Kauermann:

Modeling longitudinal data with ordinal response by varying coefficients

Sonderforschungsbereich 386, Paper 144 (1999)

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

Modeling longitudinal data with ordinal response by varying coefficients

Göran Kauermann

Ludwig-Maximilians-University Munich
Institute for Statistics
Akademiestraße 1
80799 Munich, Germany

Abstract

The paper presents a smooth regression model for ordinal data with longitudinal dependence structure. A marginal model with cumulative logit link (McCullagh 1980) is applied to cope for the ordinal scale and the main and covariate effects in the model are allowed to vary with time. Local fitting is pursued and asymptotic properties of the estimates are discussed. A data example demonstrates the exploratory flavor of the smooth model. In a second step, the longitudinal dependence of the observations is considered. Cumulative log odds ratios are fitted locally which provides insight how the dependence of the ordinal observations changes with time.

KEYWORDS: Kernel smoothing, local estimating equations, longitudinal data, marginal model, ordinal data, varying coefficient models.

1 Introduction

Let y_{ir} be the r-th measurement taken together with covariates x_{ir} on the i-th individual at timepoint t_{ir}, where $i = 1, \ldots, n$ and $r = 1, \ldots, n_i$. A convenient model for the mean response at time point t_{ir} is the marginal model

$$E(y_{ir}|t_{ir}, x_{ir}) \quad = \quad h\{z(t_{ir}, x_{ir})\beta\} \tag{1}$$

where the covariates are linked to the mean response via the link function $h(\cdot)$. The design matrix $z(t_{ir}, x_{ir})$ in (1) is allowed to depend on both, the time t and the covariates x. This accommodates time variation as well as interactive time covariate effects. For instance in the linear interaction model

$$E(y_{ir}|t_{ir}, x_{ir}) \quad = \quad h(\beta_0 + t_{ir}\beta_t + x_{ir}\beta_x + t_{ir}x_{ir}\beta_{tx}),$$

as special case of (1), time enters as linear shift and the effect of the covariates changes linearly with time. A priori it is however unknown how time enters the model, i.e. how main and covariate effects vary with time. Moreover, a solely parametric model can hide complex interaction structures which are not represented by simple parametric functions. Therefore it seems desirable to extend (1) in that time enters the model nonparametrically. This is fulfilled by modeling

$$E(y_{ir}|t_{ir}, x_{ir}) \quad = \quad h\{\beta_0(t_{ir}) + x_{ir}\beta_x(t_{ir})\} \tag{2}$$

where $\beta_0(t)$ is a smooth function in time, i.e. the smooth main effect, and $\beta_x(t)$ is the covariate effect which is allowed to vary smoothly with time. Models of type (2) have been introduced by Hastie & Tibshirani (1993) as varying coefficient models. The focus of this paper is to discuss model (2) for longitudinal data with ordinal response variable.

We assume in the following that the response y_{ir} takes values $1, \dots, q+1$ which allow for an ordered interpretation. A widespread model for ordinal data is the cumulative model as introduced by McCullagh (1980). As varying coefficient model this is written as

$$P(y_{ir} \leq k | x_{ir}, t_{ir}) \quad = \quad F\{\beta_{0k}(t_{ir}) + x_{ir}\beta_x(t_{ir})\} \tag{3}$$

for $k = 1, \dots, q$ and $F(\cdot)$ as known continuous distribution function. Frequently $F(\cdot)$ is chosen as logistic distribution function. The q main effects $\beta_{0k}(t)$ are smooth functions fulfilling the restrictions $\beta_{0k}(\cdot) \leq \beta_{0k+1}(\cdot)$. As previously $\beta_x(t)$ gives the covariate effect which is allowed to vary smoothly with time. It should be noted that (3) has a rather general form since no parametric specification is made for the influence of time.

We apply the varying coefficient model (3) to analyze data collected at patients suffering from prostate cancer. The patients were treated with radiation, which was given in three different dose levels. The patients were observed over a five years follow up, where drop out effects were tested but did not occur significant. As response variable we consider the severeness of side effects of the therapy, like pain or bleeding, which is measured on an ordinal scale. One of the objective of study was to investigate how the dose of radiation affects the severeness of side effects and moreover, if and how this effect varies over the time of follow up. We analyze this point by fitting model (3) to the data and considering the shape of the covariate effect $\beta_x(t)$.

Estimation of model (3) is done by local estimating equations, see e.g. Carroll, Ruppert & Welsh (1988). In the setting considered here local estimation can be seen as a weighted generalized estimating equation (GEE) with working independence used in the fitting. In a solely parametric framework working independence is known to provide consistent but not necessarily efficient estimates. For smooth estimation however efficiency arguments are less focussed than bias-variance trade-off properties. One reason for this is that the asymptotic order of the bias of smooth estimates typically dominates the parametric bias order. The bias-variance trade-off in turn guarantees consistent estimates. For longitudinal data it also has to take the time dependence of the observations into account. We therefore apply a "leaving out one individual" cross validation as suggested by (Rice & Silverman 1991).

At a second step of the analysis we consider the longitudinal dependence structure of the observations in more detail. We suggest a smooth modeling by allowing the dependence between two observations y_{ir} and y_{is} to vary smoothly with the time lag $|t_{ir} - t_{is}|$. Moreover, the longitudinal dependence is allowed to depend on additional covariates, where their effect may also vary with the time lag $|t_{ir} - t_{is}|$. The longitudinal dependence here is modeled by cumulative log odds ratios which preserves the ordinal structure. Local fitting finally allows for further insight in the time variation.

