Threshold Models with Correlated Random Components

Ayoub Saei

Abstract

Threshold models for ordinal longitudinal/repeated measurements data need to account for the correlation between observations within each subject. Two variance-covariance structures are introduced for two types of threshold models. Estimation of the parameters in the linear predictor and components of variance are derived in the general form. A small simulation study is carried out to support the estimation and inference methods. These methods are then applied to two practical examples.

S3RI Methodology Working Paper M04/07

Threshold Models with Correlated Random Components

Ayoub Saei

Southampton Statistical Sciences Research Institute, University of Southampton,

Highfield, Southampton SO17 1BJ, UK

Summary. Threshold models for ordinal longitudinal/repeated measurements data need to account for the correlation between observations within each subject. Two variance-covariance structures are introduced for two types of threshold models. Estimation of the parameters in the linear predictor and components of variance are derived in the general form. A small simulation study is carried out to support the estimation and inference methods. These methods are then applied to two practical examples.

Keywords: Correlation; Longitudinal; ML; Ordinal; REML

1. Introduction

Ordinal or ordered categorical data often occur as response variables in statistical applications, although some arise from grouping an underlying continuous random variable. McCullagh and Nelder (1989) have summarised different situations, which result in such polytomous data. A particular polytomous data are ordinal response variables. A flexible and widely used model in analysing ordinal responses is the threshold model (ordinal regression model McCullagh 1980).

In many applications, the dependence structure is more complex than the independent observations assumed by the ordinal regression model of McCullagh (1980). In particular, observations often exhibit a clustering, and observations within clusters are correlated. A common instance of this occurs with longitudinal and repeated measurements where a cluster consists of the set of observations for a given subject. In this paper, we develop two types of models for the ordinal response variables, time dependent and time independent threshold models with correlated random effects.

Random component threshold models are applied to dairy produce testing by Zhaorong et al (1992) and Saei and McGilchrist (1996). These models are very versatile for analysing two complex data sets in Saei et al (1996) and Saei and McGilchrist (1997). Jansen (1990 and 1992) use random effects threshold models for extra variation and nested errors in analysing ordinal response data. Threshold model with random effects following the Hougaard (1986) family distribution is given in Crouchley (1995). Hedeker and Gibbons (1994) apply the multilevel random effects threshold models and use Gauss-Hermite quadrature within a Fisher scoring algorithm. Ten Have (1996) use cumulative comp-lementtary log-log link and a log-gamma random effects distribution. The longitudinal ordinal response data are analysed by maximum likelihood and generalized estimating equations (GEE) in Kenward, et al (1994) and Miller et al (1993) by GEE and weighted least squares.

Longitudinal and repeated measures models need to account for the correlation between observations within subjects. The models also need to show possible changes of the threshold parameters (cut-points) over time/space. This paper introduces two random components threshold models in analogue to Saei and McGilchrist (1998). The models allow the observations within a subject to be correlated. Sections 2 and 3 give models and estimation procedures. In sections 4 and 5 estimation equations are simplified for two particular types of variance-covariance matrices. Section 6 presents a small simulation study. Sections 7 and 8 apply the estimation methods to the unpublished microbial plaque data and respiratory disorder data of Koch et al (1990). The results of the paper are discussed in the last section.

2. Models

Suppose that repeated observations of an ordinal response variable Y are obtained (Y_{it}) at times t = 1,2,...,n from subject i, i=1, 2,...,N; Y_{it} can take on values 1, 2, ..., M. The distribution of Y_{it} depends on the linear predictor $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}$, where $\mathbf{x}_{it}$ is a vector of p known regression variables with fixed regression coefficient $\boldsymbol{\beta}$ and u_{it} is a random variable. The random variables u_{it} and u_{is} for a subject are correlated; those from distinct subjects are independent. The random vector $\mathbf{u}_i = [u_{i1}, u_{i2}, \ldots u_{in_i}]'$ is assumed to follow multivariate normal distribution with variance-covariance matrix that is characterised by variance and correlation parameters φ and ρ. The observable variable Y_{it} is the categorised value of an unobservable continuous variable V_{it} with conditional mean of η_{it} =E($\mathrm{V}_{it}|u_{it}$). Let θ_k be the threshold parameter for the response category k and θ_{kt} be the corresponding parameter at time t. Two cumulative distribution functions for Y_{it} conditional on u_{it}, are then

$$\mathrm{P}(\mathrm{Y}_{it} \le k) = \mathrm{G}(\theta_k - \eta_{it}) \tag{2.1}$$

$$\mathrm{P}(\mathrm{Y}_{it} \le k) = \mathrm{G}(\theta_{kt} - \eta_{it}) \tag{2.2}$$

where k = 1, 2,..., M and G(.) is cumulative distribution function for an unobservable continuous random variable V_{it} with conditional mean η_{it}. Different choices of the distribution function G(.) result in different threshold models. Four commonly considered threshold models correspond to the four common distributions; standard normal, logistic, extreme minimal and extreme maximal for G(.). The results of threshold model with logistic distribution for G(.) are reported in this paper. Models (2.1) and (2.2) are called *time independent* and *time dependent random component threshold models* respectively.

