Journal of Statistical Research *ISSN 0256 - 422 X*
2016, Vol. 48-50, No. 1, pp. 15-33

Analysis of ordinal longitudinal data using semi-parametric mixed models

KALYAN DAS
Department of Statistics, University of Calcutta, India
Email: kalyanstat@gmail.com

SURUPA ROY
Department of Statistics, St. Xavier's College, Kolkata, India
Email: surupachakraborty@yahoo.co.in

ASIS KUMAR CHATTOPADHYAY
Department of Statistics, University of Calcutta, India
Email: akcstat@gmail.com

SUMMARY

A spline mixed item response theory model that allows for three-level multivariate ordinal outcomes and accommodates multiple random subject effects is proposed for analysis of ordinal outcomes in longitudinal studies. Assuming cumulative logit model with proportional odds, maximum marginal likelihood estimation for model parameters is proposed utilizing Monte Carlo Metropolis Hastings Newton Raphson (MCMHNR) algorithm. An iterative Fisher scoring solution, which provides standard errors for all model parameters, is considered. The performance of the estimates of the model parameters in finite samples has been looked into. A longitudinal orthodontic data set, where plaque content in teeth is repeatedly measured over time, is used to illustrate application of the proposed model.

Keywords and phrases: ordinal response, proportional odds model, spline, Monte Carlo EM, Metropolis-Hastings, orthodontic data.

1 Introduction

Many interesting problems in Biomedical, industrial and other experiments involve the study of how an ordered response variable depends on a set of regressors. In psychometric and educational testing literature, a large amount of research has been devoted to developing mixed-effects models for subject-specific comparisons of multivariate ordinal responses. In longitudinal studies, information from the same set of subjects is measured repeatedly over time. Multivariate data arise when different item responses, related to a single underlying outcome, are measured to provide more complete and reliable information. The aim of such studies is to estimate the mean or individual response at a

certain time, to relate time-invariant or time-dependent covariates to repeatedly measured response variables, or to relate the response variables to each other.

One way to model ordinal regression data is to assume that the observed response is the discrete version of a continuous latent variable for which a linear regression model holds. Alternatively, an index model of the discrete probabilities may be written for a given transformation, called link function as in the seminal paper of McCullagh (1980). It is well known that the latent variable approach and the index model approach are essentially equivalent (see Greene, 2004 and Wooldridge, 2003). Examples of such related models are obtained by assuming the logistic distribution for the errors in the latent variable and the ordered logit model, or the normal distribution for the latent error and the ordered probit model.

The restricted version of the generalized logit model is the standard ordered logit model discussed in most statistics textbooks and it is known in the statistical literature as the proportional odds model (see McCullagh, 1980). Especially when the number of possible ordinal values is large, the model may require many more parameters than the simple ordered logit model. This may be justified for example when it is reasonable to assume that the threshold between adjacent categories depends on subjective judgments, as for instance in the analysis of the determinants of health status, happiness etc. As the ordered logit model may be seen as a properly constrained generalized logit model, the effect of covariates on threshold parameters may be tested by imposing appropriate linear constraints. When the dependence of threshold parameters on individual covariates is not justified by the nature of the response variable, the rejection of the proportional odds assumption should be taken as a warning that the latent model is not properly specified, like when, for instance, the distribution of the error is heteroscedastic or the covariate is not exogenous.

Often in longitudinal studies it is required to characterize the temporal trends exhibited by some real data. The mean trajectory appears to show curvature. In fact individual series shows more curvature. In a situation where the primary focus of the analysis is to relate disease progression at different time points to the subject's habit/nature, it is of practical interest to develop an appropriate method that truly incorporates the temporal patterns as well as the covariate information. Certainly, a less restrictive assumption on the time functions might be more desired than imposing some parametric assumptions, which might be incorrect.There has been a tremendous advancement in statistical research on non parametric function estimation. In many situations a semi parametric generalized partially linear mixed model (GPLMM) is considered for handling the covariate effects (time) non-parametrically. Such a model is essentially a compromise between the GLMM and a fully nonparametric model. This kind of model is popular in longitudinal studies such as human viral dynamics, pharmacokinetic analyses and studies of growth and decay. On the other hand the inclusion of a nonparametric covariate in an otherwise GLMM raises the high dimension problem. In order to avoid this, we consider a generalized partial ordinal longitudinal model (GPOLM) that can be viewed as a compromise between GLMM and a fully nonparametric model.

Considerable studies have been done on partially linear models (see Hardle et al., 2000). In order to analyze discrete outcomes, where the influential covariates and the outcome have definite functional relationship (monotone), it is natural to extend the model to a partial semi parametric Generalized linear model. Previously, Severini and Staniswalis (1994), Hardle et al. (1998) and Muller

(2001) have looked into the influential aspects of GPLM. Later Lin and Carroll (2001a), Wang et al. (2005) and He et al. (2005) have considered the GPLMM in the context of clustered/longitudinal data. Lin and Carroll (2001b) address that the conventional profile kernel-based approach is incapable of producing a $\sqrt{n}$ consistent estimator of the parameters unless the non-parametric function is under-smoothed or working independence is assumed for the GEE methodology. These limitations can be avoided if regression spline approximation is considered in GPLMM.To the best of our knowledge, no literature has yet been published for the analysis of GPOLM. Our attempt is to show that in the regression spline approximation under GPOLM, the spline approach results in the optimal rate of convergence for estimating the unknown function and the parameters of interest. The primary focus of our paper is to use a spline mixed regression model for analyzing ordinal longitudinal data. Such a model accommodates longitudinal dependence and subject specific variation in the data through random effects. We consider a data on oral hygiene where 220 individuals consisting of students and staff members of medical schools in and around the city of Kolkata were selected randomly irrespective of age, sex and oral hygiene status and their plaque scoring was recorded according to Turesky et al. (1970). The reduction in the thickness of plaque for subjects are usually recorded as belonging to four different categories, viz 'no reduction', 'slight reduction', 'moderate reduction' and 'vast reduction' (to a great extent). In addition, auxiliary information on age, sex, food habit, smoking habits etc were also observed for each subject. The purpose of the study is to see whether the progression of the plaque reduction is truly effective with the use of a solution (kept in mouth for 1 minute followed by a thorough rinse with water to remove any excess of disclosing solution) and if so, to what extent such progression depends on the covariates taken.

The article is organized as follows. In Section 2 we introduce the spline mixed cumulative logit model with proportional odds setup. In Section 3 we consider estimation of model parameters using MCMHNR approach. In section 4 an asymptotic study is given. An exact sample study has been carried out in Section 5, to see the performance of the estimator under the proposed approach. Data arising from an orthodontic study have been analyzed in Section 6. Finally, conclusion and discussion are made in Section 7.