An overview about parametric models for longitudinal data is found in Diggle, Liang & Zeger (1994). Parametric marginal models of type (1) for binary and categorical response are treated for instance in Liang & Zeger (1986), Prentice (1988), Liang, Zeger & Qaqish (1992) or Fitzmaurice, Molenberghs & Lipsitz (1995). Generalized estimating equations (GEE) are thereby of common use and extensions for ordinal response variables have been suggested by Heagerty & Zeger (1996) and Fahrmeir & Pritscher (1996). Smooth estimation for continuous longitudinal response with time as single covariate is treated e.g. in Hart & Wehrly (1986) or Rice & Silverman (1991). Moyeed & Diggle (1994) and Zeger & Diggle (1994) apply semi-parametric modeling to longitudinal data with continuous response and additional covariates. Models for correlated categorical data with smooth components are proposed in Wild & Yee (1996), Gieger (1997) or Fahrmeir, Gieger & Heumann (1999). The first paper focus on smooth additive components while Gieger and Fahrmeir et al. also consider varying coefficients. All three papers apply spline fitting in a GEE framework while we here concentrate on local estimation. The latter allows for asymptotic consideration of the estimates including bandwidth selection. Moreover it provides a simple fitting routine if the timepoints of measurements t_{ir} are not grouped and if the cluster size n_i differs among the individuals.

2 Marginal Varying Coefficient Model

2.1 Local Estimation

We rewrite model (3) in matrix form. Let $\tilde{y}_{ir} = (\tilde{y}_{ir,1}, \ldots, \tilde{y}_{ir,q})^T$ be the indicator vector with elements $\tilde{y}_{ir,k} = 1$ if $y_{ir} = k$ and $\tilde{y}_{ir,k} = 0$ otherwise. This yields the vector of cell probabilities $\mu_{ir} = E(\tilde{y}_{ir}|x_{ir}, t_{ir})$. The cumulative model (3) is now written as

$$\mu_{ir}\{\beta(t_{ir})\} \quad = \quad E(\tilde{y}_{ir}|x_{ir}, t_{ir}) = h\{Z_{ir}\beta(t_{ir})\} = h(\eta) \tag{4}$$

with

$$Z_{ir} = \begin{pmatrix} 1 & 0 & \cdots & 0 & x_{ir} \\ 0 & 1 & \cdots & 0 & x_{ir} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_{ir} \end{pmatrix} \quad \text{and } \beta(t_{ir}) = \begin{pmatrix} \beta_{01}(t_{ir}) \\ \vdots \\ \beta_{0q}(t_{ir}) \\ \beta_x(t_{ir}) \end{pmatrix}.$$

The link function $h()$ is invertible where the k-th component of $h^{-1}()$ equals $F^{-1}\{P_{ir}(k)\}$ for $k = 1 \ldots q$ with $P_{ir}(k)$ abbreviating $P(y_{ir} \leq k | x_{ir}, t_{ir})$.

We estimate the varying coefficients $\beta(\cdot)$ in (4) by local estimation. Let $\omega_{ir,0}$ denote the kernel weight $K\{(t_{ir} - t_0)/\lambda\}$ with $K(\cdot)$ as unimodal kernel density function and λ as smoothing parameter. At a target point t_0 we obtain the estimate $\hat{\beta}(t_0)$ by solving the local estimating equation

$$0 \;=\; \sum_i^n \sum_r^{n_i} \omega_{ir,0} Z_{ir}^T \frac{\partial h\{Z_{ir}\hat{\beta}(t_0)\}}{\partial \eta}^T \mathrm{Var}(\tilde{y}_{ir})^{-1} \left[\tilde{y}_{ir} - \mu_{ir}\{\hat{\beta}(t_0)\}\right]. \tag{5}$$

The solution of (5) does not necessarily provide a valid estimate since it is not guaranteed that $\hat{\beta}_{0k}(t_0) \leq \hat{\beta}_{0k+1}(t_0)$. Fahrmeir & Tutz (1994) suggest the simple reparameterization $\theta_{01} := \beta_{01}, \theta_{0k} := \log(\beta_{0k} - \beta_{0k-1})$ for $k = 2, \ldots, q$ to overcome this point. For simplicity of notation however we neglect this reparameterization in the sequel.

For notational simplicity we abbreviate the component $Z_{ir}^T \partial h\{Z_{ir}\beta(t)\}/\partial \eta^T \mathrm{Var}(\tilde{y}_{ir})^{-1}$ in (5) by $W_{ir}^T\{\beta(t)\}$ in the following. Asymptotic properties of the estimate $\hat{\beta}(t_0)$ can be derived by expanding (5) about the true value $\beta(t_0)$. As shown in the appendix in first order approximation this leads to

$$\begin{aligned} \hat{\beta}(t_0) - \beta(t_0) \;&=\; \left\{\sum_i^n \sum_r^{n_i} \omega_{ir,0} W_{ir}^T \mathrm{Var}(\tilde{y}_{ir}) W_{ir}\right\}^{-1} \\ &\quad \times \left[\left\{\sum_i^n \sum_r^{n_i} \omega_{ir,0} W_{ir}^T \{\tilde{y}_{ir} - \mu_{ir}(\beta(t_{ir}))\}\right\} + b(t)\right] \end{aligned} \tag{6}$$

where $W_{ir} = W_{ir}\{\beta(t_{ir})\}$. The component $b(t)$ contains the dominating part of the bias which can be approximated by $\sum_i^n \sum_r^{n_i} \omega_{ir,0} W_{ir}^T\{\beta(t_0)\}\{\mu_{ir}(\beta(t_{ir})) - \mu_{ir}(\beta(t_0))\}$. It appears that the bias is not affected by the correlation structure which corresponds to results given e.g. in Hart (1991). In the appendix it is shown that under general regularity conditions one obtains