The parameter θ_0 is always taken as -∝, so that G(θ_0 - η_{it}) = 0 while θ_M is taken to be +∞, so that G(θ_0 - η_{it}) = 1. If $\boldsymbol{\eta}$ contains a constant term there is a lack of identifiability since any quantity added to all $\boldsymbol{\theta}$ values can be compensated by adding it also to the constant terms $\boldsymbol{\eta}$. This lack of identifiability is resolved by setting θ_1 = 0. Thus there are M - 2 and n(M - 2) unknown threshold parameters (θs) under models (2.1), (2.2) respectively, where n is the minimum of the n_is. The threshold parameters are collected into a vector $\boldsymbol{\theta}$. In general $\boldsymbol{\eta}$ can be expressed as $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, where $\mathbf{X}$ is a known matrix of regression variables, $\mathbf{Z}$ is the incidence matrix for the random effect vector $\mathbf{u}$ and $\mathbf{u} = [\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_N]'$ ~N($\mathbf{0}$, $\varphi\mathbf{A}$) with $\mathbf{A} = \mathrm{diag}(\mathbf{A}_i)$. The $\mathbf{A}_i$ are $n_i \times n_i$ matrices; their elements are functions of the correlation parameter ρ.

3. Estimation

In linear mixed models, best linear unbiased predictors (BLUP), Henderson (1963 1973,1975) are used to obtain maximum likelihood (ML) and restricted or residual maximum likelihood (REML) in Harville (1977), Thompson (1980), Fellner (1986, 1987) and Speed (1991). McGilchrist (1994) extends this approach to generalized linear mixed models for a broader class of problems. The method has elements in common with Schall (1991), Breslow and Clayton (1993), Wolfinger (1993), Nelder and Lee (1996) and Saei and McGilchrist (1998). Lee and Nelder (2001a, 2001b) further extend work of Nelder and Lee (1996) to the correlated non-normal data. An outline of the extension of Saei and McGilchrist (1998) to the correlated ordinal data is given here. Let l_1 be the loglikelihood function of ordinal observations conditional on the random component vector $\mathbf{u}$ taken to be fixed and l_2 the logarithm of the probability density function of $\mathbf{u}$. The functions l_1 and l_2 are

$$l_1 = \sum_{i=1}^{N} \sum_{t=1}^{n_i} \ln \Delta_{it}$$
$$l_2 = -(1/2)[\text{const.} + N_0 \ln \varphi + \ln|\mathbf{A}| + \varphi^{-1}\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}]$$

where $\Delta_{is} = \mathrm{G}(\theta_{y_{is}} - \eta_{is}) - \mathrm{G}(\theta_{y_{is}-1} - \eta_{is})$ and $\Delta_{is} = \mathrm{G}(\theta_{sy_{is}} - \eta_{is}) - \mathrm{G}(\theta_{s(y_{is}-1)} - \eta_{is})$ undser models (2.1) and (2.2) and $N_0 = \sum_{i=1}^{N} n_i$. The following steps obtain the ML estimators and their approximate variance-covariance matrices. Step 1. Estimates $\tilde{\theta}$, $\tilde{\beta}$ and $\tilde{\mathbf{u}}$ are obtained by maximising $l = l_1 + l_2$ for the given initial values of φ and ρ. These estimates are then used as an initial step in finding ML and REML estimates of φ and ρ via Anderson (1973) and Henderson (1973) algorithm. Starting from initial values $k = 0$, $\boldsymbol{\theta}_0$, $\boldsymbol{\beta}_0$, $\mathbf{u}_0$, φ_0 and ρ_0 (hence $\mathbf{A}_0$), the estimating equations set out by using the Newton-Raphson method