2 The Model and Likelihood

Consider a trial involving n individuals in which each individual is to be examined at K assessment times. Let y_{ijk} denote the ordinal response that has $L+1$ distinct levels, $0, \ldots, L$ (say) for individual i within the cluster j at the assessment time $(k, i = 1, \ldots, n;\ j = 1, \ldots, r; k = 1, \ldots, K)$. This gives rise to a hierarchical data structure where the assessment times (level 1) are nested within the clusters (level 2) which in turn are nested within the individuals (level 3). Further suppose, associated with the ordinal response, x_{ijk} denote the covariate vector for individual i in cluster j at time k. The covariates are completely known and may be assumed to be fixed across the entire observation times. Let u_{ij} denote the subject and cluster specific random component vector corresponding to individual i in the cluster j. The random component reflects the unobserved heterogeneity in the data. Dummy variables are often used to represent categorical variables in estimation of parameters. Let us denote Y_{ijk} as a vector of $L+1$ indicator variables, given by, $Y_{ijk} = (Y^0_{ijk}, \ldots, Y^L_{ijk})'$ with $Y^l_{ijk} = 1$, if

$y_{ijk} = 1$ and 0, otherwise ($l = 0, \ldots, L$). Further suppose, the vector of probabilities and cumulative probabilities are respectively denoted by $\pi_{ijk} = (\pi^0_{ijk}, \ldots, \pi^L_{ijk})'$ and $\eta_{ijk} = (\eta^0_{ijk}, \ldots, \eta^L_{ijk})'$, where π^l_{ijk} and η^l_{ijk} are given by,

$$\pi^1_{ijk} = P\left(Y^1_{ijk} = 1 \mid x_{ijk}, u_{ij}\right) = P\left(y_{ijk} = l \mid x_{ijk}, u_{ij}\right), \tag{2.1}$$

$$\eta^1_{ijk} = P\left(y_{ijk} \leq l \mid x_{ijk}, u_{ij}\right) = \sum_{i0}^{1} \pi^1_{ijk}. \tag{2.2}$$

Corresponding to the individual i, the multivariate ordinal data can be represented as $(y_{i11} = c_{11}, \ldots, y_{ijk} = c_{jk}, \ldots, y_{irK} = c_{rK})'$, where c_{jk} $(j = 1, \ldots, r; k = 1, \ldots, K)$ can take the ordinal scores $0, \ldots, L$. Conditional on the subject and cluster specific random components u_{ij} and given the covariates, the associated probability follows from (2.1) and can be written as,

$$\begin{aligned} P_{ij} &= \prod_{k=1}^{K} \prod_{l=0}^{L} \{P(y_{ijk} \leq l \mid u_{ij}, x_{ijk}) - P(y_{ijk} \leq l-1 \mid u_{ij}, x_{ijk})\}^{l(y_{ijk}=l)} \\ &= \prod_{k=1}^{K} \prod_{l=0}^{L} (\eta'_{ijk} - \eta^{l-1}_{ijk})^{l(y_{ijk}=l)}, \end{aligned} \tag{2.3}$$

where $I(y_{ijk} = l) = 1$, if $y_{ijk} = l$ and 0 otherwise, $\eta^{-1}_{ijk} = 0$ and $\eta^L_{ijk} = 1$. To model the dependence of the response on the covariates and the random component we use cumulative logit model with proportional odds assumptions. Typically such a model is written as,

$$\text{log it}(\eta'_{ijk}) = \log\left(\frac{\eta'_{ijk}}{1 - \eta'_{ijk}}\right) = \lambda_l + x'_{ijk}\beta + z'_{ijk}u_{ij} + f_0(t_{ijk}), \tag{2.4}$$

where λ_l $(l = 0, \ldots, L-1)$ is the intercept in the lth logit model which satisfy the relationship $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{L-1}$ and β denotes the p dimensional vector of covariate effects corresponding to x_{ijk}. The random component vector u_{ij} is a subject and cluster specific random effect of dimension q associated with the completely specified design vector z_{ijk}. For subject i in cluster j, we write the random component vector u_{ij} as, $u_{ij} = (u^1_{ij}, \ldots, u^q_{ij})'$ and assume that $u_{ij} \sim N_q(0, I_q)$. Let us further write $u_i = (u'_{i1}, \ldots, u'_{ir})'$, where

$$u_i \sim N_{rq}(0, I_q \otimes \Sigma), \quad \Sigma = \sigma^2[(\mathbf{1} - \rho)I_r + \rho\mathbf{1}\mathbf{1}']. \tag{2.5}$$

In model (2.5), σ^2 and ρ denotes the intra cluster variability and correlation coefficient respectively. They are treated as nuisance parameters and are estimated along with the other regression parameters. In model (4), t_{ijk} may be simply time or in general any time dependent covariate and $f_0(.)$ is an unknown smooth function.

We use the basis of cubic B-splines with q preselected knots to approximate the unspecified smooth function f_0 in which the rth knot corresponds to the $r/(q+1)$th sample quantile of the distinct values of $t_{ijk} (i = 1, \ldots, n; j = 1, \ldots, r; k = 1, \ldots, K)$. Let $B_1(t), \ldots, B_{q+4}(t)$ be the cubic B-spline basis for the space of cubic splines with q preselected knots. For details on computing

of B-splines and their mathematical properties we refer to Boor (2001). The cubic B-splines space includes a constant function, and the constant is given in the parametric component of the model (4), so to model $f_0()$ one of the $(q+4)$ B-spline basis functions needs to be dropped so that the resulting parameterization is of full rank. Any one of them can be dropped, but for convenience Li (2011) models f_0 as a linear combination of the first $q+3$ fixed-knot cubic B-spline basis functions. In this paper in order to approximate f_0 by a regression spline, we consider a set of knots on $[0,1]$ with $0 = s_0 < s_1 < \cdots < s_{k_n} = 1$ and generate $N = k_n + l$ normalized B-spline basis functions of degree $l+1$ that span the linear space. We then express $f_0(t) \approx v'(t)\gamma$, where, $v(t) = (B_1(t), \ldots, B_N(t))'$ is the vector of basis functions and $\gamma \in R^N$ is the spline coefficient vector. Let us denote the vector of parameters by $(\theta', \phi')'$ where, $\theta = (\lambda_0, \ldots, \lambda_{L1}, \beta', \gamma')'$ and $\phi' = (\sigma^2, \rho)'$. Then in view of (3) and (5), the likelihood for subject i can be written as,

$$L_i(\theta, \phi) = \int \prod_{j=1}^{r} \prod_{k=1}^{K} \prod_{l=1}^{L} \left[\eta'_{ijk} - \eta^{l-1}_{ijk}\right]^{I(y_{ijk}=l)} g(u_i) du_i, \tag{2.6}$$

where $g(u_i)$ denotes the density function of u_i given in (2.5). Here our primary focus lies in estimating and making inference on the parameter vector θ although the vector of nuisance parameter ϕ is also estimated in the study simultaneously.

