Tutz:

Modelling of repeated ordered measurements by isotonic sequential regression

Sonderforschungsbereich 386, Paper 298 (2002)

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

Modelling of repeated ordered measurements by isotonic sequential regression

Gerhard Tutz

Ludwig-Maximilians-Universität München, Institut für Statistik

Akademiestrasse 1, 80799 München

tutz@stat.uni-muenchen.de

Summary

The paper introduces a simple model for repeated observations of an ordered categorical response variable which is isotonic over time. It is assumed that the measurements represent an irreversible process such that the response at time t is never lower than the response observed at the previous time point $t-1$. Observations of this type occur for example in treatment studies when improvement is measured on an ordinal scale. Since the response at time t depends on the previous outcome, the number of ordered response categories depends on the previous outcome leading to severe problems when simple threshold models for ordered data are used. In order to avoid these problems the isotonic sequential model is introduced. It accounts for the irreversible process by considering the binary transitions to higher scores and allows a parsimonious parameterization. It is shown how the model may easily be estimated by using existing software. Moreover, the model is extended to a random effects version which explicitly takes heterogeneity of individuals and potential correlations into account.

Key words:
Ordinal data, cumulative model, sequential model, repeated measurements, isotonic ordinal regression, random effects models.

1 Introduction

In many studies in clinical and epidemiologic research a response variable is measured repeatedly for each subject. Quite frequently the response variable is categorical with an ordering of the categories refering to severity of disease or level of improvement following an intervention. The problem considered here is of the latter type based on a data set given by Davis (1991). In a randomized study 60 children undergoing surgery were treated with one of four dosages of an anaesthetic. Upon admission to the recovery room and at minutes 5, 15 and 30 following admission, recovery scores were assigned on a categorical scale ranging from 1 (least favourable) to 7 (most favourable). Therefore one has four repetitions of a variable having 7 categories. Of course each individual may be assigned to one of the 2401 cells of the $7 \times 7 \times 7 \times 7$ contingency tables arising from the four measurements. However, since most of the cells will be empty it is of no use to try to model the total response . One approach is to model the marginal distribution of this contingency table. This *marginal approach* may be used for example if one wants to know how dosage determines the recovery level at differing times. It has been used by Stram, Wei & Ware (1988), Moulton & Zeger (1989), Liang, Zeger & Qaqish (1992), Stram & Wei (1988) and others.

The data structures considered here have the special feature that observed response categories are isotonic over time. In our example transitions from measurement to measurement take place only in the direction of higher recovery scores. In consecutive measurements patients never get a lower score in the measurement taken at a later point in time. This characterizes the essential structure of data that are considered in this paper. Scores taken at a later time have always to be equal or higher than in the previous measurement. Marginal modelling of this sort of data faces severe problems since the contingency tables with margins corresponding to measurements contain structural zeros. For example all the two dimensional tables have zeros in the lower (or upper) off-diagonal triangles. Straightforward marginal modelling which does not take care of these structured zeros will fail since the association structure is modelled inadequately. Fahrmeir, Gieger & Heumann (1999) gave on marginal approach which takes account of the structural zeros. In the present paper an alternative approach is proposed which focusses on the transition process. This approach, known as *transitional or conditional mod-*

elling does account not only for the dependence of marginal distributions on explanatory variables, but also involves the analysis of patterns of change, i.e. the conditional response given the previous one. For example, Bonney (1987) considered regressive logistic models which are of the transitional type. For the type of data considered here where observed categories of later measurements are higher than for earlier measurement the transitions take place only into the direction of higher scores. Thus the underlying process is irreversible. Transitional modelling aims directly at modelling this irreversible process. The objective is to model the effect of covariates on the first measurement and the following transitions between categories. It is of interest to know e.g. how dosage determines the recovery level at admission and the successive improvements following admission.

2 Isotonic modelling of irreversible processes

In the following let Y_t denote the ordered categorical measurement taken at time t with values from $\{1, \ldots, k\}$. Let $x' = (x_1, \ldots, x_m)$ denote a vector of covariates. Firstly, in the next section ordinal models for one response variable are sketched shortly.

2.1 Ordinal models for separate time points

Let us first consider the case of fixed time t, i.e. the case of cross-sectional data. Then the effect of x upon the response Y_t may be modelled by an ordinal regression model. Since fixed time is considered the index t is omitted. The most widespread type of model is the *threshold* or *cumulative model*

$$P(Y \leq r|x) = F(\theta_r + x'\beta) \tag{2.1}$$

where F is a distribution function, e.g. the logistic function $F(u) = 1/(1 + \exp(-u))$. For the logistic function one gets the proportional odds model

$$\log \frac{P(Y \leq r|x)}{P(Y > r|x)} = \theta_r + x'\beta,$$

for the extreme value distribution $F(u) = 1 - \exp(-\exp(u))$ one gets the so-called proportional hazard model. A motivation for the cumulative model

is the construction of a latent continuous variable $u = x'\beta + \varepsilon$ where ε has distribution function F. Then the observed value Y may be considered as a coarser version of the latent variable given by $Y = r$ if $\theta_{r-1} \leq u \leq \theta_r$ for thresholds $-\infty = \theta_0 < \theta_1 < \ldots < \theta_{k-1} < \theta_k = \infty$. For details and extensions see McCullagh (1980), Cox (1988), Brant (1990).