$$\begin{bmatrix}\boldsymbol{\theta}_{k+1}\\ \boldsymbol{\beta}_{k+1}\\ \mathbf{u}_{k+1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\theta}_{k}\\ \boldsymbol{\beta}_{k}\\ \mathbf{u}_{k}\end{bmatrix}+\mathbf{V}^{-1}\begin{bmatrix}\mathbf{I}_{\boldsymbol{\theta}} & 0\\ 0 & \mathbf{X}'\\ 0 & \mathbf{Z}'\end{bmatrix}\begin{bmatrix}\partial l_1/\partial\boldsymbol{\theta}_k\\ \partial l_1/\partial\boldsymbol{\eta}_k\end{bmatrix}-\mathbf{V}^{-1}\begin{bmatrix}0\\ 0\\ \mathbf{A}_0^{-1}\mathbf{u}_k\end{bmatrix} \tag{3.1}$$

where

$$\mathbf{V}=\begin{bmatrix}\mathbf{I}_{\boldsymbol{\theta}} & \mathbf{0}\\ \mathbf{0} & \mathbf{X}'\\ \mathbf{0} & \mathbf{Z}'\end{bmatrix}\left(-\begin{bmatrix}\partial^2 l_1/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}'_k & \partial^2 l_1/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\eta}'_k\\ \partial^2 l_1/\partial\boldsymbol{\eta}_k\partial\boldsymbol{\theta}'_k & \partial^2 l_1/\partial\boldsymbol{\eta}_k\partial\boldsymbol{\eta}'_k\end{bmatrix}\right)\begin{bmatrix}\mathbf{I}_{\boldsymbol{\theta}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{X} & \mathbf{Z}\end{bmatrix}+\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & \varphi_0^{-1}\mathbf{A}_0^{-1}\end{bmatrix},$$

$\partial l_1/\partial\boldsymbol{\theta}_k$, $\partial l_1/\partial\boldsymbol{\eta}_k$, $\partial^2 l_1/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}'_k$, $\partial^2 l_1/\partial\boldsymbol{\theta}_k\partial\boldsymbol{\eta}'_k$ and $\partial^2 l_1/\partial\boldsymbol{\eta}_k\partial\boldsymbol{\eta}'_k$ are evaluated at initial estimates of the parameters.

Step 2. Once iterations of (3.1) have converged to $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$, let $\tilde{\boldsymbol{\eta}}=\mathbf{X}\tilde{\boldsymbol{\beta}}+\mathbf{Z}\tilde{\mathbf{u}}$, $\mathbf{B}=-\partial^2 l_1/\partial\tilde{\boldsymbol{\eta}}\partial\tilde{\boldsymbol{\eta}}'$, $\mathbf{T}^*=[\varphi_0^{-1}\mathbf{A}_0^{-1}+\mathbf{Z}^{\mathrm{T}}\mathbf{B}\mathbf{Z}]^{-1}$. The ML estimate of the variance component parameter φ is estimated by

$$\hat{\varphi}=(\textstyle\sum_{i=1}^{N}n_i)^{-1}[\mathrm{tr}(\mathbf{A}^{-1}(\rho_0)\mathbf{T}^*)+\tilde{\mathbf{u}}'\mathbf{A}^{-1}(\rho_0)\tilde{\mathbf{u}}] \tag{3.2}$$

The evaluation of (3.1) and (3.2) is then iterated with initial values set to $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\beta}}$, $\tilde{\mathbf{u}}$ and $\hat{\varphi}$. Step 3. After convergence of steps 1 and 2, the ML estimate of the correlation parameter ρ is

$$\hat{\rho}=h(.) \tag{3.3}$$

where $h(.)$ is a function of $\mathrm{tr}\{(\partial\mathbf{A}^{-1}/\partial\rho_0)\mathbf{A}(\rho_0)\}$, $\hat{\varphi}_{(\mathrm{ML})}^{-1}[\mathrm{tr}\{(\partial\mathbf{A}^{-1}/\partial\rho_0)\mathbf{T}^*\}]$ and $\tilde{\mathbf{u}}'(\partial\mathbf{A}^{-1}/\partial\rho_0)\tilde{\mathbf{u}}$. Note that evaluation of the $h(.)$ components depends on varinac-covariance matrix of the random effect $\mathbf{u}$. The new values $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\beta}}$, $\tilde{\mathbf{u}}$, $\hat{\varphi}$ and $\hat{\rho}$ are substituted for $\boldsymbol{\theta}_0$, $\boldsymbol{\beta}_0$, $\mathbf{u}_0$, φ_0 and ρ_0 in (3.1), and then (3.1) - (3.3) are evaluated again. At convergence, $\hat{\varphi}$ and $\hat{\rho}$ are ML estimates ($\hat{\varphi}_{(\mathrm{ML})}$ and $\hat{\rho}_{(\mathrm{ML})}$) of φ and ρ with associated asymptotic variance-covariance matrix of