The critical issue for getting a rigorous model selection criterion can be based on estimating the relative expected Kullback-Leibler $(K - L)$ information. Akaike (1973) found that the maximized log likelihood value was a biased estimate of $K - L$ information but this bias was approximately equal to 'p', the number of estimable parameters in the approximating model. Thus an approximately unbiased estimator of $K - L$ information for large samples and good models is given by Akaike's Information Criterion (AIC), where

$$\text{AIC} = 2 \log L(\hat{\theta}, \hat{\phi}) + 2p. \tag{2.7}$$

In (2.7) above, $(\hat{\theta}, \hat{\phi})$ is the maximum likelihood estimator of the parameter vector arising in model (2.6) and $L(\cdot)$ denotes the likelihood function given the data vector. Minimizing the AIC over a set of possible models can thus be seen as minimizing the average distance of an approximating model to the underlying truth.

3 Parameter Estimation

The likelihood function given in (2.6) is difficult to maximize because of the multidimensional integral over u_i which is the consequence of a mixed effects modelling. Numerical integration techniques like Gauss Hermite quadrature or adaptive Gaussian quadrature (Pinheiro and Bates, 1995) can be used to approximate the above integral to any practical degree of accuracy. Diverse methodologies in both Bayesian and Classical paradigm are available in the literature for fitting GLMM. In Bayesian perspective Markov Chain Monte Carlo (MCMC) method is implemented via Gibbs sampling techniques (Zeger and Karim, 1991) to generate repeated samples from the posterior distribution of the random effects. In the classical approach Breslow and Clayton (1993)

proposed the penalized quasi likelihood (PQL) for approximating the high dimensional integration using Laplace approximation. However, as reported by several authors PQL estimates are biased downwards for some variance components. Later Breslow and Lin (1995) and Lin and Breslow (1996) gave bias corrected PQL. McCulloch (1994) investigated GLMM with a probit link using Monte Carlo EM (MCEM). He extended MCEM to the logit model and introduced the Monte Carlo Newton Raphson (MCNR) and simulated maximum likelihood methods. For simple models it was found that the MCNR estimates inherits the properties of the exact ML estimates. Natarajan et al. (2000) and Zhou and Liu (2008) used the Monte Carlo version of EM to calculate ML estimates of parameters. Meza et al. (2009) and Davier and Sinharay (2010) proposed an alternative to MCEM via the Stochastic Approximation EM (SAEM) of Deylon et al. (1999). We could have considered any one of the three stochastic versions (SEM, SAEM and MCEM) to analyze our data. Since all three lead to similar conclusions (Celeux et al., 1995), we preferred to work with MCEM method here.

In this paper we adopt the MCNR approach to calculate the fully parametric Maximum likelihood estimates based on the likelihood (6). The Monte Carlo approach calls for generating random observations from the posterior distribution of the random effects which however is not in a closed form. To circumvent this difficulty Metropolis Hastings algorithm (see Chib and Greenberg, 1995) is used to generate data from the posterior distribution of the random effects which does not require the exact form of the conditional distribution. Moreover a good starting solution is needed for the MCNR method. In our analysis moment estimates are used. McCulloch (1997) pointed out that although this approach is computationally intensive it provides feasible solutions for a variety of data configurations. In presence of influential points in the data this method can be extended to the Robust Monte Carlo Newton Raphson method of Sinha (2004).

3.1 The MCMHNR Approach

To set up the EM algorithm, we consider the random effects to be missing. We write the observed data for individual $i \;\; (i = 1, \ldots, n)$, as $D_{0i} = \{y_{ijk}, x_{ijk}, t_{ijk}; j = 1, \ldots, r;\, k = 1, \ldots, K\}$ and the complete data is denoted by $D_{ci} = \{y_{ijk},\; x_{ijk},\; t_{ijk};\, j = 1, \ldots, r, k = 1, \ldots, K\}$. Further suppose $f(D_{oi} \mid u_i)$ denotes the conditional distribution of the observed data given the random component. Then using (3) and (5) the complete data log likelihood for all the subjects is given by,

$$\begin{aligned} l_c(\theta, \phi) &= \sum_{i=1}^{n} \log(f(D_{0i} \mid u_i; \theta)) + \sum_{i=1}^{n} \log(g(u_i; \phi)) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{r} \sum_{k=1}^{K} \sum_{l=1}^{L} I(y_{ijk} = l) \log\left[\eta_{ijk}^{l} - \eta_{ijk}^{l1}\right] + \sum_{i=1}^{n} \log(g(u_i; \phi)) \\ &= l_{c1}(\theta) + l_{c2}(\phi). \end{aligned} \tag{3.1}$$

From (3.1) it is to be noted that since θ enters only the first term so the M step of EM algorithm with respect to θ uses only $L_{c1}(\theta)$. The second term in (8) involves only the distribution of u_i which is assumed to be normal and so maximizing the likelihood $l_{c2}(\phi)$ gives the standard maximum

likelihood estimates of ϕ after replacing u_i's with their conditional expected values. Writing $D_0 = \{D_{0i}, i = 1, \ldots, n\}$ and $u = (u_1', \ldots, u_n')'$, the score functions for θ and ϕ can be expressed as:

$$\xi_\theta(\theta) = E_u\left[\frac{\partial l_{c1}(\theta)}{\partial \theta}\Big| D_0\right] = 0; \quad \xi_\phi(\phi) = E_u\left[\frac{\partial l_{c2}(\phi)}{\partial \phi}\Big| D_0\right] = 0. \tag{3.2}$$

In order to solve for θ and ϕ from equation (3.2), we propose a Monte Carlo Newton Raphson (MCNR) algorithm. Using MCNR, the updated estimate of θ and ϕ at $(t+1)$th step is given by,

$$\theta^{(t+1)} = \theta^{(t)} - \Lambda_1^{-1(t)}\xi_\theta(\theta^{(t)}), \quad \phi^{(t+1)} = \phi^{(t)} - \Lambda_2^{-1(t)}\xi_\phi(\phi^{(t)}) \tag{3.3}$$

where $\Lambda_1^{(t)} = \partial\xi_\theta(\theta)/\partial\theta|_{\theta^{(t)}}$ and $\Lambda_2^{(t)} = \partial\xi_\theta(\phi)/\partial\phi|_{\phi^{(t)}}$. The expressions for first and second order derivatives are given in Appendix A1. The MCNR approach gives an iterative computational scheme, where the maximization step becomes automatic. However the conditional expectations in (3.2) cannot be computed in a closed form. This is because the conditional distribution of u involves the marginal distribution of the data which in fact is the likelihood in equation (2.6) that we are trying to avoid calculating directly. To circumvent this difficulty we use Metropolis Hastings algorithm (Smith and Roberts, 1993) to produce random draws from the conditional distribution of $u \mid D_0$. Then we can approximate the required expectation in (3.2) by Monte Carlo approach.

