Fahrmeir, Gieger, Heumann:

An application of isotonic longitudinal marginal regression to monitoring the healing process

Projektpartner

gsf

MAX-PLANCK-GESELLSCHAFT

TUM

An application of isotonic longitudinal marginal regression to monitoring the healing process

Ludwig Fahrmeir[1], Christian Gieger[2] and Christian Heumann[3]

Institut für Statistik, Ludwig-Maximilians-Universität München

Ludwigstr. 33, 80539 München, Germany

Summary

This paper discusses marginal regression for repeated ordinal measurements that are isotonic over time. Such data are often observed in longitudinal studies on healing processes where, due to recovery, the status of patients only improves or stays the same. We show how this prior information can be used to construct appropriate and parsimoniously parametrized marginal models. As a second aspect, we also incorporate nonparametric fitting of covariate effects via a penalized quasi-likelihood or GEE approach. We illustrate our methods by an application to injuries from sporting activities.

Keywords: marginal regression, isotonic ordinal repeated measurements, non-parametric predictors, penalized generalized estimating equations, iterative proportional fitting.

1   Introduction

Marginal regression models for repeated or clustered ordinal measurements have recently been proposed by several authors, e.g. Heagerty and Zeger (1996), Molenberghs and Lesaffre (1994), Fahrmeir and Pritscher (1996). Here, we consider the case where observed response categories are isotonic over time, that is Y_{is} does not have higher rank than Y_{it} for $s < t$ and each individual i. This situation is not uncommon in longitudinal studies on healing processes. Our development has been motivated by a clinical trial conducted at the Technical University of

---

[1] email: fahrmeir@stat.uni-muenchen.de

[2] email: gieger@stat.uni-muenchen.de

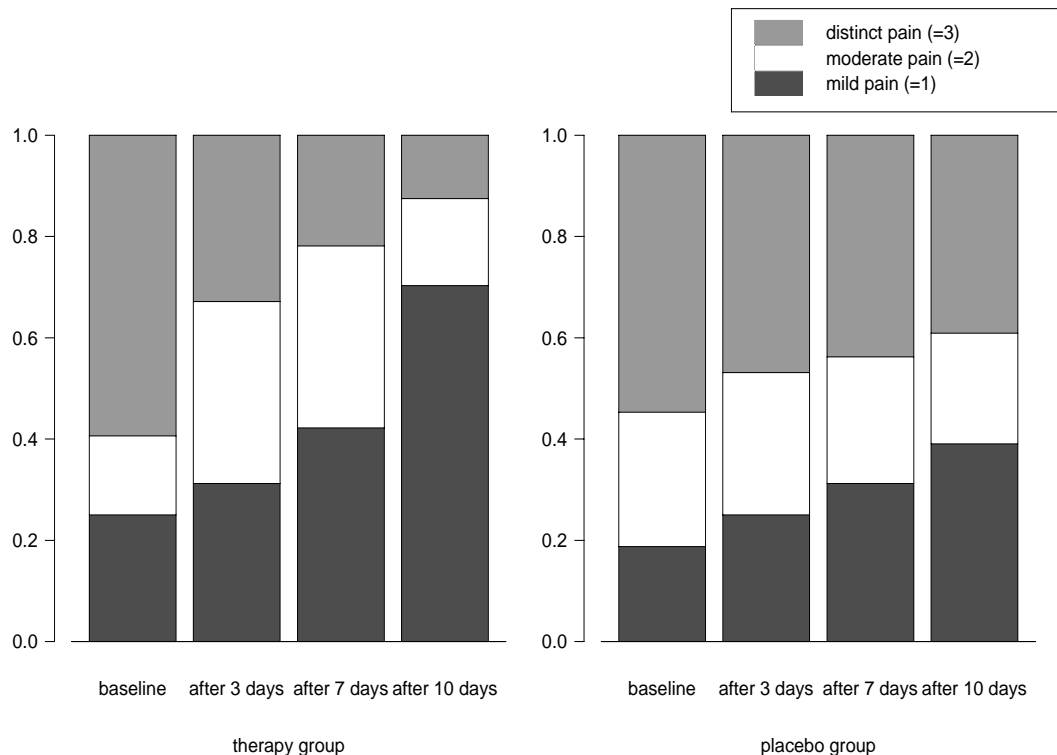[3] email: heumann@stat.uni-muenchen.de

Figure 1: Distribution of "pain from pressure" for both groups.