$$\mathrm{Var}\left(\begin{bmatrix}\hat{\varphi}_{(\mathrm{ML})}\\ \hat{\rho}_{(\mathrm{ML})}\end{bmatrix}\right)=2\begin{bmatrix}\varphi^{-2}(\mathrm{N}_0-2r_1)+\varphi^{-4}r_{11} & \varphi^{-1}(2k_1-\nu_1-\varphi^{-1}k_{11})\\ . & (\varphi^{-2}k_{11}^{(11)}+\nu_{11}-2k_1^{(11)})\end{bmatrix}^{-1} \quad (3.4)$$

where $N_0=\sum_{i=1}^{N}n_i$; r_1, r_{11}, k_1, v_1, v_{11}, k_{11}, $k_{11}^{(11)}$ and $k_1^{(11)}$ are given in appendix A.

Let $\mathbf{V}=\begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12} & \mathbf{V}_{13}\\ \mathbf{V}_{21} & \mathbf{V}_{22} & \mathbf{V}_{23}\\ \mathbf{V}_{31} & \mathbf{V}_{32} & \mathbf{V}_{33}\end{bmatrix}$ and $\mathbf{V}^{-1}=\begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12} & \mathbf{T}_{13}\\ . & \mathbf{T}_{22} & \mathbf{T}_{23}\\ . & . & \mathbf{T}_{33}\end{bmatrix}$ be the partitions of the matrix $\mathbf{V}$ and its inverse corresponding to the dimensions of the $\boldsymbol{\theta}$, $\boldsymbol{\beta}$ and $\mathbf{u}$. Replacing $\mathbf{T}^*$ by $\mathbf{T}_{33}$ in all above three steps (3.1 - .3.3) yields REML estimates of the parameters and variance components φ and ρ. Similarly, replacing $\mathbf{T}^*$ by $\mathbf{T}_{33}$ in (3.4) gives the asymptotic variance-covariance matrix for the REML estimators $\hat{\varphi}_{(\mathrm{REML})}$ and $\hat{\rho}_{(\mathrm{REML})}$.

4. Exchangeable and AR(1) Models

A simple form of dependence arises when the random components of $\mathbf{u}$ have an exchangeable correlation (constant correlation) ρ. The variance covariance matrix for the random components $\mathbf{u}$ is $\rho\mathbf{A}$ where $\mathbf{A}$ is a block diagonal matrix with blocks $\mathbf{A}_i$ = $(1-\rho)\mathbf{I}_i+\rho\mathbf{J}_i$ and i = 1,2,..., N; $\mathbf{I}_i$ is an $n_i\times n_i$ identity matrix and $\mathbf{J}_i$ is a $n_i\times n_i$ matrix with all elements equal to 1. Let $\mathbf{T}^*=\left[\mathbf{T}_{ij}^*\right]$ with $\mathbf{T}_{ij}^*$ as a $n_i\times n_j$ matrix, the ML estimation equation for ρ, i.e., $h(.)$ in (3.3) is then

$$\hat{\rho}=-(\mathrm{a}\rho_0^2+\mathrm{b}\rho_0+\mathrm{c})^{-1}\mathrm{d}\ ,$$

where

$$\mathrm{a}=\sum_{i=1}^{N}n_i(n_i-1)^2,\ \mathrm{b}=\sum_{i=1}^{N}[\varphi^{-1}((n_i-1)^2(\mathrm{b}_{1i}+\mathrm{b}_{2i})-(n_i-1)(\mathrm{b}_{3i}+\mathrm{b}_{4i}))+n_i(n_i-1)(2-n_i)]$$
$$\mathrm{c}=2\sum_{i=1}^{N}[\varphi^{-1}(\mathrm{b}_{1i}+\mathrm{b}_{2i})-n_i(n_i-1)],\ \mathrm{d}=\sum_{i=1}^{N}\varphi^{-1}[(\mathrm{b}_{1i}+\mathrm{b}_{2i})-(\mathrm{b}_{3i}+\mathrm{b}_{4i})],\ \mathrm{b}_{1i}=\tilde{\mathbf{u}}_i'\,\tilde{u}_i,$$
$$\mathrm{b}_{2i}=\mathrm{tr}(\mathbf{T}_{ii}^*),\ \mathrm{b}_{3i}=\tilde{\mathbf{u}}_i'\mathbf{J}_i\tilde{\mathbf{u}}_i \text{ and } \mathrm{b}_{4i}=\mathrm{tr}(\mathbf{J}_i\mathbf{T}_{ii}^*).$$

