}(\hat{\beta}) &= \{\sum_{i=1}^{n}\sum_{r=1}^{q} d_{ir} Z_{1,ir}^T F_{ir} Z_{1,ir}\}^{-1} \qquad (23)\end{aligned}$$

with $F_{ir} = -E(l_{\eta\eta,jr}(\beta, \gamma_j)$. One should note that the components d_{jr}, $j = 1, \ldots, n$ and $r = 1, \ldots, q$ in (22) and (23) are random variables which makes the variance formulae of non-standard structure. Details are found in the appendix.

3.3 Example

Example (continued)

We revisit the knee pain example form above and model the pain as sequential model

$$P(Y = 1|Y \geq r, treatment, age) \quad = \quad F\{\beta_{0r} + treatment\ \beta_t + \gamma_0(age)\}$$

with $F(\cdot)$ as logistic link. Parameter estimates are shown in Table 3. For comparison, we also fit a parametric sequential model with quadratic *age* effect. The treatment effect shows to be the same in the parametric and semiparametric model, both with comparable standard deviations. The threshold parameters β_{0r} now have a different interpretation than those in the cumulative model. Here, they give the effect upon the non-transition to a higher level of the pain. In Figure 4 the smooth age effect is shown for both smooth models, i.e. for the cumulative and for the sequential model. It is seen that the effects are rather similar in both models. This is not surprising since there are specific link functions, e.g. the link function resulting from the extreme value distribution, where cumulative and sequential model are identical (see e.g. Läärä & Matthews, 1985 or Tutz, 1991).

4 Discussion

Both types of models, the cumulative model as well as the sequential model, are extended to the semiparametric case above. The derivation is given for the case that the thresholds do not depend on covariates, which in both cases may be derived from a latent variable framework with nice interpretations. Assuming the thresholds themselves to depend on covariates one assumes that the predictor incorporates category specific effects like $\eta_r = \beta_{0r} + z_1^T\beta_r + \gamma_{0r}(x) + z_2^T\gamma_{zr}(x)$ for the r-th category.

in this case not only the intercepts but also the effects of z and x are specific to the category r. Of course, any combination of category specific and global effects is possible. The only modification in comparison to the case treated in the paper is, that the design matrices Z_1 and Z_2 are now constructed differently to cope for possible category specific parameterization, e.g. Z_2 gets a diagonal type structure. The fitting procedure and the asymptotic results itself remain unchanged. For the case of the sequential model, purely category specific effects are of less interest, since this corresponds to fitting separate binary models.

Acknowledgments

The authors gratefully acknowledge support in various aspects from the Sonderforschungsbereich 386 funded by the Deutsche Forschungsgemeinschaft.

A Technical Details

The Cumulative Model

For notational convenience we write $l_{\eta,j}$ for $l_{\eta,j}(\beta,\gamma_j)$, this means we neglect the parameter argument if we refer to the true parameters. Moreover with $l_{\eta\eta,j} = \partial l_{\eta,j}/(\partial\eta)$ we denote the second order derivative of the j-th likelihood contribution and $F_j{=}{-}E(l_{\eta\eta,j}){=}E(l_{\eta,j}l_{\eta,j}){=}\{\partial g^{-1}(\eta_j)/\partial\eta\}\mathrm{cov}(y_j)^{-1}\{\partial g^{-1}(\eta_j)/\partial\eta\}^T$ denotes the Fisher contribution. We use a bold notation if we refer to sums of elements, e.g. $\mathbf{F}_{12} = \sum_i Z_{1,i}F_iZ_{2,i}$, and we additionally use subscript (i) if we refer to weighted sums, e.g. $\mathbf{F}_{22,(i)} = \sum_j w_{h,ij}Z_{2,j}F_jZ_{2,j}$. We do not formally write down the technical assumptions required for the following statements, but essentially what is needed is that all Fisher matrices like $\mathbf{F}_{11}$ or $\mathbf{F}_{22,(i)}$ are invertible for $i = 1,\dots n$. This holds for n sufficiently large, if the design density $f(z_1, z_2, x)$ is not degenerated, and if h tends sufficiently slow to zero, i.e. $nh \to \infty$. For a more thorough discussion of these points we refer to Kauermann & Tutz (2000). We expand (8) about γ_j by considering β as true parameter, which is supposed to be known. This yields