Munich. The goal in this study was to assess the effect of an antiinflammatory drug on injuries from sporting activities, mostly on legs, compared to placebo. The drug was applied as a spray, containing 10% Ibuprofen, while the placebo was the same spray without Ibuprofen. Altogether 140 patients took part in this double-blind randomized study, 70 patients in each treatment arm. They had visits to the physician prior to the treatment (baseline) and after 3, 7 and 10 days. At all visits, the severity of injuries and the healing process were assessed by several indicators, with "pain from pressure" as the variable of primary interest. It was measured in 5 ordered categories by pressing increasing weights on the injured spot, until it became too painful for the patient. In addition to this response variable, the variables gender, age, height and weight were available. For our analysis, we aggregated the response variable into the three categories "mild pain" (= 1), "moderate pain" (= 2) and "distinct pain" (= 3). We also deleted 13 patients with missing values. All results are therefore based on 127 patients, with 64 patients in the drug group and 63 in the placebo group. Figure 1 shows the distribution of severity of injuries, measured by the response, say Y, "pain from pressure" for both groups.

Figure 1 indicates a global positive healing effect of the therapy compared to placebo as well as an increase of this effect over time. Since the goal of the study was a confirmatory analysis of drug effects, a marginal regression model appears to be an appropriate choice to estimate and test the effect of the therapy.

However, direct application of existing methods for repeated ordinal measurements becomes problematic because of the particular data structure: There is no patient in the sample with increasing "pain from pressure" between two successive visits. The two-dimensional contingency tables below contain corresponding data for responses Y_0 (baseline), Y_1 (visit 1) and Y_2 (visit 2), Y_3 (visit 3):

		Y_1						Y_3			
		1	2	3				1	2	3	
	1	28	0	0	28		1	47	0	0	47
Y_0	2	5	22	0	27	Y_2	2	21	18	0	39
	3	3	19	50	72		3	2	7	32	41
		36	41	50	127			70	25	32	127

The same pattern appears for all pairs (Y_s, Y_t) with $s < t$ and, consequently, for data stratified by covariates. Without taking care of that special data structure, marginal modelling becomes problematic because probabilities corresponding to zeros in contingency tables will either be badly fitted or, if they are correctly estimated close to or by zero, association measures like global or local odds ratios will tend to infinity. This implies serious numerical problems concerning existence and convergence of parameter estimates. In this situation it seems reasonable to assume the same structure for probability tables of pairs (Y_s, Y_t), $s < t$: entries in the northeast corner are assumed to be zero as in Figure 2a, or at least so close to zero that they are better neglected for parsimoniously parametrized modelling. In the following we develop a marginal regression approach that is tailored to this problem.

We will discuss ideas in the context of the concrete study under consideration, but extensions to other and more general settings are obvious. Since the scientific goal was analysis of the marginal response probabilities with association as a nuisance, we also restrict discussion to a GEE1 approach. However, extensions to GEE2 or full likelihood analysis can surely be reasonable in other cases. As an additional feature, we incorporate the possibility of nonparametric modelling

3

a)

Y_s	Y_t 1	2	3	
1	π_{11}	0	0	π_{1+}
2	π_{21}	π_{22}	0	π_{2+}
3	π_{31}	π_{32}	π_{33}	π_{3+}
	π_{+1}	π_{+2}	π_{+3}	1

b)

	Y_t 1	2	3	
1	π_{1+}	0	0	π_{1+}
2	π_{21}	$\pi_{2+}-\pi_{21}$	0	π_{2+}
3	$\pi_{+1}-\pi_{21}-\pi_{1+}$	$\pi_{+2}+\pi_{21}-\pi_{2+}$	$1-\pi_{+1}-\pi_{+2}$	$1-\pi_{1+}-\pi_{2+}$
	π_{+1}	π_{+2}	$1-\pi_{+1}-\pi_{+2}$	1

Figure 2: Pairwise 3 x 3 probability tables: a) Joint probabilities, b) Reparametrized table.

and fitting of components in the predictors for mean or (working) association structures via a penalty approach (see Gieger, 1997). Compared to purely parametric modelling, this allows a refined and more flexible specification of the mean structure and a gain in efficiency due to improved working associations. Wild and Yee (1996) presented an additive extension of generalized estimating equation methods for correlated binary data. Semiparametric modelling of predictors in estimating equations based on local regression has recently been considered by Carroll, Ruppert and Welsh (1997).