The autoregressive first order, AR(1) assumes that the dependence of the observations for a subject decreases with their distance in time. In this case, the variance covariance matrix of the random components **u** is $\varphi\mathbf{A}$, where **A** is a block diagonal matrix with the blocks $\mathbf{A}_i = (1-\rho^2)^{-1}[a_{st}]$, $a_{st} = \rho^{|s-t|}$, and s, t = 1, 2,…, n_i. Let $\mathrm{a} = 2(1-\rho^{-2})^{-1}N$,

$\mathrm{b} = \sum_{i=1}^{N} \mathrm{tr}(\mathbf{T}^*_{ii}\mathbf{\Lambda}_i) + \sum_{i=1}^{N} \tilde{\mathbf{u}}'_i \mathbf{\Lambda}_i \tilde{\mathbf{u}}_i$ and $\mathrm{c} = \sum_{i=1}^{N} \mathrm{tr}(\mathbf{T}^*_{ii}\mathbf{\Gamma}_i) + \sum_{i=1}^{N} \tilde{\mathbf{u}}'_i \mathbf{\Gamma}_i \tilde{\mathbf{u}}_i$, the h(.) in (3.3) is

$$\hat{\rho} = -(\mathrm{a} + 2\hat{\varphi}^{-1}\mathrm{b})^{-1}\hat{\varphi}^{-1}\mathrm{c}$$

where $\mathbf{\Lambda}_i$ is a $n_i \times n_i$ diagonal matrix with diagonal elements (0,1,…,1,0) and $\mathbf{\Gamma}_i$ is a $n_i \times n_i$ matrix and has -1 above and below principal diagonal and zero all other elements. The REML counterparts are obtained by replacing $\mathbf{T}^*$ by $\mathbf{T}_{33}$.

5. Simulation

A limited simulation study was undertaken to examine the performance of the method for several parameter combinations. Observations were generated from time independent threshold model $\mathrm{P}(\mathrm{Y}_{\mathrm{it}} \leq \mathrm{y}) = \mathrm{G}(\theta_y - \eta_{it})$ where G(.) is the logistic function and $\eta_{it} = \beta_0 + x_{it}\beta_1 + u_{it}$. The u_{it} are generated from an AR(1) with parametees ρ and φ. The x_{it} are randomly assigned to 1 or 0 and observations are obtained at time t = 1, 2,…, 5 for i = 1, 2,…, 30. Estimates of parameters are obtained, and the process is replicated 1000 times. Table 1 shows the results by REML method for the different combinations of the parameters in the variances in which ρ and φ are allowed to change. The threshold and fixed parameters are held constant. The quantities are reported in Table 1 are the true parameters values (TV), the average of the biases (AB), standard error over 1000 simulations (SE1) and the average asymptotic standard error of 1000 simulations (SE2).

The results indicate that both ML (not reported in the paper) and REML estimators of threshold parameters and variances components ρ and φ are negatively biased. The

REML method reduces biases of the estimators. With increased variance components the bias of the estimator φ is increased. The relative bias (bias/(true value)) of estimating ρ is much smaller than for φ. The bias of the REML estimator of ρ is not significant for the parameter set 3. The biases of the estimators of β_0 and β_1 are small and their standard errors over simulation SE1 are in agreement with SE2 in both ML and REML methods. This is also true for the correlation parameter ρ. However, the SE2 of the estimator of φ is greater than SE1. This indicates that the method overestimates the standard error of the estimator of φ. However, the results of other simulations (not reported) show that the difference between SE1 and SE2 decreases by increasing the number of subjects from 30 to 100.

6. Application To Microbial Plaque Data

The preceding theory is applied to unpublished microbial plaque data. The data are obtained from Nasr Isfahani. A study was planned to determine the effective time of brushing teeth according to Bass method by Nasr Isfahani (1999). A total of 60 dentist students were randomly selected and assigned to 6 groups of 10 each students. The students were asked not to use any procedure to remove plaque 24 hours before the visit. At the baseline and after brushing teeth for 1, 2,…, 6 minutes, the plaque indices (sinless and loe plaque index) were recorded on a four point ordinal scale (0 = excellent, 1 = good, 2 = not bad and 3 = bad) for all four tooth surfaces (mesial, distal, fasial and lingual). Data on tooth number 1 in the upper-left quadrant are used here. The model fitted is

$$\eta_{is} = \beta_0 + \beta_t(\text{time}_i) + \omega_j(\text{base}_i) + u_{is}$$

where β_0 is an intercept, β_t is the time effect, time_i is the brushing time for ith student, ω_j is the initial value effect, bas_i is the initial value of plaque of the ith student at base-line and u_{is} is the sth tooth surface (s =1= mesial, s = 2 =distal, s = 3 = fasial and s = 4 = lingual) random effect for student i. Random vector $\mathbf{u}_i = (u_{i1}, u_{i2}, ..., u_{i4})'$ is a multivariate normal with zero mean and exchangeable variance-covariance structure. Note that the AR(1) model is not fitted since the distance between two-tooth surfaces can not be meaningfully defined.