To implement the Metropolis algorithm, we first specify the candidate distribution $h(u)$ from which potential new values are drawn and then compute the acceptance function that gives the probability of accepting the new value (as opposed to keeping the previous value). In our case, the target density can be expressed as proportional to the product of the density $g(u;\phi)$ that can be sampled and the conditional density $f(D_0 \mid u, \theta)$ that is uniformly bounded. Thus following Chib and Greenberg (1995) we set the proposal density to be equal to $g(.)$ (as in the independence chain) to draw candidates. In this case the acceptance probability takes a simplified form and requires the computation of $f(D_0 \mid u, \theta)$ only. Let u^0 denote the previous draw and u^{can} is a new value from the candidate distribution. Then we accept u^{can} as a potential observation from the conditional distribution of with probability of acceptance given by,

$$A(u^0, u^{com}) = \min\left\{\frac{f(D_0 \mid u^{com}, \theta)}{f(D_0 \mid u^0, \theta)}, 1\right\}. \tag{3.4}$$

Incorporating the Metropolis step in MCNR method results in MCMHNR algorithm which can now be stated as follows:

Step 1: Choose starting values θ^0, ϕ^0. Set $t = 0$.

Step 2: Generate R values $u^{(1)}, u^{(2)}, \ldots, u^{(R)}$ from the conditional distribution $f(u \mid D_0, \theta, \phi)$ using the Metropolis Hastings algorithm and use them to form the Monte Carlo estimates of the expectations.

Step 3: Compute:

$$\theta^{(t+1)} = \theta^{(t)} - \hat{\Lambda}_1^{-1(t)}\hat{\xi}_\theta(\theta^{(t)})$$
$$\phi^{(t+1)} = \phi^{(t)} - \hat{\Lambda}_1^{-1(t)}\hat{\xi}_\phi(\phi^{(t)}).$$

Replacing the expectations in (3.2) by Monte Carlo estimates and using (3.1),it follows that,

$$\hat{\xi}_\theta(\theta) = \frac{1}{R}\sum_{r=1}^{R} \frac{\partial}{\partial\theta} \log f(D_0 \mid u^{(r)}; \theta); \hat{\xi}_\phi(\phi) = \frac{1}{R}\sum_{r=1}^{R} \frac{\partial}{\partial\phi} \log g(u^{(r)}; \phi)$$

$$\hat{\Lambda}_1 = \frac{\partial}{\partial\theta}\hat{\xi}_\theta(\theta); \hat{\Lambda}_2 = \frac{\partial}{\partial\phi}\hat{\xi}_\phi(\phi)$$

Set $t = t + 1$.

Step 4: If convergence is achieved, declare $\theta^{(i+1)}$ and $\phi^{(i+1)}$ as the maximum likelihood estimates of θ and ϕ respectively. Otherwise return to Step 2.

3.2 Knot Selection

An important aspect of spline smoothing is knot selection. Since we are mainly concerned with the efficiency of the covariate effect estimates, we opt for convenient choices of knot placements. For the Knot selection we have applied a data adaptive scheme which is briefed below:

Step 1: We at first consider $Q_1 = 10$ largest local maxima and $Q_2 = 10$ smallest local minima.

Step 2: We have identified the time points corresponding to these $Q = Q1 + Q2$ points. These Q points have been chosen as the initial knots. Let $q = Q + k + 1$, for cubic spline $k = 3$. These k points are determined based on the quantiles.

Step 3: We removed the ith knot and evaluated the residual sum of squares (RSS_i), for $i = 1, 2, \ldots$

Step 4: We have chosen that model for which RSS_i is minimum and set $q = q - 1$.

Step 5: We have continued Steps 2-4 till $q = k + 1$.

4 Asymptotics

In this section, to ensure consistency of the proposed estimates, the asymptotic properties of the solution to score equations in (3.2) have been investigated. The asymptotic distribution of the estimators of θ and ϕ would be separately looked into as in view of (3.1), l_{c_1} involves only θ and l_{c_2} involves only ϕ. Essentially, here this section, we would consider only the asymptotic distribution of $\hat{\theta}$ as that of $\hat{\phi}$ is straightforward. We consider a sequence of consistent estimators $\hat{\theta}_n (= \hat{\theta}$ say) in the sense that as

$$n \to \infty, \sup_{t\in[0,r]} \mid v'(t)\hat{\gamma} - f_0(t) \mid \xrightarrow{P} 0, \hat{\lambda} - \lambda^0 \xrightarrow{P} 0 \text{ and } \hat{\beta} - \beta^0 \xrightarrow{P} 0,$$

where λ^0 and β^0 are true unknown values of λ and β respectively. The required basic assumptions are given below.

A.1 The distinct values of $t_{ijk}, 0 \leq t_{ijk} \leq \tau$ form a quasi-uniform sequence that grows dense on $[0, 1]$.

A.2 For every $i, Max\{\|X_i\|\} \leq B_0$ for some non-random constant B_0, where $X_i = ((x_{ijk}))$ $i = 1, \ldots, n; j = 1, .., r; k = 1, \ldots, K$.

A.3 $\mid f_0^{(s)}(.) \mid < A_0$, for some non-random value A_0 for $s \geq 2$.

A.4 Conditional on data and for every $i, \sup_{i\geq 1} E\|S_{ic}\|^{2+\delta} < \infty$, for some $\delta > 0$, where $S_{ic} = \frac{\partial}{\partial\theta} l_{ic}(\theta)$ and $l_{ic}(\theta) = \sum_{i=1}^{r} \sum_{k=1}^{K} \sum_{l=1}^{L} I(y_{ijk} = l) \log[\eta'_{ijk} - \eta^{l-1}_{ijk}]$. In fact, $E_D E_{u|D} \left(\frac{\partial^2}{\partial\theta\partial\theta'} l_{ic}(\theta) \right) = B_i$, with $\sup_{i\geq 1} \|B_i\| < \infty$ and D stands for the whole data set.