2 Regression models

For a GEE1 approach, we have to specify two generalized estimating equations: one for the mean structure, that is for marginal probabilities, and a second for pairwise associations. Reparametrization of joint probabilities in Figure 2a by marginal probabilities and remaining joint probabilities as in Figure 2b shows that a marginal model for $\pi_{l+} = \mathrm{pr}(Y_s = l)$, $l = 1, 2$, and $\pi_{+r} = \mathrm{pr}(Y_t = r)$, $r = 1, 2$, has to be supplemented only by a model for the joint probability $\pi_{21} = \mathrm{pr}(Y_s = 2, Y_t = 1)$ or a corresponding measure of pairwise association.
Preliminary data analysis indicated a positive and time-varying effect of the therapy and a possibly nonlinear effect of age on the patients' status. Effects of other covariates appeared to be negligible in comparison. We first considered the cumulative logistic main effects model

$$\mathrm{logit}\ \mathrm{pr}(Y_{it} \le r) = \theta_{rt} + \beta_1 T H_i^{(1)} + \beta_2 T H_i^{(2)} + \beta_3 T H_i^{(3)} + f(A_i), \qquad (1)$$

$r = 1, 2$, $t = 0, \ldots, 3$, for the mean structure. The time-varying threshold param-

eters model the trend of the healing process in the placebo group. The indicator variables $TH^{(j)}$, $j = 1, 2, 3$ are defined by

$$TH^{(j)} = \begin{cases} 1, & \text{therapy is applied and } t = j, \\ 0, & \text{else,} \end{cases}$$

and are included to account for the additional possibly time-varying effects β_1, β_2, β_3 of therapy compared to placebo. The effect of age A is incorporated additively in form of an unknown smooth function f that will be fitted nonparametrically by a natural cubic smoothing spline. In a further step, we extended this generalized additive model (see Hastie and Tibshirani, 1990) to a varying coefficient model (see Hastie and Tibshirani, 1993) of the form

$$\text{logit pr}(Y_{it} \leq r) = \theta_{rt} + f_1(A)TH_i^{(1)} + f_2(A)TH_i^{(2)} + f_3(A)TH_i^{(3)} + f_4(A_i), \quad (2)$$

$r = 1, 2$, $t = 0, \ldots, 3$. Compared to (1) it additionally allows for interactions between therapy and age in form of nonlinear functions $f_1(A)$, $f_2(A)$ and $f_3(A)$ that are also fitted nonparametrically by cubic smoothing splines. A general form for the marginal mean structure is

$$\text{logit pr}(Y_{it} \leq r) = x'_{itr}\beta + u'_{itr}f(v), \quad (3)$$

where β is the vector of fixed effects, $f(v) = (f_1(v_1), f(v_2), \ldots)'$ is a vector of unknown smooth functions of covariates $v_1, v_2, \ldots$, and x_{itr}, u_{itr} are design vectors constructed from basic covariates.

As a measure of pairwise association, we take the (local) odds ratio

$$\psi_{st} = \frac{\text{pr}(Y_s = 2, Y_t = 1)\,\text{pr}(Y_s = 3, Y_t = 2)}{\text{pr}(Y_s = 2, Y_t = 2)\,\text{pr}(Y_s = 3, Y_t = 1)} = \frac{\pi_{21}\,\pi_{32}}{\pi_{22}\,\pi_{31}}. \quad (4)$$

Together with marginal probabilities this provides an appropriate reparametrization of pairwise joint probabilities.

In our application, we will work with a so-called unspecified working association assumption

$$\log \psi_{ist} = \alpha_{st}, \qquad i = 1, \ldots, n, \ s < t.$$

This is a special case of usual linear parametric models

$$\log \psi_{ist} = \tilde{x}'_{ist}\alpha, \quad (5)$$

with a vector α of association parameters and a design vector $\tilde{x}_{ist}$. Similarly as in (3), an additive nonparametric component $\tilde{u}'_{ist}g(v)$ could be included in (5), see

Gieger (1997) and, for a related idea, Heagerty and Zeger (1997). However, we do not make use of this possibility here and restrict discussion to the parametric model (5).