$$E\{\hat{\beta}(t_0) - \beta(t_0)\} = O(\lambda^2) \tag{7}$$

$$\begin{aligned}\operatorname{Var}\{\hat{\beta}(t_0)\} = & \left\{\sum_i^n \sum_r^{n_i} \omega_{ir,0} W_{ir}^T \operatorname{Var}(\tilde{y}_{ir}) W_{ir}\right\}^{-1} \\ & \times \left\{\sum_i^n \sum_r^{n_i} \sum_s^{n_i} \omega_{ir,0}\omega_{is,0} W_{ir}^T \operatorname{Cov}(\tilde{y}_{ir}, \tilde{y}_{is}^T) W_{is}\right\} \\ & \times \left\{\sum_i^n \sum_r^{n_i} \omega_{ir,0} W_{ir}^T \operatorname{Var}(\tilde{y}_{ir}) W_{ir}\right\}^{-1}. \end{aligned} \tag{8}$$

The inner part of the variance (8) reflects the correlation between observations taken at one individual. The variance has order $O(n^{-1})$ so that $\hat{\beta}(t_0)$ is consistent for smoothing parameter $\lambda \to 0$ and sample size $n \to \infty$ (see appendix).

Equation (5) can be seen as a weighted generalized estimating equation (GEE) with independence assumed as working correlation. In a solely parametric setting this is known to provide consistent but not necessarily efficient estimates (see e.g. Liang & Zeger 1986). In the smoothing context however efficiency is less focused. Instead bias-variance trade-off properties are of primary interest. This is because the bias of smooth estimates usually has order $O(\lambda^2)$ which dominates the standard parametric bias order $O(n^{-1})$. It is therefore necessary to select the bandwidth λ such that the mean squared error of the estimates is minimized. For dependent data this approach is particularly relevant since the weights $\omega_{ir,0}$ in (5) have to take both into account, the smoothness of $\beta(t)$ and the correlation among the observations. A simple routine for selecting a suitable bandwith for dependent data is a "leaving out one individual " cross validation as suggested by Rice & Silverman (1991). Let therefore $d(y_{ir}, \mu_{ir})$ denote the Kullback-Leibler distance defined by $d(y_{ir}, \mu_{ir}) = \sum_k^{q+1} \tilde{y}_{ir,k} \log P(y_{ir} = k)$. The bandwidth λ is then selected by maximizing the cross validation function

$$cvl(\lambda) \quad = \quad \sum_i \sum_r^{n_i} d[y_{ir}, \mu_{ir}\{\beta^{-i}(t_{ir})\}] \tag{9}$$

with $\widehat{\beta}^{-i}(t_{ir})$ solving (5) by neglecting all observations from the i-th individual. It is shown in the appendix that this approach tends to minimize the integrated mean squared error of the estimates and hence automatically takes the correlation among the observations into account.

The cumulative model (4) allows for a further interpretation. The ordinal response y_{ir} can be seen as coarser version of a latent score variable u_{ir}, say, see e.g. Fahrmeir & Tutz (1994). By setting $u_{ir} = -x_{ir}\beta_x(t_{ir}) + \varepsilon_{ir}$ with ε_{ir} latent and distributed according to distribution $F(\cdot)$ in (3), we can interpret the main effects $\beta_{0k}(t_{ir})$ as thresholds. This means we get $y_{ir} = k$ if $\beta_{0k-1}(t_{ir}) < u_{ir} \leq \beta_{0k}(t_{ir})$ for $k = 1, \ldots q$ and $\beta_{00}(t_{ir}) \equiv -\infty$. If $\beta_{0k}(t)$ are parallel curves, i.e. $\beta_{0k}(t) = \beta_0(t) + \theta_k$ for $\theta_k \in \Re$ and $\beta_0(t)$ some smooth function, we can determine the score by $u_{ir} = -\beta_0(t) - x_{ir}\beta_x(t_{ir}) + \varepsilon_{ir}$. This in turn implies the categorization

$$y_{ir} = k \Leftrightarrow \theta_{k-1} < u_{ir} \leq \theta_k \tag{10}$$

with $\theta_0 = -\infty$. The categorization mechanism is now independent of both, covariates x and time t. Taking $F(\cdot)$ as logistic distribution function, property (10) is equivalently expressed by the proportional log odds assumption $\text{logit}\{P(y_{ir} \leq k)\} - \text{logit}\{P(y_{ir} \leq l)\} = const$, where $const$ is a constant depend on k and l only.

Exploratory analysis of the shape of $\beta_{0k}(\cdot)$ therefore allows to investigate whether proportional log odds can be assumed, i.e. whether a categorization like (10) holds.

2.2 Example

We analyse data collected at the University of Chicago Hospitals. Patients with prostate cancer were treated with radiation, where one of three dose levels (D) of radiation was given to each patient. Further covariates are the stage of the tumor at the beginning of the therapy (S, with three levels) and the hospital in which the patient was treated and followed up (H, two hospitals). In each of the two hospitals a physician assessed the side effects of the radiation therapy on the ordinal scale "no problems" (y=1), "minor problems" like pain (y=2) and "severe problems" like bleeding (y=3). All assessments in the corresponding hospital were made by the same doctor so that the hospital effect can also be interpreted as a physician effect which compensates the subjective character of the response variable. The patients ($n = 196$) were followed up over 5 years, roughly three to five visits a year. The timepoints of measurement (t, measured in months) thereby differ from patient to patient. If a patient did not visit the doctor at least once every half year, subsequent information was neglected to avoid intermediate drop out effects. We model dose D and stage S linearly which leads to the varying coefficient model

$$P(y \leq k|D, S, H, t) \quad = \quad F\{\beta_{0k}(t) + D\beta_D(t) + S\beta_S(t) + H\beta_H(t)\} \qquad (11)$$

for $k = 1, 2$ where $F()$ is chosen as logistic distribution function.