A.5 True parameter vector $\theta^0 = (\lambda^{0l}, \beta^{0l}, \gamma^{0l})'$ satisfies $\|\theta^0\| \leq M_0$ for some known constant $M_0 (> 0)$.

Assumption A.1 essentially indicates that we have only local dependence in the sample. Assumption A.2 is the compact support for covariates. The smoothness condition on f0 given in assumption A.3 determines the rate of convergence of the spline estimate $\hat{f} = v'(t)\hat{\gamma}$. Both the assumptions A.2 and A.4 are natural and are easy to check. Assumption A.5 is basically a technical condition required to justify consistency.

It is true that, in our model, the covariates x_{ijk} may be time dependent and hence must depend on t_{ijk}. Such dependence can be taken into account through some relationship (either linear or non-linear). For example, we can express covariates as,

$$X_{ijku} = \Psi_u(t_{ijk}) + \varepsilon_{ijku}; i = 1, \ldots, n; j = 1, \ldots, r; k = 1, \ldots, K; u = 1, \ldots, p. \tag{4.1}$$

where $\Phi_u(\cdot)$ are p functions for each of which sth derivative is bounded and ε_{ijku}'s are independent random variables with mean zero and also independent of y_{ijk}'s. In view of the fact that γ is the nuisance parameter vector, for clear representation we modify equation (3.3) as,

$$\hat{\theta} = \hat{\theta}^0 - \left[\Lambda_1^{*-1} \xi_\theta^*(\theta)\right]_{\theta=\hat{\theta}_0} \tag{4.2}$$

where

$$\begin{aligned}
&A_1^* = E\left[\frac{\partial}{\partial\theta'}(X^{*\prime} W Y_0) \mid D\right], \xi_\theta^* = E\left[X^{*\prime} W Y_0 \mid D\right], X^* = (I - H)X, \\
&H = P(P'P)^{-1}P', P = 1_L \otimes v'(t_{ijk}), \\
&Y_0 = (y_{ijkl}, i = 1, \ldots, n; j = 1, \ldots, r; k = 1, \ldots, K; l = 0, \ldots, L-1)', \\
&R_n = (X^{*\prime} X^*), W = \text{ Diag } (\ldots, 1 - \eta'_{ijk} - \eta^{l-1}_{ijk}, \ldots) \text{ and} \\
&X = \begin{pmatrix} 1_L \otimes 1'_L & x'_{111} & v'_{111} \\ 1_L \otimes 1'_L & x'_{112} & v'_{112} \\ \vdots & \vdots & \vdots \\ 1_L \otimes 1'_L & x'_{nrk} & v'_{nrk} \end{pmatrix}
\end{aligned}$$

For the existence of Fisher Information, the following assumptions are further made,

$$\text{A.6 } (i) \lim_{n\to\infty} \frac{k_n}{n}(P'P) = Q, \quad (ii) \lim_{n\to\infty} R_n = R.$$

In assumption, A.6 (i), k_n is the number of knots, Q and Rare positive definite matrices with all eigen values bounded. Assumption A.6 (i) is a very standard property of B-spline basis functions and holds true under general design conditions (He and Shi, 1996). A.6 (ii) is a prerequisite for the existence of asymptotic distribution of the proposed estimator. The asymptotic distribution of $\hat{\beta}_n$ then follows from the following theorem:

Theorem 1. *Under assumptions A.1-A.6, the MLE $\hat{\theta}$ of θ^0 is consistent i.e. $\|\hat{\theta} - \theta^0\| \xrightarrow{P} 0$ as $n \to \infty$. Specifically as $n \to \infty$,*

$$\left(\hat{\beta} - \beta^0\right) \xrightarrow{P} 0, \qquad \sup_{t\in[0,r]} \mid v'(t)\hat{\gamma} - f_0(t) \mid \to 0. \tag{4.3}$$

The sketch of the proof is given in Appendix A2.

5 Simulation Study

In the simulation study we focus on a setting where $L = 4, K = 4, r = 4$ and $n = 100$. We simulate the clustered longitudinal ordinal response from a model with,

$$\text{logit}(\eta'_{ijk}) = \lambda_l + \beta x_i + u_{ij} + \sin(\pi t_{ijk}), \tag{5.1}$$

where the monotone difference intercepts $(\lambda_0, \lambda_1, \lambda_2)$ are assigned the value $(-2.0, -1.5, -1.0)$ and the regression parameter β is chosen to be 0.5. The time dependent covariate t_{ijk} is simulated from Uniform $(-1, 1)$ while the baseline covariate x_i is generated from $N(0, 1)$. The random component $u_i = (u_{i1}, \ldots, u_{ir})'$ is generated from a r-variate normal distribution with mean zero and variance-covariance matrix given by $\sigma_u^2 \left[(\mathbf{1} - \rho)I_r + \rho\mathbf{1}\mathbf{1}'\right]$, where the true values of σ_u^2 and p are taken to be 1.0 and 0.6 respectively. During the estimation process the function $\sin(\pi t_{ijk})$ is approximated by the normalized cubic B spline basis function. The data adaptive scheme outlined in Section 3.2 is applied and the number of internal knots is chosen to be 4. The knot points are taken as the 20th, 40th, 60th and 80th percentile values of $t_{ijk}; i = 1, \ldots, n; j = 1, \ldots, r; k = 1, \ldots, K$. Metropolis Hastings (MH) algorithm is employed for generating observations from the conditional distribution of u_i given the data. For simplicity and time saving purpose, the MH sample size R is chosen to be 500. The number of iterations needed in the Newton Raphson method within the Metropolis algorithm is predetermined to be 30. This resulted in about two-decimal accuracy in the simulation study. The simulation is repeated 100 times. For each parameter θ_i associated with the outcome model the goodness of fit measures namely bias and mean square error (MSE) are computed. Suppose $\hat{\theta}_{u'}$ denote the estimate of θ_i in the $t'th$ simulated data. Then Bias and MSE are

given by,

$$\text{Bias}_i = \frac{1}{100}\sum_{u=1}^{100}(\hat{\theta}_u - \theta_i); MSE_i = \frac{1}{100}\sum_{u=1}^{100}(\hat{\theta}_u - \theta_i)^2. \tag{5.2}$$

The first measure assesses the accuracy of $\hat{\theta}_i$ and the second measure assesses the precision. We also compared the efficiency of the proposed model with the naïve model. For a naïve model the ordinal responses are generated using (5.1), but we fit a model after replacing the nonlinear function of time by t_{ijk} simply. The estimated values of the parameters along with the bias and MSE of the estimates of the parameters are presented in Table 1 for the naïve model as well as for the proposed model which accounts for the longitudinal effect through spline function. The program has been implemented in R 2.14.1.