Semiparametric estimation of the mean structure is based on penalized generalized estimating equations (PGEE; Gieger, 1997). In the following, y_i is the vector of indicator variables $y_{irt} = I(Y_{it} = r)$ for observed categories at visit t, $\pi_i = \pi_i(\beta, f)$ is the corresponding vector of probabilities $\pi_{irt} = \mathrm{pr}(y_{irt} = 1)$ derived from model (3), and f is a generic symbol for the vector of function evaluations or spline basis coefficients. The PGEE is

$$\sum_{i=1}^{N} D_i V_i^{-1}(y_i - \pi_i) - \Lambda \mathrm{P}\gamma = 0, \tag{6}$$

where $\gamma = (\beta', f')'$. The first term has the usual form of GEE's, where $D_i = D_i(\beta, f)$ is the first derivative of $\pi_i(\beta, f)$ with respect to (β, f) and V_i is a working covariance matrix, with elements depending on marginal probabilities as well as on odds ratios and thus on (β, f) as well as on α. The second term is the first derivative of the quadratic penalty term $\Lambda\gamma'\mathrm{P}\gamma$, known from penalized (quasi-)likelihood estimation for cubic smoothing splines. The diagonal matrix Λ contains smoothing parameters and P is a diagonal penalty matrix with a zero on the diagonal if the corresponding parameter is not penalized, e.g. for a fixed effect. To estimate association parameters α together with β and f, we augment (6) as usual by a GEE

$$\sum_{i=1}^{N} C_i U_i^{-1}(w_i - \nu_i) = 0, \tag{7}$$

for association parameters. In (7) w_i is the vector of centered products $w_{ist} = (y_{i2s} - \pi_{i2s})(y_{i1t} - \pi_{i1t})$ and ν_i the vector of corresponding expectations $\mathrm{E}\, w_{ist} = \nu_{ist} = \pi_{i,21,st} - \pi_{i2s}\pi_{i1t}$. Note that we get a very parsimonious parametrization compared to full parametrized model. The joint probability $\pi_{i,21,st} = \mathrm{pr}(Y_{is} = 2, Y_{it} = 1)$ is related to $\log\psi_{ist}$ by (4) and (5). Therefore ν_i is a function of β, f and α. The matrix C_i is the first derivative of ν_i with respect to α and U_i is a further working covariance matrix. As in the binary case (Prentice, 1988) simple but useful choices are $U_i = I$ and $U_i = \mathrm{diag}(\mathrm{var}(w_{ist}))$.

The algorithm for computing estimates $(\hat{\beta}, \hat{f}, \hat{\alpha})$ of this PGEE1 approach can be summarized as follows:

1. Obtain initial values $(\beta^{(0)}, f^{(0)}, \alpha^{(0)})$. One can use $(\beta^{(0)}, f^{(0)})$ resulting from a regression assuming independence and $\alpha^{(0)} = 0$.

2. Use a modified version of the iterative proportional fitting algorithm (IPF), which was originally introduced by Deming and Stephan (1940) and has also been used by Heagerty and Zeger (1996) and others, to obtain the joint probabilities in the bivariate marginal tables. That is, get the current estimates of the local odds ratios, $\psi_{ist}^{(k)}$, from the current estimate $\alpha^{(k)}$ and construct bivariate tables having this odds ratios. In our special case one can use e.g.

$$
\begin{array}{c|ccc}
 & \multicolumn{3}{c}{Y_t} \\
 & 1 & 2 & 3 \\
\hline
1 & p_{ist}^{(k)} & 0 & 0 \\
Y_s \quad 2 & p_{ist}^{(k)}\psi_{ist}^{(k)} & p_{ist}^{(k)} & 0 \\
3 & p_{ist}^{(k)} & p_{ist}^{(k)} & p_{ist}^{(k)}
\end{array}
$$

with $p_{ist}^{(k)} = 1/(5 + \psi_{ist}^{(k)})$ as *initial* tables. Then apply IPF to get tables with marginal probabilities according to the current estimate $(\beta^{(k)}, f^{(k)})$. IPF in general preserves the local odds ratios and automatically accounts for the structural zeros, also in more general cases than the one considered here. The resulting bivariate probabilities can now be used to obtain $V_i^{(k)}, U_i^{(k)}$ and $\nu_i^{(k)}$.