$$\begin{aligned} 0 &= \sum_j w_{h,ij} Z_{2,j} \{ l_{\eta,j} + l_{\eta\eta,j} Z_{2,j} (\widehat{\gamma}_i - \gamma_j) + \ldots \} \\ \Leftrightarrow \widehat{\gamma}_i - \gamma_i &= \mathbf{F}_{22,(i)}^{-1} \{ \sum_j w_{h,ij} Z_{2,j} \{ l_{\eta,j} + F_j Z_{2,j} (\gamma_i - \gamma_j) \} \\ &\quad + O_p(n^{-1} h^{-1}) + O_p(n^{-1/2} h^{-1/2}) + O(h^4) \end{aligned} \tag{24}$$

where the asymptotic correction components follow by simple but tedious expansion (see Kauermann & Tutz, 2000). Since $l_{\eta,j}$ and $l_{\eta,i}$ are independent for $i \neq j$ and have zero mean one directly obtains (12). Moreover, the bias $E(\widehat{\gamma}_i - \gamma_i)$ equals $\mathbf{F}_{22,(i)}^{-1} \mathbf{b}_{2,(i)}$ with $\mathbf{b}_{2,(i)} = \sum_j w_{h,ij} Z_{2,j} F_j Z_{2,j} (\gamma_i - \gamma_j)$. Using standard smoothing arguments one can show that $\mathbf{F}_{22,(i)}$ has order $O(nh)$ and $\mathbf{b}_{2,(i)} = O(nh^3)$ so that $E(\widehat{\gamma}_i - \gamma_i) = O(h^2)$.

With $\widehat{\gamma}_{i|\beta}$ we now denote the solution of (8) for β fixed, i.e. we make the dependence of $\widehat{\gamma}_i$ on β explicit in the notation. Hence $\widehat{\gamma}_{i|\widehat{\beta}}$ denotes the final estimate. Expanding (9) about the true parameters β and γ_i yields

$$\begin{aligned} 0 &= \sum_i \tilde{Z}_{1,i}^T l_{\eta,i} (\widehat{\beta}, \widehat{\gamma}_{i|\widehat{\beta}}) \\ &= \sum_i \tilde{Z}_{1,i}^T [l_{\eta,i} + l_{\eta\eta,i} \{ Z_{1,i} (\widehat{\beta} - \beta) + Z_{2,i} (\widehat{\gamma}_{i|\widehat{\beta}} - \gamma_i) \} + \ldots] \\ &= \sum_i \tilde{Z}_{1,i}^T [l_{\eta,i} + l_{\eta\eta,i} \{ \tilde{Z}_{1,i} (\widehat{\beta} - \beta) + Z_{2,i} (\widehat{\gamma}_{i|\beta} - \gamma_i) \} + \ldots] \\ \Leftrightarrow \widehat{\beta} - \beta &= \tilde{\mathbf{F}}_{11}^{-1} [\sum_i \tilde{Z}_{1,i} \{ l_{\eta,i} - F_i Z_{2,i} (\widehat{\gamma}_{i|\beta} - \gamma_i) \}] + \ldots \end{aligned}$$

where $\tilde{\mathbf{F}}_{11} = \sum_i \tilde{Z}_{1,i}^T F_i \tilde{Z}_{1,i}$. It is shown in Severini & Wong (1992) that the expectation of the latter component above has negligible order $O(h^4)$. We give a short sketch of the statement here. Using expected second order derivatives in (10) gives