Table 1: Parameter estimates, simulated biases and mean square error of the parameter estimates for the proposed model and naive model.

Parameters	True	Naive Model			Proposed Model		
	values	Estimates	Bias	MSE	Estimates	Bias	MSE
λ_0	-2.0	-1.8059	0.1933	0.3802	-2.0319	0.0319	0.0435
λ_1	-1.5	-1.419	0.1065	0.3334	-1.507	-0.0906	0.0227
λ_2	-1.0	-1.155	-0.1563	0.3866	-1.003	-0.0032	0.0161
β	0.5	0.5178	0.0178	0.0152	0.4968	-0.0131	0.0097
σ_u^2	1.0	0.9137	-0.0232	0.0128	0.9677	-0.0962	0.0020
ρ	0.6	0.6000	0.0000	.0000	0.6000	0.0000	0.0000

Table 1 shows that the monotone difference estimates and the regression coefficients are biased under the naïve model, whereas the proposed model recovers the estimates well. However the estimates of the parameters associated with the distribution of the random component remains robust under model misspecification. In the naïve model we have 6 parameters while the proposed model involves 13 parameters. The AIC factor under naïve model comes out to be 3642.622 while that under the proposed model is 3552.039. During the estimation process under the proposed model the spline coefficients $\gamma = (\gamma_1, \ldots, \gamma_N)$ are also estimated along with the other parameters of interest. The fitted function is then given by,

$$\hat{f}_0(t) = \sum_{m=1}^{N} \hat{\gamma}_m B_m(t), \tag{5.3}$$

where $\hat{\gamma}_m$ is the estimated value of γ_m and $B_m(t)$ denotes the B-spline basis function. The calculation of basis functions for the cubic B-spline is done using the $\{$splines$\}$ package in R. With four internal knot points and spline of order 3 and intercept =False the $bs(.)$ function in R returns $N = 7$

basis functions. Figure 1 displays the graph of the fitted function given by (5.3) and the true function given by $\sin(\pi t)$against the different values of t_{ijk}. The graph reveals that the cubic B-spline basis function approximates the true function $\sin(\pi t)$ well.

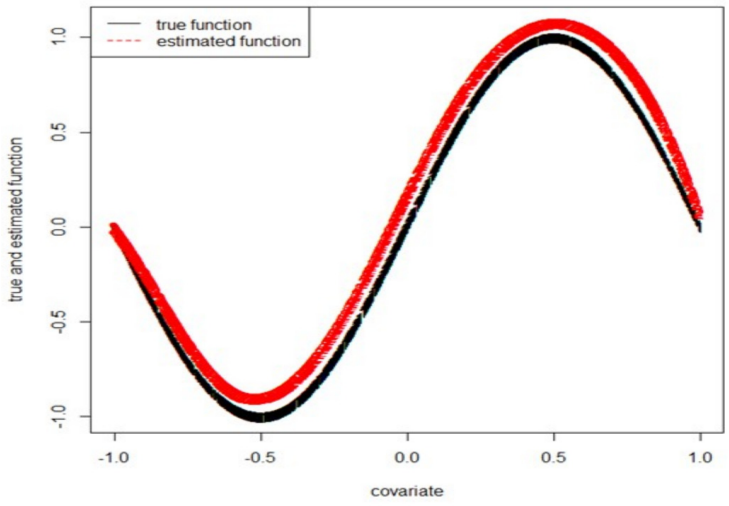

Figure 1: Plot of true function and estimated function against the time- dependent covariate.

For justification of the working of MCMHNR algorithm a simpler set up is chosen. Here we assume that in model $(16), \rho = 0$. This leads to uncorrelated random components and hence the multidimensional integration over $u_i = (u_{i1}, u_{i2}, u_{i3}, u_{i4})'$ is reduced to one dimensional integrals. The score equations now involve integrals over the random component u_i, which are evaluated using Gauss Hermite quadrature. Alternatively we apply MCMHNR algorithm as outlined in Section 3 under this simple set up. The likelihood estimates of the parameters are computed for each case. The AIC for the exact approach comes out as 2641.332, while that on application of EM algorithm is 2599.236. This shows that the MCMHNR method approximates the exact likelihood approach well.

6 Data Analysis

In this section we motivate the proposed model through an analysis of orthodontic data. Oral hygiene is of severe concern as a significant proportion of world population is highly susceptible to some destructive periodontal diseases. The data are the result of a study of 220 individuals consisting of staff members and students of medical schools in and around the city of Kolkata. These individuals have been selected at random irrespective of age, gender and oral hygiene status. A detailed history of each subject was recorded a week prior to the beginning of the study to collect information

like age, gender, occupation, food habits and smoking habits. Plaque scoring was done according to Tureskey et al. (1970). The teeth selected for scoring of plaque were the maxillary right first permanent molar, maxillary left permanent central incisor, maxillary left first premolar, mandibular left first permanent molar, mandibular right central incisor and mandibular right first premolar which we shall denote as teeth 1-6. Ordinal score of 0-2 was assigned as: 0 (No plaque), 1 (A thin band of plaque up to 1 mm at the cervical margin of the crown of the tooth.), 2 (A band of plaque wider than 1 mm of the crown of the tooth).