Table 2 shows the results for the time dependent threshold model (2.2). There is a significant variation between students with REML estimate value of φ = 5.55 and standard error of 2.03. The random effects for a student are highly correlated.

The paired comparisons show that there is a statistically significant changes from β_1 to β_2 but not from β_2, β_3, …, β_6. This indicates that the effective time to remove plaque is between 2 and 3 minutes. The initial value of plaque has no statistically significant effect.

The results also shows that the threshold parameters θ_{ks} (k =1,2 and s = 1, 2, …, 4) have statistically significant changes over four tooth surfaces (mesial, distal, fasial and lingual). The model (2.2) yields the same predicted values as observed in 88.3%.

7. Application to Respiratory Disorder Data

A second application is to respiratory disorder given by Koch et al (1990). A total of 111 patients within two centres were randomly assigned to two treatments (active, placebo). At the baseline, status of each patient was recorded according to a five point ordinal response scale (0 = terrible, 1 = poor, 2 = fair, 3 = good, 4 = excellent), and also at each of four visits (visit 1, visit 2, visit 3, visit 4) during the time period over which the treatments were administrated. Patients' characteristics like age and gender were also recorded at the time of entry to the study. The model is

$$\eta_{it} = \beta_0 + \alpha(\mathrm{c}_i) + \beta_t(\mathrm{treat}_{it}) + \gamma(\mathrm{age}_i) + \lambda(\mathrm{gen}_i) + \omega_j(\mathrm{stut}_i) + u_{it}$$

where risk variables are centre (c_i, 1 = centre 1, 0 = centre 2), age in years at base-line (age_i), gender (gen_i, 1 = male, 2 = famle), status at base-line ($stut_i$, j = 1, 2,…, 5), treatment ($treat_{it}$, 1 = active, 0 = placebo) and u_{it} is ith patient random effect. The results of fitting model (8.1) via REML method for AR(1) are given in Table 3.

The results show that the random components are highly correlated for a patient with REML estimate of $\rho = 0.9$ and standard error of 0.04. The center, gender effects α, λ and age regression coefficient γ are not significant. The initial status of patients has a statistically significant effect. It indicates that patients with status in lower category (terrible) at base-line tend to give response in lower category in the end of study.

Although the treatment effect β_t increases from the first visit to the second and then decreases, the paired comparisons show that the changes are not statistically significant.

By using asymptotic distribution of the estimators β_t and assuming that the β_t are constant over time, it can be seen that the treatment is highly statistically significant. The estimate of treatment effect is positive, indicating that patients in the treatment active group are more likely to respond in the higher categories. The results (not reported in the paper) of fitting model (2.2) show that the changes of the threshold parameters $\boldsymbol{\theta}$ over time are not statistically significant. Under model (2.1), the observed and predicted values agree in 66.8% and 72.8% for an exchangeable and an AR(1) models respectively. These percentages show that the model (2.1) with AR(1) correlation structure is a suitable model for this data.

8. Discussion

A simple method of analysing the correlated ordinal response data is presented. Two types of modelling, time independent and time dependent threshold models are used. The latter model takes into account possible changes of the threshold parameters over time. The estimation and inference approaches have been applied to two data sets. The results (Tables 2 and 3) show that observations are highly correlated within clusters in both applications. The threshold parameters have significant changes over four tooth surfaces in the first application. These parameters are constant over time period of study in the second application. Although the threshold parameters are changing over four tooth surfaces, the conclusions do not change from the model (2.1) to the model (2.2). However, in another application (submitted for the publication) the significant changes of the threshold parameters yielded different conclusions from the models (2.1) and (2.2).

The results of a limited simulation (Table 1) show that the REML estimates of the fixed effect and threshold parameters are very good. The REML estimates are even unbiased for some of those parameters. The REML estimator of the variance parameter φ is negatively biased. However, the bias of the REML estimator for the correlation parameter ρ is very small. The method provides very good estimates of the standard error of $\boldsymbol{\beta}$ and correlation parameter ρ. It overestimates the standard error of the estimator of φ.