An alternative type of ordinal model is the *sequential model*

$$P(Y = r|Y \geq r, x) = F(\theta_r + x'\beta) \tag{2.2}$$

where F is again a distribution function. For the logistic distribution function (2.2) is the continuation ratio logit model

$$\log \frac{P(Y = r|x)}{P(Y > r|x)} = \theta_r + x'\beta, \tag{2.3}$$

(e.g. Agresti (2002)). In the special case where F is the extreme value distribution, the sequential model is equivalent to the cumulative model (2.1) (see Läärä & Matthews (1985), Tutz (1991)). To motivate the use of the sequential model for irreversible processes it is useful to derive the model as a process model. The reasoning behind (2.2) is that of a latent sequential or step-wise mechanism, for example the sequential recovering after treatment. The process starts in category 1 (lowest score). The first step is the transition from category 1 to category 2. This transition is determined by the dichotomous variable $Y^{(1)}$, more general

$$Y^{(r)} = \begin{cases} 1 & \text{process stops, no transition to category } r+1 \\ \\ 0 & \text{process continues, transition to category } r+1. \end{cases}$$

If the first (latent) step fails ($Y^{(1)} = 1$), the process stops and we have the observation $Y = 1$. If the first (latent) step succeeds ($Y^{(1)} = 0$), we will have an observed score $Y \geq 2$. The second step is the transition from category 2 to category 3 determined by $Y^{(2)}$. In general $Y^{(r)}$ determines the rth (conditional) step given the $r-1$ previous steps were successful. We observe $Y = r$ if $Y^{(1)} = \ldots = Y^{(r-1)} = 0, \quad Y^{(r)} = 1$, meaning that $r-1$ transitions to a higher score have been made but the rth step has failed. Thus model (2.2) may be seen as a simple model for the successive transitions of dichotomous variables $Y^{(1)}, Y^{(2)}, \ldots$.

The sequential model is particularly suited for response categories which can be arrived only step by step. In fact, in the recovery problem the stages

of recovery at each time point can only be reached successively. Even at admission the observed score is the result of a sequential process. Although the transitions are not observed directly, from the observations $Y = r$ it can be concluded which transitions actually were successful. In general cumulative and sequential models have different motivation, the former by considering the observed categories as a coarser version of a latent variable, the latter as a process model. However, it should be noted that in special cases the two types of models are equivalent. If F is chosen as the extreme value distribution the models can be shown to be the same (Läärä & Matthews (1985), Tutz (1991)). For various extensions of the models and comparisons between the two types of models see Armstrong & Sloan (1989), Greenland (1994), Barnhart & Sampson (1994), Agresti (1999) and Fahrmeir & Tutz (2001).

In the case of repeated measurements Stram, Wei & Ware (1988) used a cumulative type model separately for each time point and gave an estimate of the asymptotic covariance matrix by combining the estimates. In the same way the sequential model can be used separately for each time point. However, this marginal type of modelling seems not to be the best choice for data where $Y_t \geq Y_{t-1}$, because it does not account for the fact that higher levels are reached successively from measurement to measurement. Therefore, in the next section an alternative method is developed.

2.2 A simple model for irreversible processes

Let us now consider all the repeated measurements $Y_t,\ t = 1, \ldots, m$, rather than just one measurement at fixed time. For simplicity let Y_t be an ordered response variable having the same number of categories, i.e. $Y_t \in \{1, \ldots, k\},\ t = 1, \ldots, m$. The specific assumption made here is that the process $Y_1, Y_2, \ldots, Y_m$ is *irreversible* or *isotonic* in the sense that $Y_t \geq Y_{t-1},\ t = 1, \ldots, m$.

The total probability has the form

$$\begin{aligned} P(Y_1 = r_1, \ldots, Y_m = r_m) &= P(Y_1 = r_1)P(Y_2 = r_2|Y_1 = r_1)\cdots \\ &\quad P(Y_m = r_m|Y_1 = r_1, \ldots, Y_{m-1} = r_{m-1}) \quad (2.4) \\ &= \prod_{t=1}^{m} P(Y_t = r_t|(Y_1, \ldots, Y_{t-1}) = (r_1, \ldots, r_{t-1})) \end{aligned}$$

The representation (2.4) focuses on the process nature of the successive measurements instead of considering the marginal distribution of the responses

$Y_1, \ldots, Y_m$. Although it is not restricted to the case of irreversible processes it is easy to incorporate the restrictions arising from irreversibility. For simplicity here the Markov property $P(Y_t = r_t|Y_1, \ldots, Y_{t-1}) = P(Y_t = r_t|Y_{t-1})$ is assumed to hold.
When modelling the conditional response one option is to use the cumulative model. This approach faces the problem that the number of response categories depends on the previous outcome, so that

$$Y_t|\{Y_{t-1} = s\} \in \{s, s+1, \ldots, k\}. \tag{2.5}$$

Thus assuming a cumulative model for the conditional response implies that the number of response categories is varying. If the ordered categories are modelled by

$$P(Y_t \leq r|Y_{t-1} = s, x) = F(\theta_{tsr} + x'\beta),$$

$t = 1, \ldots, m,\ s = 1, \ldots, k-1,\ r = s, \ldots, k-1$, one has at time t $k-1$ different cumulative models, namely one for each starting category $s = 1, \ldots, k-1$. Moreover, the $k-1$ sets of thresholds have to fulfill $\theta_{tsr} < \ldots < \theta_{ts,k-1},\ s = 1, \ldots, k-1$.

The sequential model circumvents the problem of varying response categories since it is local in the sense that it models the local transition to higher categories given a category is reached. So in the following we will consider the *isotonic sequential model for repeated measurements*

$$P(Y_t = r|Y_t \geq r, Y_{t-1} = s, x) = F(\theta_{tsr} + x'\beta), \tag{2.6}$$

$t = 1, \ldots, m,\ s = 1, \ldots, k-1,\ r = s, \ldots, k-1$.

In model (2.6) the response of the first measurement is modelled as a simple sequential model. For the consecutive measurements the same step-wise process is assumed but now starting from the category that has been reached in the previous measurement. The restriction of irreversibility is simply incorporated by reducing the possible categories in (2.6) to $r = s, \ldots, k-1$. Another advantage of model (2.6) is that it reduces to a dichotomous response model and therefore for estimation any program package that is able to handle binary regression models like SAS, S-PLUS or GLIM may be used. Estimation and goodness of fit are considered in the next section.