We assume a missing completely at random drop out process (p-value 0.72 when testing grouped data against missing at random, see e.g. Diggle 1989) and choose

$\lambda = 10$ as bandwidth for a Gaussian kernel by cross validation. Figure 1 shows the fitted varying coefficients. The confidence bands are calculated from (8) using local sandwich type estimate, i.e. we apply (8) with plug-in estimates and replace the covariance $\mathrm{Cov}(\tilde{y}_{ir}, \tilde{y}_{is})$ by its empirical version $(\tilde{y}_{ir} - \hat{\mu}_{ir})(\tilde{y}_{is} - \hat{\mu}_{is})^T$ with $\hat{\mu}_{ir} = \mu_{ir}\{\hat{\beta}(t_{ir})\}$. The two main effects are plotted in one plot, upper left plot, to investigate their parallel shape. It appears that proportional log odds can be assumed which means that the response y_{ir} can be seen as classified version of a latent score u_{ir}, classified according to (10). The covariate effects are shown in the remaining three plots. As reference the zero line is given. Beside the fitted coefficients $\hat{\beta}_x(t)$, the bias reduced estimates $\hat{\beta}_x(t) - \hat{b}_x(t)$ are also shown, where $\hat{b}_x(t)$ is a plug in estimate of the corresponding subvector of the bias $b(t)$.

Stage and hospital effect do not show substantial time variation. In contrast, the dose effect clearly varies over time and after about three years a high dose therapy leads to an increase of the side effects. This effect becomes also visible from Figure 2 where the proportion of patients with side effects ($y = 1$ or 2 for minor and severe side effects) is plotted for different subgroups of the patients. The two right plots of Figure 2 extract two groups from the left plot but now with .9 confidence bands being added. The effect of dose varies over time and separates the selected groups after about 3 years of follow up.

3 Longitudinal Dependence Structure

3.1 Local Estimation

We next consider the longitudinal dependence of the observations in more detail. To accompany the ordinal scale of the response we model cumulative log odds ratios as suggested by Heagerty & Zeger (1996) or Fahrmeir & Pritscher (1996) in a solely parametric framework. For variables y_{ir} and y_{is} let

$$\vartheta_{irs}^{kl} = \log \left\{ \frac{P(y_{ir} \leq k, y_{is} \leq l) P(y_{ir} > k, y_{is} > l)}{P(y_{ir} \leq k, y_{is} > l) P(y_{ir} > k, y_{is} \leq l)} \right\} \tag{12}$$

define the cumulative log odds ratios, where $k, l = 1, \ldots, q$. We assume that the log odds ratios depend on some (time constant) covariates x_i, say, and we allow the resulting effect to vary smoothly with the time lag $\Delta t = |t_{ir} - t_{is}|$. For instance in the above example we model ϑ_{irs}^{kl} to depend on the hospital where the resulting hospital effect may vary smoothly in Δt. We generally set $\vartheta_{irs}^{kl} = \alpha_0^{kl}(\Delta t) + x_i \alpha_x^{kl}(\Delta t) =: (1, x_i)\alpha^{kl}(\Delta t)$ with $\alpha^{kl}(\cdot) = (\alpha_0^{kl}(\cdot)^T, \alpha_x^{kl}(\cdot)^T)^T$. Here $\alpha_0^{kl}(\Delta t)$ serves as smooth main effect on the longitudinal dependence and $\alpha_x^{kl}(\Delta t)$ is the covariate effect which may change for different timelags.

The diagonal elements ϑ_{irs}^{kk} of (12), i.e. for $k = l$, correspond to the covariance $\mathrm{Cov}\{\delta(y_{ir} \leq k), \delta(y_{is} \leq k)\}$ where $\delta(y_{ir} \leq k)$ equals 1 for $y_{ir} \leq k$ and 0 otherwise. Due to the longitudinal structure it appears natural to assume that the diagonal log odds ratios ϑ_{irs}^{kk} will decrease monotonely for increasing time lag Δt. This assumption however seems not justified for off diagonal element ϑ_{irs}^{kl} with $k \neq l$. Smooth fitting of $\alpha^{kl}(\Delta t)$ will therefore provide exploratory insight in the longitudinal dependence structure.

We write the vector of log odds ratios as $\vartheta_{irs} = (\vartheta_{irs}^{11}\ , \ldots , \vartheta_{irs}^{1q}\ , \vartheta_{irs}^{21}\ , \vartheta_{irs}^{qq})$ where the lexicographical order used to get ϑ_{irs} is also used for the vectors defined below. Let μ_{irs} be the vector of cell probabilities with components $\mu_{irs}^{kl} = P(y_{ir} = k, y_{is} = l)$ and let P_{irs} denote the joint probability vector with elements $P_{irs}(k,l) = P(y_{ir} \leq k, y_{is} \leq l)$. We obtain μ_{irs} from P_{irs} by $\mu_{irs} = BP_{irs}$ with B as matching contrast matrix. Moreover $P_{irs}(k,l)$ is obtained from the marginal distributions $P_{ir}(k)$ and $P_{is}(l)$ and the log odds ratio ϑ_{irs}^{kl} by the link

$$P(Y_{ir} \leq k, Y_{is} \leq l) \quad = \quad g\{\vartheta_{irs}^{kl}; P_{ir}(k), P_{is}(l)\}, \text{ for } k,l = 1, \ldots q.$$

The function $g()$ is available analytically and given for instance in Palmgren (1989) or Diggle, Liang & Zeger (1994, p.150). With $\tilde{g}(\vartheta_{irs}, P_{ir},, P_{is})$ we define the vector valued link function with components $g\{\vartheta_{irs}^{kl}; P_{ir}(k), P_{is}(l)\}$. Finally, v_{irs} denotes the vector of the centered products $(\tilde{y}_{ir,k} - \mu_{ir,k})(\tilde{y}_{is,l} - \mu_{is,l})$.