The categories 'moderate reduction' and 'vast reduction' were assigned the ordinal scores 1 and 2 respectively while the categories 'no reduction' and 'slight reduction' were combined and given the ordinal score 0.The plaque scoring on individual teeth was measured on four occasions separated at an interval of 1 month. Figure 2 shows the average response (plaque score) over time for each of the six teeth. The graph reveals a non-linear pattern in plaque deposit over time. The main focus of this orthodontic study is to see whether plaque reduction is truly effective with the use of a solution (kept in mouth for 1 minute followed by a thorough rinse with water to remove any excess of disclosing solution) and if so, to what extent such progression (i.e. plaque reduction) depends on the covariates taken. We consider the following model:

$$\eta^l_{ijk} = \lambda_l + \beta_A x_{Ai} + \beta_G x_{Gi} + \beta_F x_{Fi} + \beta_S x_{Si} + u_{ij} + f_0(t_K). \tag{6.1}$$

In equation (6.1) above, the baseline covariates $x_{Ai},\ x_{Gi},\ x_{Fi},\ x_{Si}\ \ (i = 1, \ldots, 220)$ correspond to age, gender, food habit and smoking habit respectively. The binary covariates x_{Gi}, x_{Fi} and x_{Si} takes the value 1 if the person is a male , non-vegetarian and a smoker. The non-linear behavior of the response over time is captured by the smooth unknown function $f_0(t_k)(k = 1, \ldots, 4)$, where $t_k = k$. In the analysis, the unknown function is approximated by a smoothing spline of order 1 with 4 internal knot points. Table 2 provides the estimated values of the parameters along with their standard errors for both the naive model and the proposed model. The naive model replaces the non-linear function by t_k. The results reveal that the smokers will have on an average less value of the response i.e. plaque reduction. Moreover food habit is not a significant factor in determining the effect of the solution (treatment) on plaque reduction.In this study 'age' does not play a significant role. The reason for this may be that the subjects considered belonged to almost the same age group. Finally it can be concluded from the results that the particular treatment applied on plaque reduction had better effect on males. The fitted function $\hat{f}_0(t) = \sum_{m=1}^{N} \hat{\gamma}_m B_m(t)$ is computed for $N = 3$. Here $\hat{\gamma}_m$ denotes the estimated value of the spline coefficients corresponding to the basis spline function $B_m(t)$ for different time points (t). The function shows a non-linear decreasing trend over time. Thus it can be inferred that in general the application of the solution helps in reducing plaque deposit over time.

7 Conclusion

In many longitudinal set up where responses are ordinal in nature, one faces the stiff challenge in expressing the dependence of such responses over time. In our present orthodontic study, it is evident from Figure 2 that average response (plaque reduction) varies nonlinearly over time. The variation

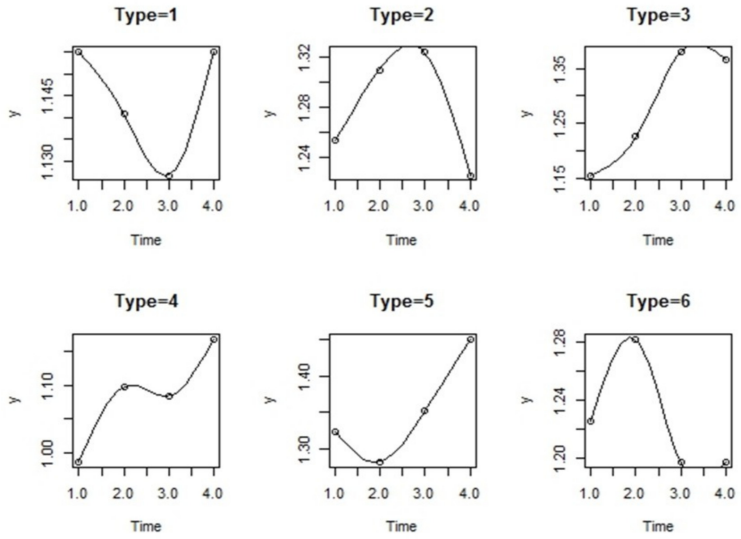

Figure 2: The average response (plaque score) over time for each of the six teeth.

also changes over the six teeth. To account for such unknown variability, we have proposed a GPOLM that can be viewed as a compromise between GLMM and a fully nonparametric model. We have approximated the non-parametric function in the GPOLM by a regression spline. A MCMHNR method has been proposed to estimate the model parameters. Simulation study indicates that the model which ignores the non-linear effect of time produces biased estimates of the intercepts and the regression coefficients. Result from the orthodontic study reveals that smoking has a negative effect in plaque reduction. However in general the application of the solution helps in reducing plaque deposit over time.

Acknowledgment

The authors are thankful to the reviewer for careful suggestions that helped to improve the clarity of the manuscript.

Appendix A1

First order derivatives:

$$\frac{\partial l_{c_1}(\theta)}{\partial \theta} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} \frac{I\left(y_{ijk}=l\right)}{\left(\eta_{ijk}^{l}-\eta_{ijk}^{l-1}\right)}\left(\frac{\partial \eta_{ijk}^{l}}{\partial \theta}-\frac{\partial \eta_{ijk}^{l-1}}{\partial \theta}\right), \qquad (A1.1)$$

Table 2: Estimated values of the covariate effects along with their standard errors.

Parameters	Naive Model		Proposed Model	
	Estimates	Standard Error	Estimates	Standard Error
λ_0	-3.091	0.2211	-3.0624	0.2186
λ_1	0.4670	0.1237	0.6944	0.1172
β_{FOOD}	0.0956	0.2082	0.0095	0.1602
β_{AGE}	-0.0030	0.1102	-0.0004	0.0080
β_{GENDER}	0.3095	0.2113	0.3744	0.1865
β_{SMOKE}	-0.1651	0.1619	-0.3023	0.1510
σ_u^2	0.9077	0.1153	0.9513	0.0629
ρ	0.6002	0.0014	0.6000	0.0014

where $I(x)$ is an indicator function, $\eta'_{ijk} =$ logit $(\lambda_1 + x'_{ijk}\beta + z'_{ijk}u_{ij} + v'(t_{ijk})\gamma)$, $\theta = (\lambda_0, \ldots, \lambda_{L-1}, \beta', \gamma')'$, and

$$\left.\begin{array}{rcl} \frac{\partial \eta'_{ijk}}{\partial \lambda_l} & = & \eta'_{ijk}(1-\eta'_{ijk}), \;\; l = 0, \ldots, L-1, \\ \frac{\partial \eta'_{ijk}}{\partial \beta'} & = & \eta'_{ijk}(1-\eta'_{ijk})x'_{ijk} \\ \frac{\partial \eta'_{ijk}}{\partial \gamma'} & = & \eta'_{ijk}(1-\eta'_{ijk})v'(t_{ijk}) \end{array}\right\} \qquad (A1.2)$$

Substituting (A1.2) in (A1.1) we get,

$$\frac{\partial l_{c_1}(\theta)}{\partial \theta} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk} = l\right)\left(1 - \eta'_{ijk} - \eta_{ijk}^{l-1}\right)\tilde{X}_{ijk}, \qquad (A1.3)$$

where $\tilde{X}_{ijk} = (1' x'_{ijk} v'(t_{ijk}))'$.