2.3 Estimation

Let for n observations x_i be the explanatory variable and $Y_{it}|x_i \in \{1, \ldots, k\}$ be the response variable. The multivariate nature of the response variable Y_{it} becomes obvious by using dummy variables

$$y_{itr} = \begin{cases} 1 & , \quad \text{if } Y_{it} = r \\ 0 & , \quad \text{otherwise.} \end{cases}$$

Instead of Y_{it} now the vector $(y_{it1}, \ldots, y_{itq})$, $q = k-1$, represents the response. For convenience let the observation at 'time 0' be defined by $Y_{i0} = 1$ (i.e. $y_{i01} = 1$ and $y_{i0s} = 0$ for $s > 1$). When modelling the responses $Y_{i1}, \ldots, Y_{im}$ given x_i the kernel of the likelihood for observation i has the form

$$\begin{aligned} L_i &= P(Y_{i1} = y_{i1}, \ldots, Y_{im} = y_{im}|x_i) \\ &= \prod_{t=1}^{m} P(Y_{it} = y_{it}|Y_{i,t-1} = y_{i,t-1}, x_i), \end{aligned}$$

where (2.4) is used and in the conditioning on previous responses the Markov property is assumed to hold. Since $Y_{it}|Y_{i,t-1} = s$ takes values $s, s+1, \ldots, k$, the essential term may be written as

$$\begin{aligned} P(Y_{it} = r|Y_{i,t-1} = s, x_i) &= P(Y_{it} = r|Y_{it} \geq r, Y_{i,t-1} = s, x_i) \\ &\quad \cdot \prod_{j=s}^{r-1} (1 - P(Y_{it} = j|Y_{i,t-1} \geq j, Y_{i,t-1} = s, x_i)). \end{aligned} \tag{2.7}$$

The right side of (2.7) is a product of the probabilities for the transition from s to $s+1$, $s+1$ to $s+2$ etc. and the probability for the failing transition from r to $r+1$. Thus, by using (2.7) the likelihood L_i has the form

$$\begin{aligned} L_i &= \prod_{t=1}^{m} P(Y_{it} = y_{it}|Y_{it} \geq y_{it}, Y_{i,t-1} = y_{i,t-1}, x_i) \\ &\quad \cdot \prod_{r=y_{i,t-1}}^{y_{it}-1} (1 - P(Y_{it} = r|Y_{it} \geq r, Y_{i,t-1} = y_{i,t-1}, x_i)) \\ &= \prod_{t=1}^{m} \prod_{r=y_{i,t-1}}^{y_{it}} P(Y_{it} = r|Y_{it} \geq r, Y_{i,t-1} = y_{i,t-1}, x_i)^{y_{itr}} \\ &\quad \cdot (1 - P(Y_{it} = r|Y_{it} \geq r, Y_{i,t-1} = y_{i,t-1}, x_i))^{1-y_{itr}} \end{aligned}$$

Since model (2.6) has the form

$$\begin{aligned} P(y_{itr} = 1 | Y_{it} \geq r, Y_{i,t-1} = s, x_i) &= P(Y_{it} = r | Y_{it} \geq r, Y_{i,t-1} = s, x_i) \\ &= F(\theta_{tsr} + x_i'\beta) \end{aligned}$$

it is obvious that L_i is the likelihood of a dichotomous model for the observations y_{itr} where $y_{itr} = 0$ stands for successful steps to higher categories and $y_{itr} = 1$ stands for failing of this step (considered for fixed time t). However, for one observation i the likelihood L_i contains several dichotomous responses according to the steps involved. This data augmentation is described in the following.

For simplicity let the responses be denoted by $Y_{it} = r_t$. Thus the response categories of the m measurements are given by $r_0 = 1 \leq r_1 \leq \ldots \leq r_m$. The measurement r_t at time t enters the likelihood L_i in the form of the dichotomous responses $y_{itr_{t-1}}, \ldots, y_{itr_t}$. The response $y_{itr_{t-1}} = 0$ denotes start in category r_{t-1} (from the previous measurement) and transition to the adjacent category $r_{t-1} + 1$. Transitions to the adjacent categories take place until category r_t is reached. That means $y_{it,r_t-1} = 0$ for the transition from $r_t - 1$ to r_t and $y_{itr_t} = 1$ for the failing of transition to category $r_t + 1$. Thus the binary transitions for individual i are given by $y_{i11}, \ldots, y_{i1r_1}, y_{i2r_1}, \ldots, y_{i2r_2}, \ldots, y_{imr_m}$. Table 2.1 shows the dichotomous responses and the corresponding predictors contributing to the likelihood L_i.

For illustration let us consider two patients (patient 1 and 11) from the anaesthesia recovery example:

Patient	r_1	r_2	r_3	r_4
1	4	6	7	7
11	1	3	3	5

The corresponding dichotomous responses y_{itr} for patient 1 are given by

$t = 1$	$y_{111} = 0$	$y_{112} = 0$	$y_{113} = 0$	$y_{114} = 1$
$t = 2$	$y_{124} = 0$	$y_{125} = 0$	$y_{126} = 1$	
$t = 3$	$y_{136} = 0$			

For patient 11 one gets

$$
\begin{array}{lllll}
t=1 & y_{11,1,1}=1 & & & \\
t=2 & y_{11,2,1}=0 & y_{11,2,2}=0 & y_{11,2,3}=1 & \\
t=3 & y_{11,3,3}=1 & & & \\
t=4 & y_{11,4,3}=0 & y_{11,4,4}=0 & y_{11,4,5}=1. &
\end{array}
$$