At a given time lag Δt_0 we estimate $\alpha(\Delta t_0) = (\alpha^{11}(\Delta t_0)^T, \ldots, \alpha^{qq}(\Delta t_0)^T)^T$ by local estimation. Let therefore $\omega_{irs,\Delta t_0} = K\{(|t_{ir} - t_{is}| - \Delta t_0)/\gamma\}$ denote some kernel weights with γ as second smoothing parameter. The estimate $\hat{\alpha}(\Delta t_0)$ is obtained by solving the weighted estimating equation

$$0 \quad = \quad \sum_{i}^{n} \sum_{r}^{n_i} \sum_{s>r}^{n_i} \omega_{irs,\Delta t_0} V_{irs}^T \{v_{irs} - E(v_{irs})\} \tag{13}$$

where $V_{irs}^T = \tilde{Z}_i \partial \tilde{g}(\vartheta_{irs}; P_{ir}, P_{is}) / \partial \vartheta_{irs}^T B^T \mathrm{Var}(v_{irs})^{-1}$. The design matrix $\tilde{Z}_i$ is constructed from $I \otimes (1, x_i)$ with $\otimes$ denoting the Kronecker product. The variance of v_{irs} is obtained from the first four moments of y_{ir} and y_{is}, as described in Heagerty & Zeger (1996). It should be noted that (13) again assumes working independence and in practice, the marginal probabilities P_{ir} in (13) have to be replaced by plug in estimates. Neglecting the additional variability resulting from this plug in substitution we approximate the variance of $\hat{\vartheta}(\Delta t_0)$ by the sandwich formula

$$\begin{aligned}
\mathrm{Var}\{\hat{\vartheta}(\Delta t_0)\} \approx & \left\{\sum_i^n \sum_r^{n_i} \sum_{s>r}^{n_i} \omega_{irs,\Delta t_0} V_{irs}^T \mathrm{Var}(v_{irs}) V_{irs}\right\}^{-1} \\
& \times \left\{\sum_i^n \sum_r^{n_i} \sum_{s>r}^{n_i} \sum_l^{n_i} \sum_{k>l}^{n_i} \omega_{irs,\Delta t_0} \omega_{ilk,\Delta t_0} V_{irs}^T \mathrm{Cov}(v_{irs}, v_{ilk}) V_{ilk}\right\} \\
& \times \left\{\sum_i^n \sum_r^{n_i} \sum_{s>r}^{n_i} \omega_{irs,\Delta t_0} V_{irs}^T \mathrm{Var}(v_{irs}) V_{irs}\right\}^{-1}. \qquad (14)
\end{aligned}$$

3.2 Example

We continue with the prostate cancer example from above. The log odds ratios ϑ_{irs}^{kl} are modelled as

$$\vartheta_{irs}^{kl} = \alpha_0^{kl}(\Delta t) + H\alpha_H^{kl}(\Delta t)$$

with H as hospital indicator. In Figure 3 we show the four log odds ratios based on a smooth fit using a Gaussian kernel with bandwidth γ fixed at value 6. For estimation we considered observations from the first four years of follow up only. The confidence bands are calculated from local sandwich type estimates based on (14) but substituting $\mathrm{Cov}(v_{irs}, v_{ilk})$ by its empirical version. For $k = l = 1$, i.e. upper left plot, a decreasing longitudinal dependence is observed. This corresponds to the preliminary consideration in that the correlation between the observations $\delta(y_{ir} \leq k)$ and $\delta(y_{is} \leq k)$ should be decreasing for increasing time lag. Moreover the longitudinal dependence does not differ in the two hospitals. The off-diagonal plots show a rather time stable dependence, in particular for Hospital 2. Hospital 1 has less patients ($n = 75$) and more extreme probabilities which explains the larger variability of the estimates. Finally we consider the lower right plot where

$k = l = 2$. The longitudinal dependence again shows a decreasing shape, but now the two hospital distinguish. For Hospital 1 the longitudinal dependence is stronger than for Hospital 2. This means that in Hospital 1 patients with severe side effects are more likely to be classified again as severe side effect patients at a subsequent timepoint than in Hospital 2. An explanation for this might lie in the subjective character of the measurement. As mentioned, in each hospital the same doctor assessed the patients over the follow up so that the hospital effect corresponds to a physician effect. In Hospital 1 the doctor seems to take previous assessments of the patient, in particular recorded severe side effects, more into account when assessing a patient than the doctor in Hospital 2 does. The effect fades away after about 1 1/2 years time lag.

4 Conclusions

We applied local estimation to fit a marginal model including its longitudinal dependence structure. The fitting procedure is numerically simple, e.g. for fitting the marginal model of Section 2 standard software which accomodates fitting weighted observations can be used. The smooth fits provide exploratory insight in the longitudinal mean and dependence structure. This in turn can also help building appropriate parametric models.

Acknowledgements

The author likes to thank Dr. Srinivasan Vijayakumar for letting him use the data

and Florin Vaida for the cooperation and the fruitful discussions.