$\tilde{Z}_{1,i} = Z_{1,i} - \mathbf{F}_{12,(i)}\mathbf{F}_{22,(i)}^{-1}Z_{2,i}$ with $\mathbf{F}_{12,(i)} = \sum_j w_{h,ij}Z_{1,j}^T F_j Z_{2,j}$. Making use of (24) and using the above definition of $\tilde{Z}_{1,i}$ one finds

$$
\begin{aligned}
&E(\widehat{\beta} - \beta) \\
= \quad & -\tilde{\mathbf{F}}_{11}^{-1}\{\sum_i Z_{1,i}^T F_i Z_{2,i}\mathbf{F}_{22,(i)}^{-1}\mathbf{b}_{2,(i)} - \sum_i \mathbf{F}_{12,(i)}\mathbf{F}_{22,(i)}^{-1}Z_{2,i}F_i Z_{2,i}\mathbf{F}_{22,(i)}^{-1}\mathbf{b}_{2,(i)}\}. \quad (25)
\end{aligned}
$$

Matrix $\mathbf{F}_{22,(i)} =: \mathbf{F}_{22,(x_i)}$ can be interpreted as a conditional mean in the sense $\mathbf{F}_{22,(x_i)}/(nh)= E_{z_1,z_2}[Z_2^T F\{Z_1\beta + Z_2\gamma(x)\}Z_2|x_i]\{1 + O(h^2)\}$, where the expectation $E_{z_1,z_2}(\cdot|x_i)$ is carried out with respect to the conditional design density $f(z_1, z_2|x_i)$ and $F\{\eta\} = -E\{l_{\eta\eta}(\eta)\}$. Note that in the simple model (3), where z_1 and z_2 does not exist, formal integration is not required and $\mathbf{F}_{22,(i)}$ in this case is a simple function in x_i. Making use of the conditional expectation interpretation it follows, for instance,

$$
\begin{aligned}
& nh\sum_i \mathbf{F}_{22,(i)}^{-1}Z_{2,i}^T F_i Z_{2,i} \\
= \quad & n\int \mathbf{F}_{22,(x)}^{-1}\left[\int\int Z_2^T F\{Z_1\beta + Z_2\gamma(x)\}Z_2 f(z_1, z_2|x)dz_1dz_2\right] f(x)dx\{1 + O(h^2)\} \\
= \quad & n\int \mathbf{F}_{22,(x)}^{-1}\mathbf{F}_{22,(x)}f(x)dx\{1 + O(h^2)\} = nI\{1 + O(h^2)\}
\end{aligned}
$$

with I as identity matrix. Using arguments similar to this shows that the two components in (25) coincide up to order $O\{bias(\widehat{\gamma})\}O(h^2) = O(h^4)$ so that $E(\widehat{\beta} - \beta) = O(h^4)$ as claimed in (11).

The calculation of the variance of $\widehat{\beta}$ uses similar arguments. We first show that the covariates $\tilde{Z}_1$ and Z_2 are asymptotically orthogonal. This follows by the definition of $\tilde{Z}_1$ since

$$
\begin{aligned}
\tilde{\mathbf{F}}_{12}/n &= \sum_i \tilde{Z}_{1,i}^T F_i Z_{2,i}/n \\
&= \mathbf{F}_{12}/n - \{\sum_i \mathbf{F}_{12,(i)}\mathbf{F}_{22,(i)}^{-1}Z_{2,i}^T F_i Z_{2,i}/n\}\{1 + O(h^2)\} = O(h^2).
\end{aligned}
$$