3. Take a (quasi-) Fisher scoring step for $\gamma = (\beta', f')'$:

$$
\gamma^{(k+1)} = \gamma^{(k)} + \sum_{i=1}^{N}\left(D_i V_i^{-1} D_i' + \Lambda\mathrm{P}\right)^{-1}\left(\sum_{i=1}^{N} D_i V_i^{-1}(y_i - \pi_i) - \Lambda\mathrm{P}\gamma^{(k)}\right).
$$

4. Take a (quasi-) Fisher scoring step for α:

$$
\alpha^{(k+1)} = \alpha^{(k)} + \sum_{i=1}^{N}\left(C_i U_i^{-1} C_i'\right)^{-1}\left(\sum_{i=1}^{N} C_i U_i^{-1}(w_i - \nu_i)\right).
$$

5. Iterate, until a specified convergence criterion is fulfilled.

To get a robust approximation for the covariance matrix of the final estimate $(\hat\beta, \hat f)$ we use a nonparametric version of the well–known sandwich matrix.

	main effects model						varying coefficient model		
	independence			unspecified association			unspecified association		
	estim.	standard errors		estim.	standard errors		estim.	standard errors	
		naive	robust		naive	robust		naive	robust
θ_{10}	-1.1342	0.2209	0.2198	-1.2153	0.2250	0.2228	-1.2135	0.2221	0.2208
θ_{20}	-0.0780	0.1875	0.1964	-0.0814	0.1931	0.1924	-0.0958	0.1909	0.1901
θ_{11}	-1.0501	0.2723	0.2747	-0.9717	0.2243	0.2420	-0.9763	0.2206	0.2401
θ_{21}	0.3978	0.2513	0.2577	0.4473	0.2101	0.2122	0.4198	0.2072	0.2078
θ_{12}	-0.7902	0.2644	0.2738	-0.7186	0.2216	0.2527	-0.7056	0.2180	0.2441
θ_{22}	0.5789	0.2528	0.2633	0.6006	0.2155	0.2203	0.5642	0.2122	0.2166
θ_{13}	-0.3058	0.2579	0.2538	-0.2534	0.2254	0.2340	-0.2506	0.2224	0.2298
θ_{23}	0.7130	0.2589	0.2583	0.7997	0.2264	0.2276	0.7620	0.2234	0.2238
β_1	0.4940	0.3347	0.3366	0.3525	0.1809	0.1881			
β_2	0.7735	0.3373	0.3415	0.6818	0.2099	0.2239			
β_3	1.4180	0.3686	0.3685	1.1555	0.2625	0.2661			
α_{01}				0.7341			0.9897		
α_{02}				1.8711			1.7607		
α_{03}				0.7728			0.6616		
α_{12}				0.8527			1.6178		
α_{13}				3.3064			3.3922		
α_{23}				4.1806			3.3268		

Table 1: Estimated fixed effects.