The goodness of fit of the model is based on a simple method of Saei and McGilchrist (1998). Further work is needed on the goodness of fit of the model. The biases in estimation of the variance component φ and its standard error are the other problems of the method.

Appendix A

The elements in the asymptotic variance-covaraince matrix (3.4) for the ML estimators of the variance components φ and ρ are;

$r_1 = \varphi^{-1}\mathrm{tr}(\mathbf{A}^{-1}\mathbf{T}^*)$, $r_{11} = \mathrm{tr}(\mathbf{T}^*\mathbf{A}^{-1}\mathbf{T}^*\mathbf{A}^{-1})$, $k_1 = \varphi^{-1}\mathrm{tr}[(\partial\mathbf{A}^{-1}/\partial\rho)\mathbf{T}^*]$,

$\nu_1 = \mathrm{tr}[(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{A}]$, $\nu_{11} = \mathrm{tr}[(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{A}^{-1}(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{A}^{-1}]$,

$k_{11}=$ $\mathrm{tr}[\mathbf{T}^*(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{T}^*\mathbf{A}^{-1}]$, $k_{11}^{(11)} = \mathrm{tr}[\mathbf{T}^*(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{T}^*(\partial \mathbf{A}^{-1}/\partial\rho)]$

and

$k_1^{(11)} = \varphi^{-1}\mathrm{tr}[(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{T}^*(\partial \mathbf{A}^{-1}/\partial\rho)\mathbf{A}]$.

The replacement of $\mathbf{T}^*$ by $\mathbf{T}_{33}$ yields corresponding components for the asymptotic variance-covariance matrix of the REML estimtors of φ and ρ.

References

Anderson, T.W. (1973) Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics* **1**, 135-141.

Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalised linear mixed models. *Journal of the American Statistical Association* **88,** 9-25.

Crouchley, R. (1995) A random effects model for ordered categorical data. *Journal of the American Statistical Association* **90,** 489-498.

Hedeker, D. and Gibbons, R.D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933-944.

Henderson, C.R. (1963) Selection index and expected genetic advance. *In Statistical Genetics and Plant Breeding* (W.D. Hanson and H.F. Robinson, eds.), 141-163. National Academy of Sciences and National Research Council Publication No. 982, Washington, D.C.

Henderson, C.R. (1973) Sire evaluation and genetic trends. *In proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush* 10-41. Amer. Soc. Animal Sci.-Amer. Dairy Sci. Assn.-Poultry Sci. Assn. Champaign, III.

Henderson, C.R. (1973) Maximum likelihood estimation of variance components, Unpublished manuscript.

Henderson, C.R. (1975) Best linear unbiased estimation and prediction under selection model. *Biometrics* **31**, 423-447.

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrics* **73**, 387-396.

Jansen, J. (1990) On the statistical analysis of ordinal data when extra variation is present. *Applied Statistics* **39**, 75-84.

Jansen, J. (1992) Statistical analysis of ordinal data from experiments with nested errors. *Computational Statistics & Data Analysis* **13**, 319-330.

Kenward, M.G., Lesaffre, E. and Molenberghs, G. (1994) An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random. *Biometrics* **50**, 945-953.

Koch,G.G., et al (1990) Categorical data analysis. *Statistical Methodology in the Pharmaceutical Sciences***.** Eds. Donald A. Berry

McCullagh, P. (1980) Regression models for ordinal data*. Journal of the Royal Statistical Society* **B, 42**, 109-142.

McCullagh, P. and Nelder, J.A. (1989) *Generalized linear models*. Second Ed. London:Chapman and Hall.

McGilchrist, C.A. (1994) Estimation in generalized mixed models. *Journal of the Royal Statistical Society Series* **B 56, 61-69**.

Miller, M.E., Davis, C.S. and Landis, J.R. (1993) The analysis of longitudinal polytomous data: generalized estimation equations and connections with weighted least squares. *Biometrics* **49**, 1033-1044.

Nelder, J.A. and Lee, Y. (1996) Hierarchical generalized linear models. *Journal of the Royal Statistical Society Series* **B 58, 619-678**.

Lee, Y. and Nelder, J.A. (2001a) Hierarchical generalized linear models: A synthesis of generalised linear models, random-effect model and structured dispersions. *Biometrika* **88**, 987-1006.

Lee, Y. and Nelder, J.A. (2001b) Modelling and analysing correlated non-normal data. *Statistical Modelling* **1**, 3-16

Nasr Isfahani, M. (1999) Effective time brushing to remove macrobial plaque by Bass method. MD thesis.