Analogously one gets $\tilde{\mathbf{F}}_{12,(i)}/(nh) = \sum_j w_{h,ij}\tilde{Z}_{1,j}^T F_j Z_{2,j}/(nh) = O(h^2)$. This implies now

$$\begin{aligned}
& \operatorname{cov}\left[\sum_j \tilde{Z}_{1,j}^T l_{\eta,j}, \{\sum_i \tilde{Z}_{1,i} F_i Z_{2,i}(\widehat{\gamma}_{i|\beta} - \gamma_i)\}^T\right] \\
= & \sum_{i,j} w_{h,ij}\tilde{Z}_{1,j}^T F_j Z_{2,j}\mathbf{F}_{22,(i)}^{-1} Z_{2,i}^T F_i \tilde{Z}_{1,i} + \dots \\
= & \sum_i \tilde{\mathbf{F}}_{12,(i)}\mathbf{F}_{22,(i)}^{-1} Z_{2,i}^T F_i \tilde{Z}_{1,i}\{1 + O(h^2)\} = O(nh^4).
\end{aligned}$$

and similarly

$$\begin{aligned}
& \operatorname{cov}\left[\sum_i \tilde{Z}_{1,i} F_i Z_{2,i}(\widehat{\gamma}_{i|\beta} - \gamma_i), \{\sum_j \tilde{Z}_{1,j} F_j Z_{2,j}(\widehat{\gamma}_{j|\beta} - \gamma_j)\}^T\right] \\
= & \sum_i \tilde{\mathbf{F}}_{12,(i)}\mathbf{F}_{22,(i)}^{-1} Z_{2,i}^T F_i Z_{2,i}\mathbf{F}_{22,(i)}^{-1}\tilde{\mathbf{F}}_{21,(i)} = O(nh^4).
\end{aligned}$$

Making use of these results one directly finds (13).

The Sequential Model

Due to the factorization (19) of the likelihood, the results above directly transfer to sequential models. It is however not ad hoc obvious that (22) and (23) hold, since in contrast to above, the likelihood (19) is not standard due to d_{ir} which is random. We show however, that this does not affect the asymptotic variance. Let $l_{\eta,i} = \sum_{r=1}^q d_{ir} l_{\eta,ir}$ where as above we neglect listing of the parameters if we refer to the true parameters. It is easily seen that $E\{l_{\eta,i}\} = 0$ which follows since $E\{d_{ir} l_{\eta,ir} | d_{ir} = 0\} = 0$ and by definition

$$E(d_{ir} l_{\eta,ir} | d_{ir} = 1) \quad = \quad \frac{\partial F(\eta_{ir})}{\partial \eta}\operatorname{var}(y_{ir}|d_{ir} = 1)^{-1}\{P(y_{ir} = 1|d_{ir} = 1) - \tilde{\pi}_{ir}\} = 0. \tag{26}$$

In the same fashion it is shown below that the Fisher matrix contributions can be estimated by

$$F_i = E(l_{\eta,i} l_{\eta,i}) \quad = \quad \sum_i \sum_{r=1}^q d_{ir} F_{ir} \tag{27}$$

with $F_{ir} = \partial F(\eta_{ir})/\partial\eta_{ir}\text{var}(y_{ir}|\eta_{ir}, d_{ir} = 1)^{-1}\partial F(\eta_{ir})/\partial\eta_{ir}$ where $\eta_{ir} = \beta_{0r}+z_{1,ir}\beta_1+\gamma_0(x_i)+z_{2,ir}\gamma_z(x_i)$. Thus, the multinomial ordinal response, modeled as sequential model, can be treated as univariate binary response resulting from the transition model. Formula (27) follows since

$$\begin{aligned} E\{\sum_{r=1}^{q}\sum_{s=1}^{q} d_{ir}d_{is}l_{\eta,ir}l_{\eta,ir}\} &= \sum_{r=1}^{q}\sum_{s=1}^{q} E_{d_{ir}\ d_{is}}\left[E\{l_{\eta,ir}l_{\eta,is}|d_{ir} = 1, d_{is} = 1\}\right] \\ &= \sum_{r=1}^{q} E_{d_{ir}}\Big\{E(l_{\eta,ir}l_{\eta,ir}|Y_i \geq r)\Big\} \qquad (28) \\ &\quad +2\sum_{r=1}^{q}\sum_{s>r}^{q} E_{d_{ir}\ d_{is}}\Big\{E(l_{\eta,ir}l_{\eta,is}|Y_i \geq s)\Big\} \end{aligned}$$