A Technical Details

Asymptotical Behavior

We base our asymptotic consideration on an increasing sample size n, i.e. we assume that the number of independent individuals tends to infinity. In particular this implies that the correlation structure among the observations taken at one individual does not affect the asymptotic behavior. Moreover, we assume that the number of observations n_i for the i-th individual and time-points t_{ir} are independent, where n_i follows some discrete distribution with finite moments and t_{ir}, $r = 1, \dots n_i$, are independently distributed according to density $f(t)$. The support of $f()$ is supposed to be bounded and connected. This transfers standard assumptions for independent data to the repeated measurement case. Let $s_{ir}\{\beta(t)\} = Z_{ir}^T \partial h\{Z_{ir}\beta(t)\}/\partial\eta^T \operatorname{Var}(\tilde{y}_{ir})^{-1}[\tilde{y}_{ir} - \mu_{ir}\{\beta(t)\}]$ be the contribution to the local estimating equation (5). Expanding $0 = \sum_i \sum_r \omega_{ir,0} s_{ir}\{\hat{\beta}(t_0)\}$ about $\beta(t_0)$ yields in first order approximation

$$\hat{\beta}(t_0) - \beta(t_0) \;=\; \left(\sum_i \sum_r \omega_{ir,0} F_{ir}\right)^{-1} \left[\sum_i \sum_r s_{ir}\{\beta(t_{ir})\} + \tilde{b}(t_0)\right]$$

where $F_{ir} = -E[\partial s_{ir}\{\beta(t_{ir})\}/\partial\beta]$. The component $\tilde{b}(t_0)$ decomposes to

$$\begin{aligned}\tilde{b}(t_0) \;=\; & \sum_i \sum_r \Big[\Big\{W_{ir}(\beta(t_0)) - W_{ir}(\beta(t_{ir}))\Big\}\Big\{\tilde{y}_{ir} - \mu_{ir}(\beta(t_{ir}))\Big\} \\ & + W_{ir}(\beta(t_0))\Big\{\mu_{ir}(\beta(t_{ir})) - \mu_{ir}(\beta(t_0))\Big\}\Big].\end{aligned}$$

The first component has zero expectation so that the second gives the dominating part of the smoothing bias. Denoting the second component by $b(t_0)$ yields the further approximation

$$
\begin{aligned}
b(t_0) &\approx nE(n_i)\int K\left(\frac{t-t_0}{\lambda}\right)W^T(\beta(t_0))\{\mu(\beta(t))-\mu(\beta(t_0))\}f(t)dt \\
&= nE(n_i)\lambda^3\int z^2K(z)dzW^T(\beta(t_0))\Big\{\frac{\partial\mu(\beta(t_0))}{\partial t}f'(t_0)+\frac{\partial^2\mu(\beta(t_0))}{\partial t^2}f(t_0)\Big\}
\end{aligned}
$$

where $W^T(\beta(t_0)) = \int Z^T(x)\partial h\{Z(x)\beta(t_0)\}/\partial\eta(x)^T\mathrm{Var}(\tilde{y})^{-1}f(x|t_0)dx$ is the asymptotic formulation for $\sum_i\sum_r \omega_{ir,0}W_{ir}^T(\beta(t_0))$. Assuming all components involved to be sufficiently smooth, the embraced term in the formula is of finite order, hence $b(t)$ has order $O(n\lambda^3)$. In the same fashion we find that $\sum_i^n\sum_r^{n_i}\omega_{ir,0}W_{ir}^T\mathrm{Var}(\tilde{y}_{ir})W_{ir} = E(n_i)O(n\lambda)$ which in turn proves (7). Finally the inner part of the variance in (8) has order $E(n_i)^2O(n\lambda^2)$ which yields the proposed order of the variance.

Cross Validation

Let $s_{ir} = s_{ir}(\beta(t_{ir}))$ and $F_{(ir)}^{-i} = \sum_{j\neq i}\sum_s \omega_{ir,js}F_{js}$ with $\omega_{ir,js} = K((t_{ir}-t_{js})/\lambda)$. From (6) we get

$$
\widehat{\beta}^{-i}(t_{ir}) - \beta(t_{ir}) = F_{(ir)}^{-i^{-1}}\left\{\sum_{j\neq i}\sum_s^{n_j}\omega_{ir,js}s_{js} + b^{-i}(t_{ir})\right\}
$$

with obvious definition for $b^{-i}(t_{ir})$. In first order approximation the expected cross validation function equals

$$
\begin{aligned}
&E\Big\{\sum_i\sum_r^{n_i} d[y_{ir},\mu\{\widehat{\beta}^{-i}(t_{ir})\}]\Big\} \\
= \;& E\Big[\sum_i\sum_r^{n_i} d[y_{ir},\mu\{\beta(t_{ir})\}]\Big\} + s_{ir}^T\{\widehat{\beta}^{-i}(t_{ir})-\beta(t_{ir})\} \\
&-\frac{1}{2}\{\widehat{\beta}^{-i}(t_{ir})-\beta(t_{ir})\}^TF_{ir}\{\widehat{\beta}^{-i}(t_{ir})-\beta(t_{ir})\}\Big] \\
= \;& const - \frac{1}{2}E\Big[\sum_i\sum_r^{n_i}\{\sum_{j\neq i}\sum_s^{n_j}\omega_{ir,js}s_{js}+b^{-i}(t_{ir})\}^T \\
&\times F_{(ir)}^{-i^{-1}}F_{ir}F_{(ir)}^{-i^{-1}}\{\sum_{j\neq i}\sum_s^{n_j}\omega_{ir,js}s_{js}+b^{-i}(t_{ir})\}\Big]
\end{aligned}
$$

$$= \quad const - \frac{1}{2}\sum_i \sum_r^{n_i} \mathrm{tr}\Big\{ F_{(ir)}^{-i^{-1}} F_{ir} F_{(ir)}^{-i^{-1}} b^{-i}(t_{ir}) b^{-i}(t_{ir})^T \Big\}$$

$$- \sum_i \sum_r^{n_i} \mathrm{tr}\Big\{ \sum_{j\neq i} \sum_s^{n_j} \sum_u^{n_j} \omega_{ir,js}\omega_{ir,ju} F_{(ir)}^{-i^{-1}} F_{ir} F_{(ir)}^{-i^{-1}} \mathrm{Cov}(s_{js}, s_{su}) \Big\}.$$

with tr denotes the trace of a matrix. The first term is determined by the squared bias while the second contains the variance of the estimates $\widehat{\beta}_{ir}^{-i}$, $r = 1, \ldots, n_i$. Hence, maximizing (9) corresponds to minimizing the mean squared error (in mean).