3 Results

After exploratory data analysis, we first considered the semiparametric main effects model (1). Table 1 shows estimation results for time–varying thresholds θ_{rt}, $r = 1, 2$, $t = 0, \ldots, 3$ and therapy effects β_t, $t = 1, 2, 3$ obtained from the PGEE (6) under a working independence assumption and under the unspecified working association assumption $\log \psi_{ist} = \alpha_{st}$. Estimates and standard errors for thresholds are in quite close agreement under both association models and show the expected results: Thresholds and, as a consequence, corresponding cumulative probabilities for the status of the healing process increase with time. Estimates for time-varying effects β_1, β_2 and β_3 provide clear evidence of an acceleration of the healing process for the therapy group. However, results for both models differ more distinctly from each other: Point estimates of these effects have
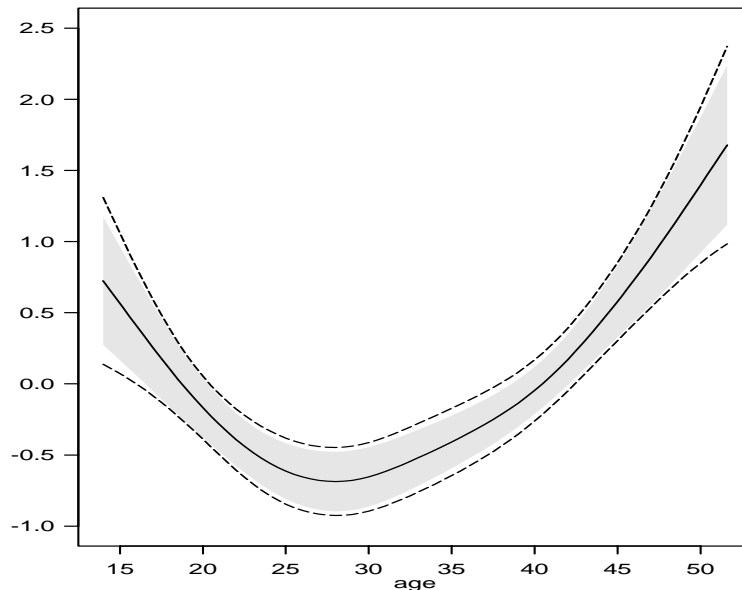
Figure 3: Main effects model: Estimated effect of age, naive standard error – dashed lines, robust standard error – boundary of shaded region.

smaller values under the unspecified association model. Also, standard errors are smaller due to gain of efficiency. As a consequence, the effect β_1, which is clearly nonsignificant under the independence assumption, becomes on the border to significance for the unspecified association model. A look at the estimates of the association parameters shows that it also seems to be important to allow for time–dependence of association parameters. A simple exchangeable association model would lead to loss of efficiency. We also experimented with association models stratified by covariates as in (5), but this did not result in any improvement.

The estimated effect $f(A)$ of age is plotted in Figure 3. The curve indicates that younger and older persons react less sensitive to pain from pressure than others. At first sight, this seems to be somewhat surprising. Therefore, in a second step, let us take a closer look at the influence of age by the varying coefficient model (2). In this model, possible interactions of age with the therapy can be explored. Table 1 and Figure 4 show that estimated thresholds and the main effect of age are still in good agreement with estimates obtained for the main effects model. Figure 5 compares the constant effects β_1, β_2 and β_3 of the therapy at $t = 1, 2, 3$ to corresponding effects $f_1(A)$, $f_2(A)$ and $f_3(A)$ varying over age. We see that refined analysis under the extended model provides additional information: The effects of therapy for younger persons, up to about 25 years, exhibit more variation
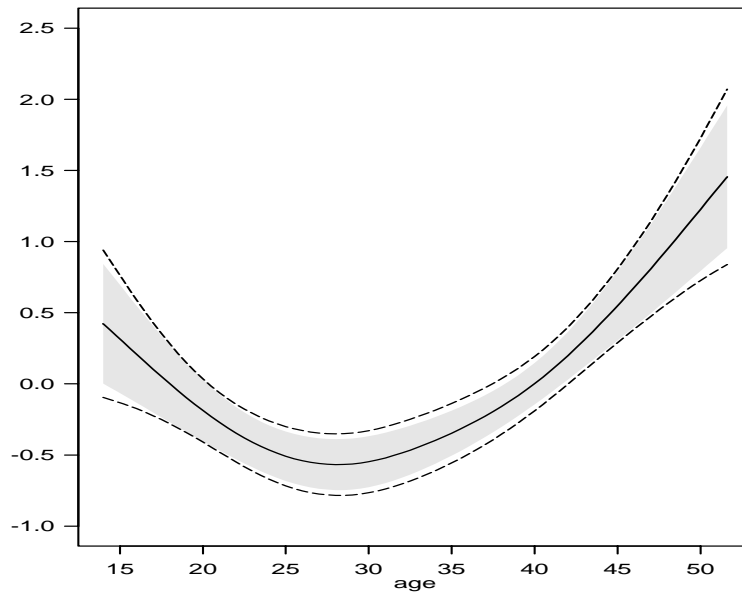
Figure 4: Varying coefficient model: Estimated main effect of age, naive standard error – dashed lines, robust standard error – boundary of shaded region.

over time than for older persons. For the young ones, the effect is still positive but still smaller at $t = 1$. However, it also increases more distinctly with time and is higher at $t = 3$. On the other side, for older patients, the effects at $t = 1$ and $t = 2$ are more or less of the same magnitude and lie between β_1 and β_2, while the effect at $t = 3$ is at about the same level as β_3. It seems that younger persons react more sensible to the therapy than others.