where we made use of the fact that $(d_{ir} = 1, d_{is} = 1)$ implies $Y_i \geq max(r, s)$. The first term in (28) consists of elements $E\{l_{\eta,ir}l_{\eta,ir}|Y_i \geq r\} = F_{ir}$. Considering the structure of $l_{\eta,ir}$, as given in the sequel of (20), the components in the second term in (28) become $E\{l_{\eta,ir}l_{\eta,is}|Y_i \geq s\}=-W_{ir}^T\tilde{\pi}_{ir}E\{l_{\eta,is}|Y_i \geq s\}=0$ with $W_{ir} = \partial F(\eta_{ir})/\partial\eta_{ir}\text{var}(y_{ir}|\eta_{ir}, d_{ir} = 1)^{-1}$, where we made use the fact that $y_{ir} = 0$ by definition, given the condition that $Y_i \geq s > r$. It is easily shown that (28) simplifies to $\sum_{r=1}^{q} E_{d_{ir}}\{E\{l_{\eta,ir}l_{\eta,ir}|Y_i \geq r\}=\sum_{r=1}^{q} P(Y_i \geq r|\eta_{ir})F_{ir}$. Since d_{ir} is the empirical estimate for $P(Y_i \geq r|\eta_{ir})$ with $E(d_{ir}) = P(Y_i \geq r|\eta_{ir})$, one obtains (27) as an estimate for the Fisher matrix contribution. Using this property it is now direct to transfer the results for the cumulative model from above to the sequential model.

References

Agresti, A. (1990). *Categorical Data Analysis.* Wiley: New York.

Barnhart, H. X. and Sampson, A. R. (1994). Overview of multinomial models for ordinal data. *Commun. Statist.- Theory Meth.* **23**, 3395–3416.

Cox, C. (1988). Multinomial regression models based on continuation ratios. *Statistics in Medicine* **7**, 435–441.

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

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

Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine* **13**, 1665–1677.

Hall, P. and Titterington, D. M. (1987). On smoothing sparse multinomial data. *Austral. J. Statist* **29**, 19–37.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Hastie, T. and Tibshirani, R. (1993). Varying–coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757–796.

Heagerty, P. and Zeger, S. (1996). Marginal regression models for clustered ordinal measurements. *J. Amer. Statist. Assoc.* **91**, 1024–1036.

Kauermann, G. (1999). Modeling longitudinal data with ordinal response by varying coefficients. *Biometrics*, (to appear).

Kauermann, G. and Tutz, G. (2000). Local likelihood estimation in varying-coefficient models including additive bias correction. *Journal of Nonparametric Statistics*, (to appear).

Läärä, E. and Matthews, J. (1985). The equivalence of two models for ordinal data. *Biometrika* **72**, 206–207.

McCullagh, P. (1980). Regression model for ordinal data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 109–127.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi–likelihood estimation in semiparametric models. *J. Amer. Statist. Assoc.* **89**, 501–511.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20**, 1768–1802.

Simonoff, J. S. (1996). *Smoothing Methods in Statistics.* New York: Springer Verlag.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. Ser. B* **50**, 413–436.

Tutz, G. (1991). Sequential models in categorical regression. *Computational Statistics & Data Analysis* **11**, 275–295.

Tutz, G. (2000). *The Analysis of Categorical Data (in German).* München: Oldenbourg Verlag.

Wild, C. J. and Yee, T. W. (1996). Additive extensions to generalized estimating equation methods. *J. Roy. Statist. Soc. Ser. B* **58**, 711–725.