Saei, A. and McGilchrist, C.A. (1996) Random component threshold models. *Journal of Agricultural, Biological and Environmental Statistics* **1**, 288-296

Saei, A. and McGilchrist, C.A. (1997) Random threshold models for inflated zero class data. *Australian Journal of Statistics* **39**, 5-16.

Saei, A., Ward, J. and McGilchrist, C.A. (1996) Threshold models in methadone program evaluation. *Statistics in Medicine* **15**, 2253-2260.

Saei, A. and McGilchrist, C. (1998) Longitudinal threshold models with random components. *The Statistician, Series* **D 47**, 365-375.

Schall, R (1991) Estimation in generalised linear models with random effects. *Biometrika* **78, 719-727**.

Speed, T. (1991) Comment on Robinson: Estimation of random effect. *Statistical Science 6*, 42-44.

Ten Have T. (1996) A mixed effects model for multivariate ordinal response data including correlated discrete failure time with ordinal responses. *Biometrics* **53**, 473-491.

Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika* **80,** 791-796.

Zhaorong, J., McGilchrist, C.A. and Jorgensen, M.A. (1992) Mixed model discrete regression. *Biometrical Journal* ***34***, 691-700.

Table 1. REML estimation for the simulated data.

		φ	ρ	θ_1	θ_2	θ_3	θ_4	β_0	β_1
	TV	0.5	0.1	0.5	2	3.5	5	0.5	1
Par 1	AB	-0.06	-0.08	0.0	-0.03	-0.06	0.07	-0.01	0.06
	SE1	0.11	0.44	0.11	0.21	0.31	0.56	0.25	0.24
	SE2	0.46	0.58	0.12	0.22	0.34	0.66	0.26	0.25
	TV	0.5	0.3	0.5	2	3.5	5	0.5	1
Par 2	AB	-0.05	-0.11	0.1	-0.05	-0.06	0.03	-0.02	0.04
	SE1	0.12	0.44	0.12	0.22	0.34	0.51	0.26	0.24
	SE2	0.46	0.56	0.12	0.22	0.34	0.66	0.26	0.25
	TV	1	0.7	0.5	2	3.5	5	0.5	1
Par 3	AB	-0.45	0.01	-0.04	-0.21	-0.33	-0.31	-0.04	0.04
	SE1	0.26	0.21	0.11	0.22	0.32	0.54	0.31	0.34
	SE2	0.48	0.25	0.12	0.22	0.31	0.54	0.34	0.33

* Par i = Parameter set i (i = 1, 2, 3).

Table 2. REML estimates of parameter (Est), standard errors (SE),z-value (Est/SE).

Par	Est	SE	Est/SE
φ	5.55	2.03	2.73
ρ	0.91	0.14	6.5
θ_{11}	4.33	0.86	5.05
θ_{12}	5.9	1.04	5.68
θ_{13}	5.66	1.02	5.58
θ_{14}	5.66	1.01	5.58
θ_{21}	4.62	0.87	5.31
θ_{22}	8.39	1.44	5.84
θ_{23}	8.12	1.37	5.92
θ_{24}	8.12	1.37	5.92
β_0	-2.22	0.76	-2.93
β_1	6.22	1.68	3.7
β_2	10.62	1.9	5.59
β_3	7.62	1.7	4.49
β_4	7.31	1.67	4.39
β_5	5.24	1.72	3.05
β_6	5.68	1.73	3.29
ω_1	-0.39	0.9	-0.43
ω_2	-0.21	0.89	-0.24
ω_3	0.23	0.83	0.28

- Par = Parameter
- ω4 is fixed at zero to achieve identifiability

Table 3. REML estimates of parameter (Est), standard error (SE),z-value (Est/SE).

Par	Est	SE	Est/SE
φ	0.75	0.37	2.01
ρ	0.9	0.04	22.5
θ_1	1.55	0.23	6.78
θ_2	4.05	0.3	13.74
θ_3	5.97	0.33	17.94
β_0	7.01	1.11	6.29
β_1	1.65	0.53	3.1
β_2	2.36	0.55	4.31
β_3	2.08	0.54	3.82
β_4	1.57	0.54	2.93
α	-0.62	0.49	-1.27
γ	-0.03	0.02	-1.52
λ	-0.56	0.61	-0.92
ω_1	-4.46	1.5	-2.97
ω_2	-3.78	0.83	-4.55
ω_3	-2.41	0.73	-3.32
ω_4	-0.66	0.74	-0.9

*ω_5 is fixed at zero to achieve identifiability