These results provide evidence for the fact that the three age groups young, middle and old show different reactions on pressure from pain in general and under the therapy. It is not clear if this is mainly caused by different subjective sensation of pain or if there is some physiological explanation.
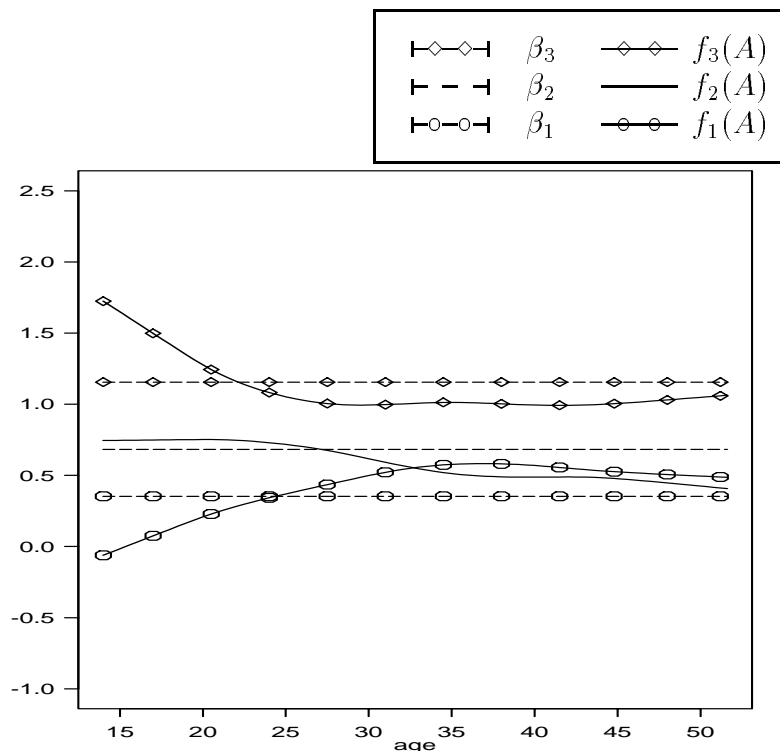
Figure 5: Effect of therapy: constant – dashed lines, varying over age – solid lines.

4   Conclusions

Inclusion of structural restrictions in bivariate or higher–order associations is an important aspect for adequate modelling in marginal regression. We discussed this for the problem at hand, but extensions to other settings are conceptually immediate. In particular, more general categorical responses, GEE2 and full like-lihood approaches (see e.g. Molenberghs and Lesaffre, 1995, Heumann, 1996,1997) are interesting topics.

Acknowledgements

References

CARROLL, R.J., RUPPERT, D., WELSH, A.H. (1997). Nonparametric Es-
timation Via Local Estimating Equations, with Applications to Nutrition

Calibration. *Discussion paper 17*, SFB 373, Humboldt–Universität, Berlin.

DEMING, W.E., STEPHAN, F.F. (1940). On a Least Squares Adjustment of Sampled Frequency Table when the Expected Marginal Totals are Known, *Annals of Mathematical Statistics*, 11, 427-444.

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.

GIEGER, C. (1997). Non- and Semiparametric Marginal Regression Models for Ordinal Response. *Discussion paper 71*, SFB 386, Ludwig–Maximilians Universität, München.

HASTIE, T., TIBSHIRANI, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

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.

HEAGERTY, P., ZEGER, S. (1997). Lorelogram: A Regression Approach to Exploring Dependence in Longitudinal Categorical Responses. Preprint.

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

HEUMANN, C. (1997). Likelihoodbasierte marginale Regressionsmodelle für korrelierte kategoriale Daten. Dissertation, Institut für Statistik, Universität München.

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.

MOLENBERGHS, G., LESAFFRE, E. (1995). Marginal Modeling of Multivariate Categorical Data. Preprint.

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

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