Second order derivatives:

$$\frac{\partial^2(\theta)}{\partial\lambda_l^2} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk}; l=0,\ldots,L-1$$

$$\frac{\partial^2 l_{c_1}(\theta)}{\partial\beta\partial\beta'} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk} x_{ijk} x'_{ijk}$$

$$\frac{\partial^2 l_{c_1}(\theta)}{\partial\gamma\partial\gamma'} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk} v(t_{ijk}) v'(t_{ijk})$$

$$\frac{\partial^2 l_{c_1}(\theta)}{\partial\beta^T\partial\lambda_l} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk} x'_{ijk}$$

$$\frac{\partial^2 l_{c_1}(\theta)}{\partial\gamma'\partial\lambda_l} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk} v'(t_{ijk})$$

$$\frac{\partial^2 l_{c_1}(\theta)}{\partial\beta'\partial\gamma} = \sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\sum_{l=1}^{L} I\left(y_{ijk}=l\right) b'_{ijk} x_{ijk} v'(t_{ijk})$$

where $b'_{ijk} = -\eta'_{ijk}(1-\eta'_{ijk}) - \eta_{ijk}^{l-1}(1-\eta_{ijk}^{l-1})$

Appendix A2

Proof of Theorem 1 : We give an outline of the proof as it is essentially based on the result of Stone (1985). Equation (4.3) can be proved following Lemma 8 and 9 in Stone (1985). In fact, it can be shown that if the number of knots $k_n \cong O\left(n^{\frac{1}{(2m+1)}}\right)$ then for $m \geq 2$,

$$\frac{1}{nrK}\sum_{i=1}^{n}\sum_{j=1}^{r}\sum_{k=1}^{K}\left(v'(t_{ijk})\hat{\gamma} - f_0(t_{ijk})\right)^2 = O_P\left(n^{\frac{-2m}{(2m+1)}}\right) \qquad (A2.1)$$

Expression (A2.1), in view of Stone (1985) can be expressed as,

$$\int\left\{\hat{f}(t) - f_0(t)\right\}^2 dt = O_P\left(n^{\frac{-2m}{(2m+1)}}\right) \qquad (A2.2)$$

The proof of equations (4.3) are rather straightforward application of Zeng and Cai (2005). Under assumptions A.1–A.6 a solution to equation (3.2) exists and with probability unity, $\hat{\theta} \to \theta^0$.

References

[1] Akaike H (1973). Information theory and an extension of the maximum likelihood principle. *In second International Symposium on Information Theory (BN Petrov and F Csaki, eds.)*. Akademiai Kiado, Budapest.

[2] Breslow NE and Clayton DG (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**(421), 9–25.

[3] Breslow NE and Lin X (1995). Bias correction generalized linear models with single component dispersion. *Biometrika* **82**(1), 81–92.

[4] Celeux G, Chauveau D and Diebolt J (1995). On Stochastic versions of the EM algorithm. *INRIA- Rapport de recherch?* **2514**.

[5] Chib S and Greenberg E (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**(4), 327–335.

[6] Davier MV and Sinharay S (2010). Stochastic Approximation methods for latent regression item response models. *Journal of Educational and Behavioral Statistics*, **35** (2), 174–193.

[7] Delyon B, Lavielle M and Moulines E (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, **27**(1), 94–128.

[8] de Boor C (2001). *A practical guide to splines*. Spring-Verlag.

[9] Greene W. H. (2004). *Econometric Analysis*. Prentice Hall.

[10] Hardle W, Mammen E and Muller M (1998). Testing parametric versus semiparametric modelling in generalized linear models. *Journal of the American Statistical Association*, **93**(444), 1461–1474.

[11] Hardle W, Liang H and Gao J (2000). *Partially linear models*. *Physica-Verlag*.

[12] He X and Shi PD (1996). Bivariate tensor product b-spline in a partly linear model. *Journal of Multivariate Analysis*, **58**(2), 162–181.

[13] He X, Fung WK and Zhu Z (2005). Robust estimation in generalized partial linear models for clustered data. *Journal of the American Statistical Association*, **100**(472), 1176–1184.

[14] Li CS (2011). A lack-of-fit test for parametric zero-inflated Poisson models. *Journal of Statistical Computation and Simulation* **81**(9), 1081–1098.

[15] Lin X and Breslow NE (1996). Bias correction generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, **91**(435), 1007–1016.

[16] Lin X and Carroll RJ (2001a). Semiparametric regression for clustered data. *Biometrics*, **88**(4), 1179–1185.

[17] Lin X and Carroll RJ (2001b). Semiparametric regression for clustered data using generalized estimating equations. *Journal of American Statistical Association*, **99**(466), 1045–1056.

[18] McCullagh P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society B*, **42**(2), 109–142.

[19] McCulloch CE (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, **89**(428), 330–335.

[20] McCulloch CE (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**(437), 162–170.

[21] Meza C, Jaffrezic F and Foulley JL (2009). Estimation in the probit normal model for binary outcomes using the SAEM algorithm. *Computational Statistics and Data Analysis*, **53**(4), 1350–1360.

[22] Muller M (2001). Estimation and testing in generalized partial linear models - a comparative study. *Statistics and Computing*, **11**(4), 299–309.

[23] Natarajan R, McCulloch CE and Kiefer NM (2000). A Monte Carlo EM method for estimating multinomial probit models. *Computational Statistics and Data Analysis*, **34**(1), 33–50.

[24] Pinheiro JC and Bates DM (1995). Approximations to the log-likelihood function in nonlinear mixed-effects models. *Journal of Computational Graphical Statistics*, **4**(1), 12–35.

[25] Severini TA and Staniswalis JG (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, **89**(428), 501–511.

[26] Sinha SK (2004). Robust analysis of generalized linear mixed models. *Journal of the American Statistical Association*, **99**(466), 451–460.

[27] Smith AFM and Roberts GO (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, **55**(1), 3–24.

[28] Stone C(1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**(2), 689–705.

[29] Turskey S, Gilmore ND and Glickman IR (1970). Reduced plaque formation by the chloromethyl analogue of Vit-C. *Journal of Periodontology*, **41**(1), 41–43.

[30] Wang N, Carroll RJ and Lin X (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, **100**(469), 147–157.

[31] Wooldridge J (2003). *Econometric analysis of cross section and panel data*. MIT press.

[32] Zeger SL and Karim MR (1991). Generalized linear model with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, **86**(413), 79–86.

[33] Zeng D and Cai J (2005). Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Annals of Statistics*, **33**(5), 2132–2163.

[34] Zhou X and Liu X (2008). The Monte Carlo EM method for estimating multinomial probit latent variable models. *Computational Statistics*, **23**(2), 277–289.