It should be noted that only six of the dummy variables may have response 0, because then the last category 7 is already reached and no further transition can be considered. Thus for patient 1 there is no contribution at time $t = 4$. Since $y_{136} = 0$ denotes that the transition to the final category 7 has already been performed at $t = 3$ no further transition is possible. The embedding of the model into models for dichotomous responses makes it easy to use programmes like SAS, or S-PLUS. One just has to transform the ordered responses into dichotomous responses as given in Table 2.1. To illustrate this data transformation let us consider again patient 1 and 11. The transformed data have the following form for $TIME$, CAT (category under consideration), $START$ (starting category from previous measurement) and $RESP$ (response):

$PATIENT$	$TIME$	CAT	$START$	$RESP$	covariates
1	1	1	1	0	x_1
1	1	2	1	0	x_1
1	1	3	1	0	x_1
1	1	4	1	1	x_1
1	2	4	4	0	x_1
1	2	5	4	0	x_1
1	2	6	4	1	x_1
1	3	6	6	1	x_1
11	1	1	1	1	x_{11}
11	2	1	1	0	x_{11}
11	2	2	1	0	x_{11}
11	2	3	1	0	x_{11}
11	3	3	3	1	x_{11}
11	4	3	3	0	x_{11}
11	4	4	3	0	x_{11}
11	4	5	3	1	x_{11}

The response variable to be modelled is $RESP$, a binary variable. The sequential model for repeated ordinal measurements can be easily fitted using

standard procedures for fitting dichotomous responses after the data has been transformed as shown above (e.g. PROC LOGISTIC in SAS or LOGISTIC REGRESSION in SPSS/PC+).

3 Application to recovery scores

In addition to the response $Y_{it} \in \{1, \ldots, 7\}$ at four time points, covariates were observed, namely $DOSE$ (4 categories: 15 mg/kg, 20 mg/kg, 25 mg/kg, 30 mg/kg), AGE (9 – 70 months) and DUR (35 – 190 minutes). For the data see Davis (1991). Since $DOSE$ is given in four levels it is considered as a categorical covariate with four categories.

3.1 Models for fixed time

In his analysis Davis (1991) used the method of Stram, Wei & Ware (1988). He found little evidence of significant effects due to $DOSE$, AGE or DUR of surgery. However, this result may be due to omitted interaction effects. We fitted various models including interactions for fixed time. Table 3.1 shows the results for three models fitted for the first response (at admission to the surgery room). The given models are the model with two-factor interactions $DOSE.AGE$ and $DOSE.DUR$, a main effect model with $DOSE.AGE$ interaction and a simple main effect model. The variable AGE is centered at 37 month which is approximately the mean in the sample. Throughout the paper the effects of $DOSE$ are given in dummy coding, i.e. $DOSE[i] = 1$ if $DOSE$ is in category i and $DOSE[i] = 0$ if $DOSE$ is not in category i, p-values correspond to Wald test.

From the estimates of the main effect model (the model used by Davis) indeed there is little evidence of significant effects. However, if the $DOSE.AGE$ interaction is included, not only the interaction is significant but also the main effects are obviously not to be neglected. This may be the effect of omitted variables which leads to estimates strongly biased towards zero (see e.g. Cramer (1991)). For example the estimate for $DOSE[1]$ goes down from -1.487 to -1.986 if the interaction $DOSE.AGE$ is omitted. Moreover, it is seen that the coefficient of AGE is negative for the models with interaction but positive for the third model without interaction. This is due to averaging if the

interaction is omitted. In the second model in Table 3.1 AGE has coefficients 0.048, 0.100, 0.014, -0.044 for $DOSE$ categories 1, 2, 3, 4 (Coefficient for $DOSE[i].AGE$ + coefficient for AGE). If the interaction between $DOSE$ and AGE is omitted the coefficient for AGE as an average is positive. Considering the deviances in Table 3.1. The DOSE.DUR interaction may be omitted but not the DOSE.AGE interaction. Table 3.2 shows the corresponding sequential models for the first measurement.

Table 3.2 about here

The results are quite similar: The main effect model shows no significance but inclusion of interactions shows significant effects. The fit of the sequential models is slightly better than for the cumulative model. Again an economic model is the second with $DOSE$, AGE, DUR and $DOSE.AGE$. The coefficients of AGE for the four $DOSE$ levels are 0.080, 0.120, 0.074, -0.036. Thus increasing AGE slightly defers transition to higher categories for $DOSE$ levels 1, 2, 3, but prompts transition for $DOSE$ level 4. From the main effects of $DOSE$ (-1.159, -0.311, -0.529, 0.0) it is seen that only the first $DOSE$ level has a significantly different effect from $DOSE$ level 4, the first one makes transition to higher categories easier. The DUR effect is weak (p-value: 0.077) with a tendency to prevent transitions with increasing DUR.

3.2 Modelling repeated measurements

Table 3.3 about here

A coarse description of the data is given in Table 3.3. Since transitions are of interest, the frequencies of transitions from t to $t+1$ for fixed time are given. From the marginal distribution it is seen that some categories have very low frequencies for all time points. There are few observations in categories 3, 5 and 6, the preferred response categories seem to be 2, 4 and 7. As is to be expected the frequencies of high response categories are increasing with time. (In the data given in Davis (1991)) of the 180 transitions from t to $t+1$ there is *one* transition to a lower category, namely from category 2 at $TIME = 2$ to category 1 at $TIME = 3$. That assumed error is corrected by using the response category 2 at $TIME = 3$).

A simultaneous analysis of the four measurements is based on the isotonic

sequential logit model given in (2.6).

$$P(Y_t = r|Y_t \geq r,\ x) = \exp\left(z'_{tsr}\beta\right)/\left(1 + \exp(z'_{tsr}\beta)\right) \tag{3.1}$$

where z_{tsr} is a design vector that depends on time t, on starting value s from which the sequential mechanism starts at time t, on category r for which the transition to higher categories is considered and of course on the explanatory variables. The former variables are referred to as $TIME$ $(t = 1, \ldots, 4)$, $START$ $(s = 1, \ldots, 6)$ and CAT (for category, $r = 1, \ldots, 7$).The covariates and interactions included are $TIME$,$DOSE$,AGE,DUR,$DOSE.AGE$. Where TIME and DOSE are treated as factors and AGE is a metric variable centered around 37.