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *J. Roy. Statist. Soc. Ser. B* **58**, 481–493.

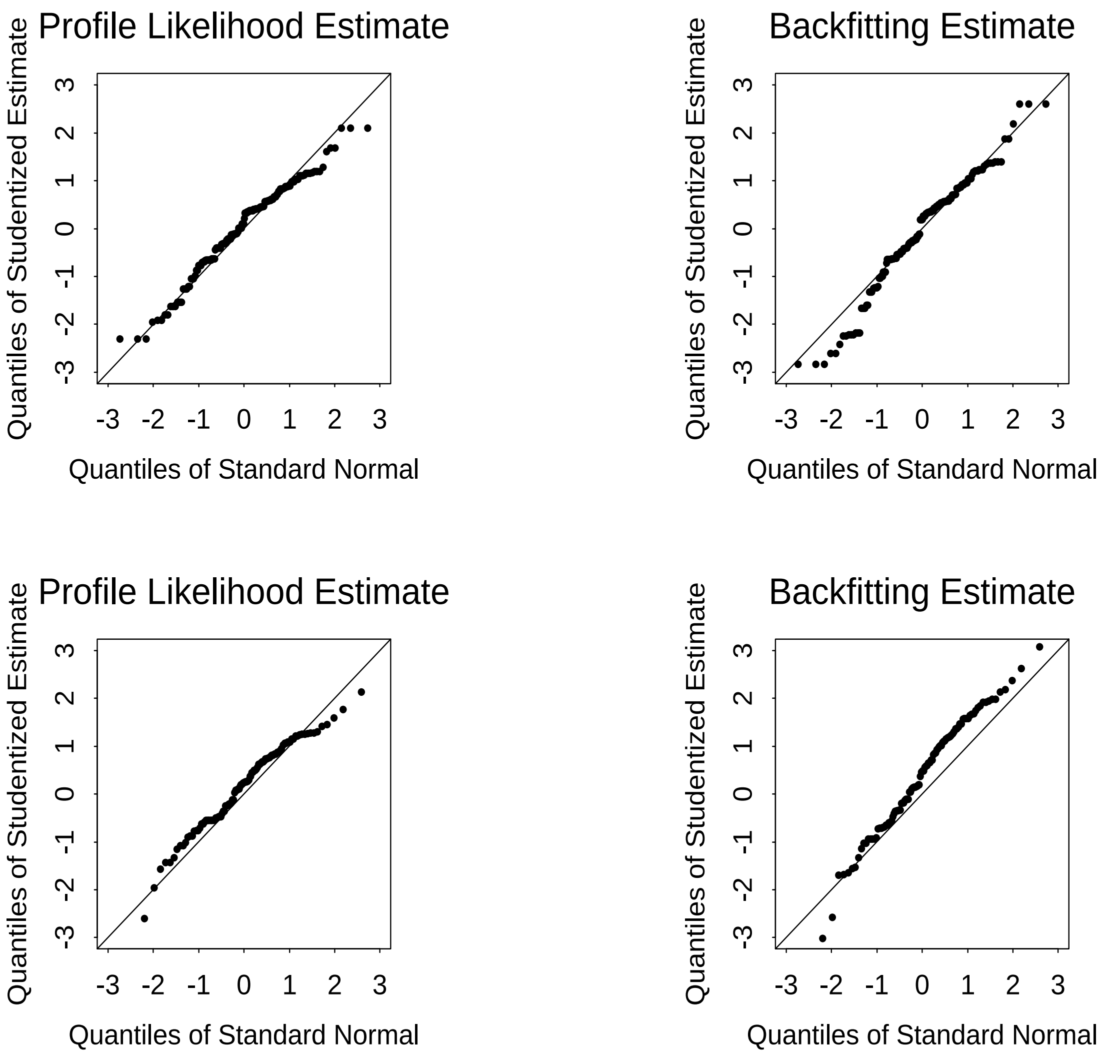

Figure 1: Simulated profile likelihood estimate $\widehat{\beta}_1$ compared with backfitting estimate for balanced design (upper row) and random design (lower row).