References

Carroll, R. J., Ruppert, D., and Welsh, A. H. (1988). Local estimating equations. *J. Amer. Statist. Assoc. 93*, 214–227.

Diggle, P. (1989). Testing for random dropouts in repeated measurement data. *Biometrics 45*, 1255–1258.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of longitudinal data.* Oxford University Press.

Fahrmeir, L., Gieger, C., and Heumann, C. (1999). An application of isotonic longitudinal marginal regression to monitoring the healing process. *Biometrics*, (to appear).

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

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

Fitzmaurice, G. M., Molenberghs, G., and Lipsitz, S. R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *J. Roy. Statist. Soc. B 57*, 691–704.

Gieger, C. (1997). Non- and semiparametric marginal regression models for ordinal response. SFB Discussion Paper 71, Institut für Statistik, Ludwig-Maximilian-Universität München,
http://www.stat.uni-muenchen.de/sfb386/publikation.html.

Hart, J. D. (1991). Kernel regression estimation with time series error. *J. Roy. Statist. Soc. B, 53*, 173–187.

Hart, J. D. and Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc. 81*, 1080–1088.

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

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

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with discussion). *J. Roy. Statist. Soc. 54*, 2–24.

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

Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Aust. J.Statist. 36*, 75–93.

Palmgren, J. (1989). Regression models for bivariate responses. Technical Report 101, Dep. Biostatist. Univ of Washington, Seattle.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics 44*, 1033–1048.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. B, 53*, 233–243.

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

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics 50*, 689–699.

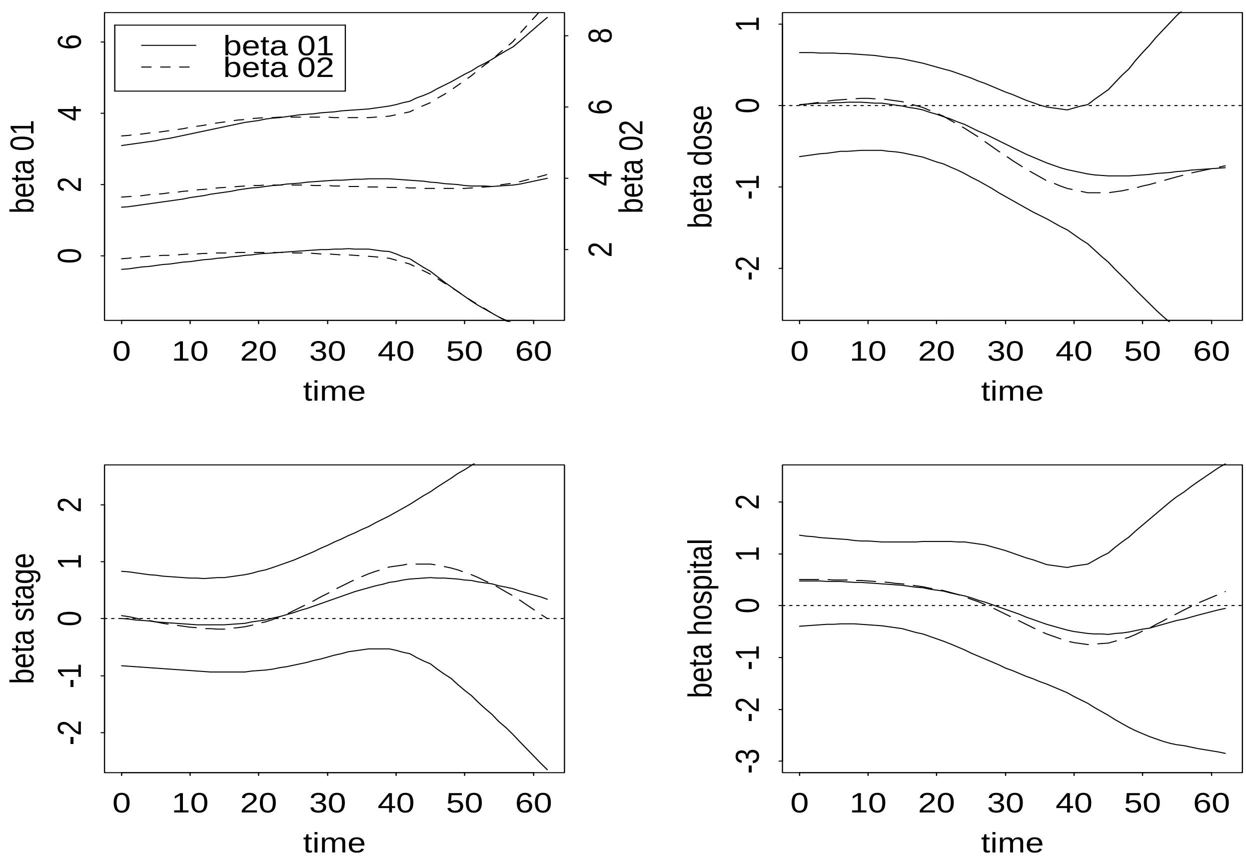

Figure 1: Varying coefficients for radiation data, dashed curves show bias reduced estimates.