When specifying z_{tsr} it is important to consider the specific data situation. First, one has to distinguish between responses at $TIME = 1$ which are determined only by covariates with $START = 1$ for all observations and responses at $TIME = 2$, 3, 4 for which the starting point is set in the previous measurement and therefore has values $START \in \{1, \ldots, 6\}$. Therefore we will develop the linear predictor in two steps.

Linear predictor for first measurement :

For $TIME = 1$ a model with explanatory terms $DOSE$, AGE, DUR and the interaction $DOSE.AGE$ has already been considered in Section 3.2. The sequential logit model (estimates given in Table 3.2)

$$\log\left(P(Y = r)/P(Y > r)\right) = \theta_r + x'\beta$$

where x stands for $DOSE$, AGE, DUR and $DOSE.AGE$ may also be given in the form

$$\log\left(P(Y = r)/P(Y > r)\right) = z'_r\beta$$

where z_r stands for CAT, $DOSE$, AGE, DUR and $DOSE.AGE$. In the latter form the constants $\theta_1, \ldots, \theta_6$ are replaced by the 'explanatory' variable CAT or more precisely by dummy variables $CAT[1], \ldots, CAT[6]$, where $CAT[r] = 1$ if $Y = r$ and $CAT[r] = 0$ else. The estimated 'category effects' for the sequential model with CAT, $DOSE$, AGE, DUR, $DOSE.AGE$ are $CAT[1] = -2.058$, $CAT[2] = -0.120$, $CAT[3] = -0.684$, $CAT[4] = 0.344$, $CAT[5] = -0.152$, $CAT[6] = -0.080$. These effects (or thresholds) represent the ease of transition

from category r to $r+1$: for large value the response is bound to stay in category r, low values signal easy transition to $r+1$. The lowest value is found for $CAT = 1$, the highest value is found for $CAT = 4$. The same effect is found for $T = 1$ in Table 3.3 where the continuation ratios are given. Continuation ratios give the proportion of observations in category r and observations in category r or higher, that means the relative frequencies in relation to the number of individuals 'under risk'.

Linear predictor for later measurements :

Since the starting category at time $T = 1$ is always the first one, $START$ has no influence in this case. This is different for $T > 1$. When modelling transitions for $T > 1$ one has to take the starting category into account. The dependence of category-specific transitions and starting category is modelled as interaction between the variables CAT and $START$. However, a simple interaction term $START.CAT$ will not work since structural zeros are implied. If $START = s$ the response $RESP$ can only take values $s, s+1, \ldots, k$. Therefore the interaction $START.CAT$ is built from the products of dummy variables

$$START[s].CAT[r], \qquad s = 1, \ldots, 6,\ r = s, \ldots, 6$$

where $START[s] = 1$ if $START = s$, $START[s] = 0$ if $START \neq s$. Assuming this reduced interaction one considers for $T > 1$ the predictor. $START.CAT, DOSE.AGE, DUR.DOSE.AGE, DOSE.DUR$. The total linear predictor in model (3.1) distinguish between $T = 1$ and $T > 0$ by use of the dummy variable $T1$ with $T1 = 1$ if $T = 1$, $T1 = 0$ if $T > 1$. The explanatory term in (3.1) is specified by a nested design where effects are considered separately for $T = 1$ and $T > 1$. The variables included are

$TIME$,
$(CAT, DOSE, AGE, DUR, DOSE.AGE).T1$,
$(START.CAT, DOSE, AGE, DUR, DOSE.AGE, DOSE.DUR).(1 - T1)$.

The first factor $TIME$ reflects the effect of the time of measurement. The second factor specifies the effects of covariates at $TIME = 1$. The third factor specifies the effects of covariates at $TIME > 1$ where the effect of starting category is taken into account. Table 3.4 gives deviances for models which are

derived by successively omitting terms. 'Full model' refers to model (3.1) with the explanatory term specified above.

Table 3.4 about here

As is seen from Table 3.4 the explanatory variables $DOSE$, AGE, DUR as well as the interactions $DOSE.DUR$, $DOSE.AGE$ may be omitted in the case $T > 1$. However, for $TIME = 1DOSE.AGE.T1$ may not be omitted and thus is kept in the model. The omission of time yields an increase in deviance of 16.5 on 3 DF. Thus time is a highly influential variable and cannot be omitted. This also holds for the (interaction) effect of starting values and category. Thus it seems that $DOSE$, AGE, DUR have an effect on the first response (immediately after surgery). But the further transitions, although depending on the level set on the first response, are not determined by the covariates. The estimates of variables for the reduced model are given in Table 3.5.

Table 3.5

It is seen from Table 3.5 that the effects for $T = 1$ are unchanged in comparison with the sequential model considered in Section 3.2 (second model in Table 3.2). This is due to the nested effects modelling. Interpretation of effects is the same as in Section 3.2. For $T > 1$ only the $START.CAT$-interaction remains in the model. They are considered as nuisance parameters which account for starting effects and are not given.