	Profile Likelihood Estimates		Backfitting Estimates	
		Balanced Design		
parameter	mean	standard deviation	mean	standard deviation
β_{02}	1.04	0.23	1.02	0.22
β_{03}	2.08	0.34	2.04	0.32
β_1	1.00	0.39	0.97	0.36
		Random Design		
β_{02}	1.00	0.21	0.97	0.20
β_{03}	2.01	0.31	1.97	0.30
β_1	1.04	0.44	1.09	0.40

Table 1: Mean and standard deviation of simulated estimates

coefficient	estimate	standard deviation	studentized value
	Semiparametric Model		
β_{02}	1.28	0.17	7.56
β_{03}	2.30	0.21	11.13
treatment	1.14	0.26	4.67
	Parametric Model		
β_{01}	3.59	1.12	3.18
β_{02}	1.25	0.20	6.35
β_{03}	2.29	0.21	10.90
treatment	1.28	0.23	5.48
age	-0.39	0.08	-4.92
age^2	0.006	0.001	5.30

Table 2: Parameter estimates in the cumulative model

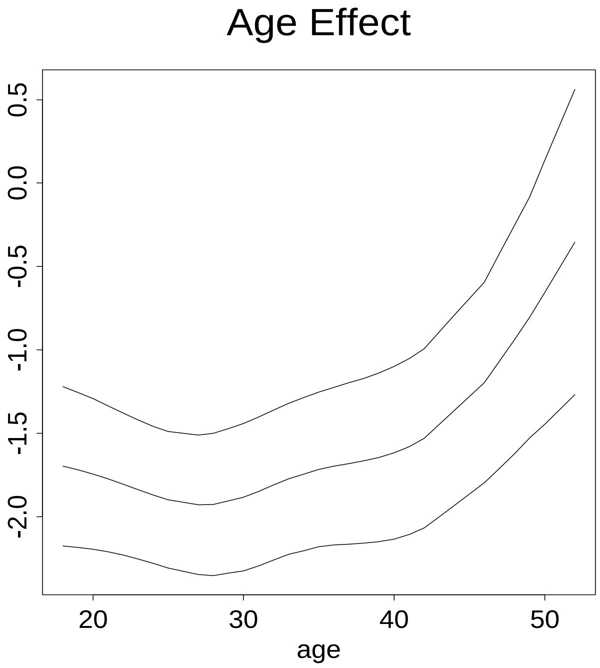

Figure 2: Fitted smooth age effect $\gamma_0(x)$ in the cumulative model

coefficient	estimate	standard deviation	studentized value
	Semiparametric Model		
β_{02}	0.54	0.31	1.72
β_{03}	1.04	0.36	2.92
treatment	0.93	0.28	3.31
	Parametric Model		
β_{01}	1.96	1.73	1.13
β_{02}	0.56	0.20	1.78
β_{03}	1.08	0.21	3.00
treatment	0.94	0.23	3.35
age	-0.27	0.11	-2.26
age^2	0.004	0.001	2.44