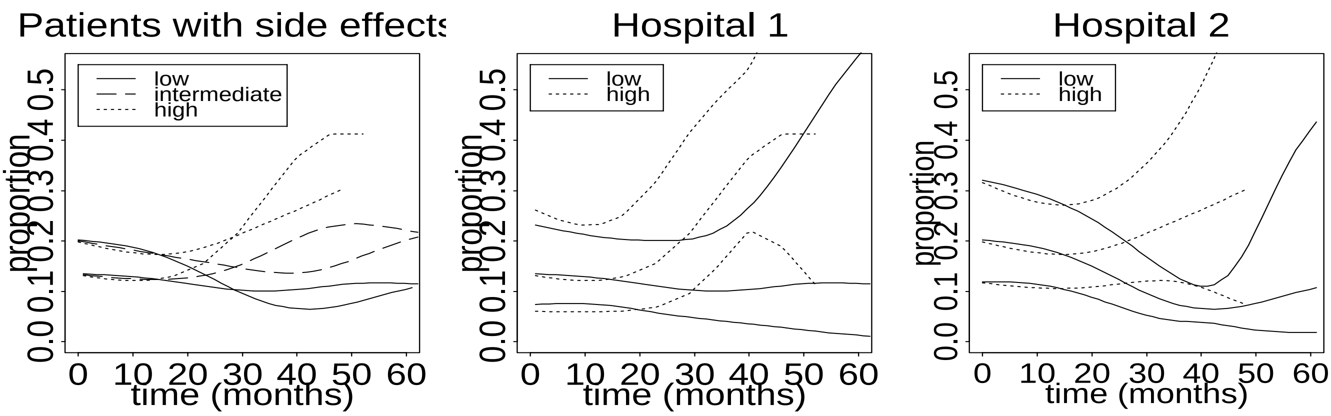

Figure 2: Proportion of Patients with side effects (stage S=intermediate) and three dose levels for both hospitals (left plot). The two right plots show the proportions separatly for the two hospitals and low and high dose levels with .9 confidence bands.

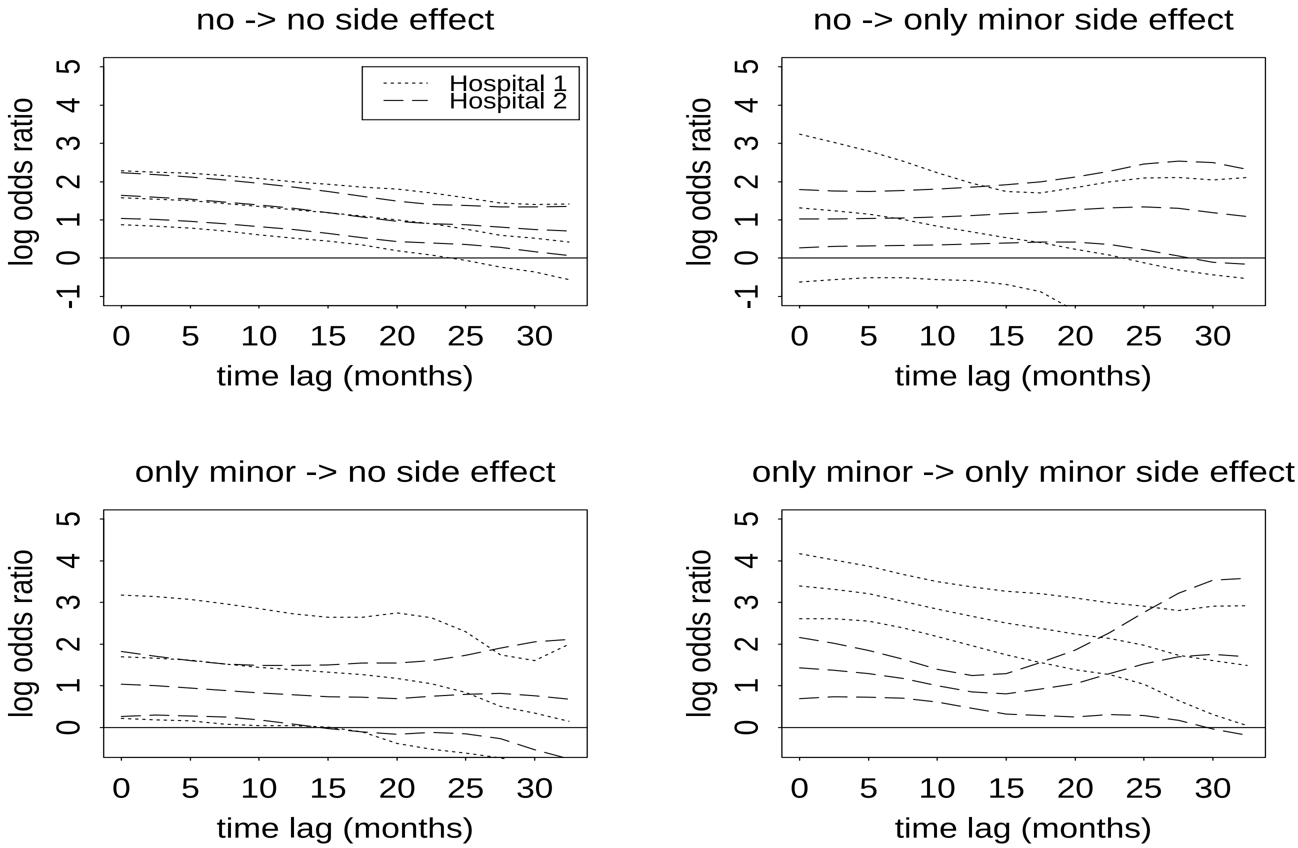

Figure 3: Log odds ratios $\vartheta^{kl}(\Delta t)$ for both hospitals