4 Incorporation of random effects

The isotonic sequential model for repeated measurements (2.6) utilizes the conditioning on covariates and responses on previous measurements. But it ignores possible heterogeneity of individuals and therefore possible correlation between the binary transitions which is due to this heterogeneity. Explicit modelling of heterogeneous reactions of individuals may be obtained by the incorporation of random effects. In the simplest case one has a random intercept b_i yielding the *random effects isotonic sequential model*

$$P(Y_{it} = r | Y_{it} \geq r, Y_{i,t-1} = s, x_i, b_i) = F(\theta_{tsr} + x_i\beta + b_i) \qquad (4.1)$$

where for simplicity the random effect is assumed to be normally distributed, i.e. $b_i \sim N(O, \sigma^2)$. In (4.1) the conditioning explicitly implies the random

effects b_i. Estimation of the structured parameters $\theta_{tsr}, \beta, \sigma^2$ is based on the marginal likelihood. Using the representation as binary transitions derived in Section 2.3 the marginal log-likelihood has the form

$$l(\{\theta_{tsr}\}, \beta, \sigma^2) = \sum_{i=1}^{n} log \int f(y_i|b_i)p(b_i; \sigma^2)db_i \tag{4.2}$$

where $p(b_i; \sigma^2)$ denotes the density of the random effects b_i and the binary transitions of individual i are collected in

$$f(y_i|b_i) = \prod_{t=1}^{m} \prod_{r=y_{i,t-1}}^{y_{it}} F(\eta_{itr})^{y_{itr}} \ (1 - F(\eta_{itr}))^{1-y_{itr}}$$

where

$$\eta_{itr} = \theta_{t,y_{i,t-1},r} + x_i'\beta + b_i$$

Since the observations are given as binary variables the framework of univariate generalized linear mixed models (GLMMs) may be used. The main problem in GLMMs is that the marginal distribution of the response obtained by integrating out the random effects, does not have closed from. This led to the development of several methods to obtain analytical approximation for the likelihood, like numerical integration based on Gauss-Hermite quadrature (e.g. Hinde (1982), Anderson & Aitkin (1985)) or Monte Carlo techniques within the EM-algorithm (McCulloch (1994), McCulloch (1997), Booth & Hobert (1999)) or approximation methods as Taylor expansions or Laplace approximation (e.g. Breslow & Clayton (1993), Wolfinger & O'Connell (1993), Longford (1994)). A more recent approach is nonparametric maximum likelihood which avoids the assumption of a fixed distribution for the random effects (Aitkin (1996), Aitkin (1999)). In the following evaluation of the integral in (4.2) is based on Gauss-Hermite quadrature within the EM-algorithm and the approach of Wolfinger & O'Connell (1993) which is implement in the SAS macro GLIMMIX. For Gauss-Hermite quadrature let the unknown parameters be collected in α. Then in the E-step of the $(p+1)th$ EM cycle one has to determine

$$M(\alpha|\alpha^{(p)}) = \sum_{i=1}^{n} k_i^{-1} \int [log\ f(y_i|b_i; \alpha) + log\ g(a_i)]\ f(y_i|b_i, \alpha^{(p)})g(b_i)db_i$$

where

$$k_i = \int f(y_i|b_i, \alpha^{(p)})g(b_i)db_i\alpha^{(p)}$$

is the estimate from the previous cycle. Gauss-Hermite quadrature yields the approximation

$$M^{GH}(\alpha|\alpha^{(p)}) = \sum_{i=1}^{n}\sum_{j=1}^{M} C_{ij}^{GH}[log\ f(y_i|d_j;\alpha) + log\ g(d_j)$$

where d_j are the quadrature points and C_{ij}^{GH} are weight factors which contain the quadrature weights. In the M-step which consists of maximizing $M(\alpha|\alpha^{(p)})$ the framework of generalized linear models may be used in the form of Fisher scoring or iterative weighted least-squares. For details see Chapter 7 in Fahrmeir & Tutz (2001). For the estimation of the random effects model a S-PLUS procedure has been written which computes estimates for Gauss-Hermite quadrature. In addition the SAS Macro GLIMMIX has been used. The model considered has the same covariate structure as the model given in Table 3.5 but with the inclusion of a random intercept. Moreover, the interaction $START.CAT.(1-T1)$ which contains nuisance parameters has been omitted. When these nuisance parameters are included the estimation of the random effects model is very unstable yielding quite different estimates for Gauss-Hermite quadrature and GLIMMIX. It is seen from Table 3.5 that the time effects have been shifted by the inclusion of a random intercept whereas the other parameter estimates remain about the same. This may be due to the low heterogeneity $\hat{\sigma} = 0.634$ for Gauss-Hermite and $\hat{\sigma} = 0.550$ in GLIMMIX. Again the interaction between AGE and $DOSE$ for $TIME > 1$ should not be neglected. The estimates resulting from Gauss-Hermite quadrature and GLIMMIX are well comparable with a slightly stronger heterogeneity resulting from Gauss-Hermite quadrature.

5 Concluding remarks

The model presented here is not restricted to Markov type models. However, modelling the dependence of y_t on all the previous outcomes increases strongly the number of parameters to be estimated. The basic advantage of the model is that the ordinal structure of the data as well as the irreversibility of the process is explicitly used. The ordinal structure is exploited by using a sequential type model for the responses at time t, the irreversibility is accounted for by considering the reduced response categories where the minimal score of the

present response is set by the outcome of the previous time point measurement. The analysis is not restricted to effects of covariates on marginal responses. It allows to analyze the effects on the conditional transitions following the first response.

An advantage of the model is that it can be embedded into models for dichotomous responses. Therefore, it can be easily fitted after suitable data transformation by using standard procedures for dichotomous responses. The incorporation of heterogeneity uses the same data transformations. With the increasing availability of programmes which handle mixed models with dichotomous responses it may be easily applied to data. Although the model is developed within a parametric framework the extension to nonparametric modelling where the linear predictor is replaced by additive or varying coefficients effects is straightforward. Semi- and nonparametric models for dichotomous models are discussed e.g. in Hastie & Tibshirani (1990).

Acknowledgement:
Support from SFB funded by Deutsche Forschungsgemeinschaft is grateful acknowledged. Thanks to L. Heigenhauser and F. Reithinger for their support with computations.

References

Agresti, A. (1999). Modelling ordered categorical data: Recent advances and future challenges. *Statistics in Medicine* **18**, 2191–2207.

Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* **6**, 251–262.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 117–128.

Anderson, D. A. and Aitkin, M. (1985). Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society* **Ser. B 47**, 203–210.

Armstrong, B. and Sloan, M. (1989). Ordinal regression models for epidemiologic data. *American Journal of Epidemiology* **129**, 191–204.

Barnhart, H. and Sampson, A. (1994). Overviews of multinomial models for ordinal data. *Comm. Stat-Theory & Methods* **23(12)**, 3395–3416.

Bonney, G. (1987). Logistic regression for dependent binary observations. *Biometrics* **43**, 951–973.

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Statist. Soc* **B 61**, 265–285.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* **46**, 1171–1178.

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

Cox, C. (1988). Multinomial regression models based on continuation ratios. *Statistics in Medicine* **7**, 433–441.

Cramer, J. S. (1991). *The Logit Model*. New York: Routhedge, Chapman & Hall.

Davis, C. S. (1991). Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. *Statistics in Medicine* **10**, 1959–1980.

Fahrmeir, L., Gieger, C., and Heumann, C. (1999). An application of isotonic longitudinal marginal regression to monitoring the healing process. *Biometrics* **55**, 951–956.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models* (2nd ed.). New York: Springer.

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

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

Hinde, J. (1982). Compound poisson regression models. In R. Gilchrist (Ed.), *GLIM 1982 Internat. Conf. Generalized Linear Models*, New York, pp. 109–121. Springer-Verlag.

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

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analysis for categorical data (with discussion). *Journal of the Royal Statistical Society* **B 54**, 3–40.

Longford, N. T. (1994). Random coefficient models. In G. Arminger, C. Glogg, & M. Sobel (Eds.), *Handbook of Statistical Modelling for the Behavioural Sciences*. New York: Plenum.

McCullagh, P. (1980). Regression model for ordinal data (with discussion). *Journal of the Royal Statistical Society* **B 42**, 109–127.

McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data. *J. Am. Statist. Assoc.* **89**, 330–335.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.

Moulton, L. and Zeger, S. (1989). Analysing repeated measures in generalized linear models via the bootstrap. *Biometrics* **45**, 381–394.

Stram, D. O. and Wei, L. J. (1988). Analyzing repeated measurements with possibly missing observations by modelling marginal distributions. *Statistics in Medicine* **7**, 139–148.

Stram, D. O., Wei, L. J., and Ware, J. H. (1988). Analysis of repeated categorical outcomes with possibly missing observations and time-dependent covariates. *Journal of the American Statistical Association* **83**, 631–637.

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

Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models; a pseudolikelihood approach. *Journal Statist. Comput. Simulation* **48**, 233–243.

	Dichotomous response			Predictor	
	y_{itr}			θ_{tsr}	x_i
$t=1$	y_{i11}	$=$	0	θ_{111}	x_i
	$\vdots$			$\vdots$	
	y_{i1,r_1-1}	$=$	0	θ_{11,r_1-1}	x_i
	y_{i1,r_1}	$=$	1	θ_{11,r_1}	x_i
$t=2$	y_{i2r_1}	$=$	0	$\theta_{2r_1r_1}$	x_i
	$\vdots$			$\vdots$	
	y_{i2,r_2-1}	$=$	0	θ_{2r_1,r_2-1}	x_i
	y_{i2,r_2}	$=$	1	θ_{2r_1,r_2}	x_i
	$\vdots$			$\vdots$	
$t=m$	$y_{im,r_{m-1}}$	$=$	0	$\theta_{mr_{m-1}r_{m-1}}$	x_i
	$\vdots$			$\vdots$	
	y_{im,r_m-1}	$=$	0	θ_{mr_{m-1},r_m-1}	x_i
	y_{im,r_m}	$=$	1	θ_{mr_{m-1},r_m}	x_i

Table 2.1: Dichotomous observations y_{itr} entering the likelihood contribution L_i for response categories $1 \leq r_1 \leq \ldots \leq r_m$

	Estimate	p-value	Estimate	p-value	Estimate	p-value
$DOSE[1]$	-3.604	1.873	-1.519	0.727	-1.009	0.680
$DOSE[2]$	-1.670	1.927	-0.513	0.758	-0.904	0.701
$DOSE[3]$	-2.995	1.935	-0.607	0.702	-0.375	0.676
AGE	-0.042	0.024	-0.044	0.024	0.011	0.015
DUR	-0.009	0.020	0.013	0.007	0.011	0.006
$DOSE[1].AGE$	0.088	0.048	0.091	0.047		
$DOSE[2].AGE$	0.143	0.047	0.146	0.046		
$DOSE[3].AGE$	0.053	0.048	0.058	0.048		
$DOSE[1].DUR$	0.028	0.023				
$DOSE[2].DUR$	0.016	0.024				
$DOSE[3].DUR$	0.032	0.024				
Deviance	178.99		181.23		193.27	
df	343		346		349	
Deviance			$DOSE.DUR$		$DOSE.AGE$	
for effects			2.24		12.04	
df			3		3	

Table 3.1: Estimates for three cumulative logistic models with varying covariate effects included ($T = 1$)

	Estimate	std error	Estimate	std error	Estimate	std error
$DOSE[1]$	-2.467	1.592	-1.159	0.562	-0.600	0.504
$DOSE[2]$	0.013	1.667	-0.311	0.616	-0.596	0.581
$DOSE[3]$	-1.928	1.662	-0.529	0.544	-0.121	0.521
AGE	-0.036	0.021	-0.036	0.021	0.016	0.013
DUR	-0.000	0.018	0.010	0.005	0.007	0.005
$DOSE[1].AGE$	0.081	0.039	0.080	0.039		
$DOSE[2].AGE$	0.116	0.039	0.120	0.039		
$DOSE[3].AGE$	0.074	0.039	0.074	0.039		
$DOSE[1].DUR$	0.017	0.020				
$DOSE[2].DUR$	-0.002	0.021				
$DOSE[3].DUR$	0.019	0.022				
Deviance	177.92		180.81		192.90	
df	343		346		349	
Deviance			$DOSE.DUR$		$DOSE.AGE$	
for effects	$T = 1$		2.89		12.09	
	$T = 2$		1.226		5.767	
	$T = 3$		4.775		1.994	
df			3		3	