Table 3: Parameter estimates in the sequential model

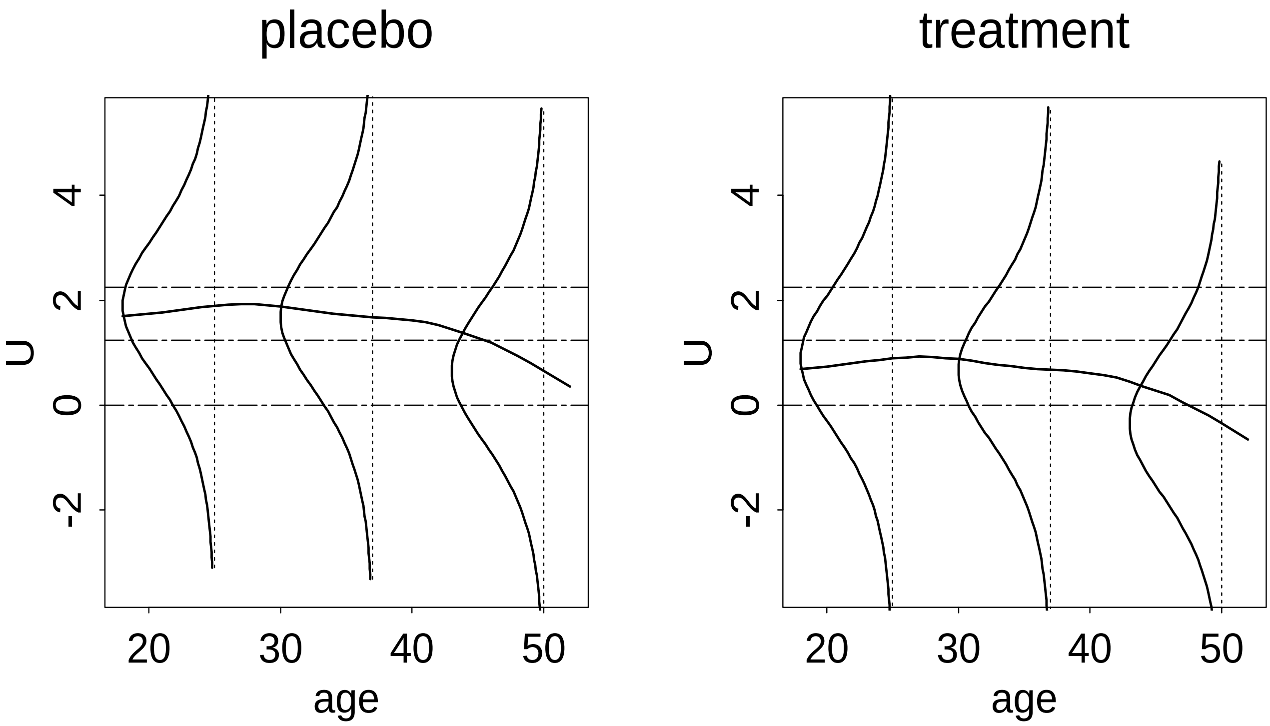

Figure 3: Fitted mean Function $E(\mathcal{U}|z,x)$ with thresholds β_{0r}

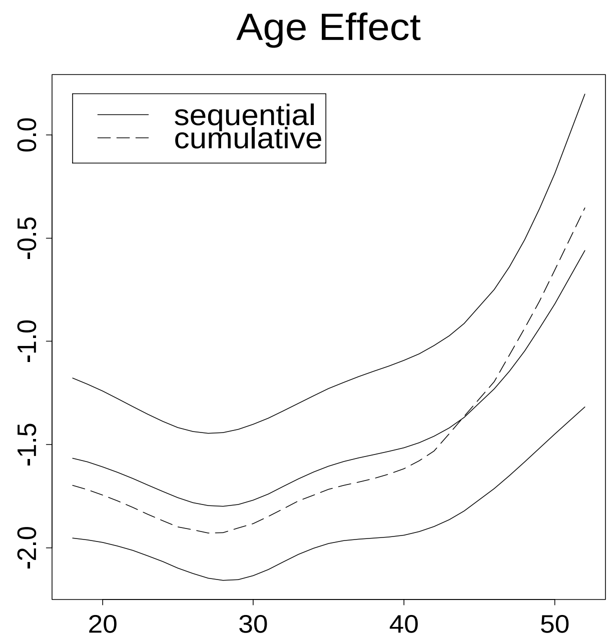

Figure 4: Fitted smooth age effect $\gamma_0(x)$ in the sequential and cumulative model