Table 3.2: Estimates for sequential logistic models with varying covariate effects included ($T = 1$)

		Response at T=2								Continuation ratios
		1	2	3	4	5	6	7	Σ	T=1
	1	5	1	2	1	0	1	0	10	0.16 (10/60)
	2	0	18	2	3	1	0	0	24	0.48 (24/50)
Response	3	0	0	2	5	1	0	0	8	0.30 (8/26)
at	4	0	0	0	2	2	3	2	9	0.50 (9/18)
T=1	5	0	0	0	0	1	1	1	3	0.33 (3/9)
	6	0	0	0	0	0	0	2	2	0.33 (2/6)
	7	0	0	0	0	0	0	4	4	1.00 (4/4)

		Response at T=3								Continuation ratios
		1	2	3	4	5	6	7	Σ	T=2
	1	4	0	0	0	1	0	0	5	0.08 (5/60)
	2	0	12	2	0	3	2	0	19	0.34 (19/55)
Response	3	0	0	3	1	0	0	2	6	0.16 (6/36)
at	4	0	0	0	5	1	3	2	11	0.36 (11/30)
T=2	5	0	0	0	0	1	1	3	5	0.26 (5/19)
	6	0	0	0	0	0	3	2	5	0.35 (5/14)
	7	0	0	0	0	0	0	9	9	1.00 (9/9)

		Response at T=4								Continuation ratios
		1	2	3	4	5	6	7	Σ	T=3
	1	2	0	0	2	0	0	0	4	0.06 (4/60)
	2	0	2	0	3	4	0	3	12	0.21 (12/56)
Response	3	0	0	1	2	1	1	0	5	0.11 (5/44)
at	4	0	0	0	0	0	4	2	6	0.15 (6/39)
T=3	5	0	0	0	0	1	0	5	6	0.18 (6/33)
	6	0	0	0	0	0	3	6	9	0.33 (9/27)
	7	0	0	0	0	0	0	18	18	1.00 (18/18)
Continuation		2	2	1	7	6	8	34	60	
ratios		0.03	0.03	0.02	0.12	0.12	0.19	1.00		
T=4		(2/60)	(2/58)	(1/56)	(7/55)	(6/48)	(8/42)	(34/34)		

Table 3.3: Frequencies of transitions and marginal distributions

Successively omitted effect	Deviance	df	Deviance of effects	df
Full Model	513.58	425		
$DOSE.DUR.(1-T)$	516.68	428	3.10	3
$DOSE.AGE.(1-T)$	520.75	431	4.07	3
$DUR.(1-T)$	522.58	432	1.83	1
$AGE.(1-T)$	524.88	433	2.29	1
$DOSE.(1-T)$	526.28	436	1.40	3
$TIME.(1-T)$	538.36	439	12.08	3

Table 3.4: Sequential models for irreversible processes including CAT, $DOSE$, AGE, DUR, $DOSE.AGE$ at $TIME = 1$, $START.CAT$, $DOSE$, AGE, DUR, $DOSE.AGE$, $DOSE.DUR$ at $TIME > 1$ and $TIME$.

Parameter	Estimate	std error	p-value
$T[2]$	0.403	0.563	0.473
$T[3]$	0.071	0.548	0.896
$T[4]$	−0.893	0.534	0.094
$CAT[1].T1$	−2.058	0.669	0.002
$CAT[2].T1$	−0.120	0.607	0.842
$CAT[3].T1$	−0.684	0.666	0.304
$CAT[4].T1$	0.346	0.721	0.631
$CAT[5].T1$	−0.152	0.898	0.865
$CAT[5].T1$	−0.080	1.070	0.940
$DOSE[1].T1$	−1.159	0.562	0.039
$DOSE[2].T1$	−0.311	0.616	0.613
$DOSE[3].T1$	−0.529	0.545	0.331
$AGE.T1$	−0.036	0.021	0.088
$DUR.T1$	0.010	0.005	0.077
$DOSE[1].AGE.T1$	0.080	0.039	0.039
$DOSE[2].AGE.T1$	0.120	0.039	0.001
$DOSE[3].AGE.T1$	0.074	0.039	0.056

Table 3.5: Estimates for the isotonic sequential model

Parameter	Gauss-Hermite	GLIMMIX	p-value GLIMMIX
$T[2]$	-0.031	0.045	0.836
$T[3]$	-0.249	-0.243	0.267
$T[4]$	-1.249	-1.297	0.000
$CAT[1].T1$	-2.089	-2.098	0.002
$CAT[2].T1$	-0.086	-0.042	0.946
$CAT[3].T1$	-0.601	-0.451	0.508
$CAT[4].T1$	0.485	0.686	0.351
$CAT[5].T1$	-0.004	0.251	0.780
$CAT[5].T1$	0.081	0.367	0.730
$DOSE[1].T1$	-1.198	-1.204	0.042
$DOSE[2].T1$	-0.336	-0.354	0.580
$DOSE[3].T1$	-0.527	-0.482	0.402
$AGE.T1$	-0.037	-0.038	0.073
$DUR.T1$	-0.010	0.010	0.091
$DOSE[1].AGE.T1$	0.081	0.080	0.047
$DOSE[2].AGE.T1$	0.124	0.124	0.002
$DOSE[3].AGE.T1$	0.072	0.064	0.118
$\hat{\sigma}$	0.634	0.550	

Table 3.6: Estimates for random effects isotonic sequential model