Log-linear Rasch-type Models for Repeated Categorical Data with a Psychological Application

Hatzinger, Reinhold; Katzenbeisser, Walter

DOI:
10.57938/f25b858d-41c2-45ac-879b-ca795393702f

Published: 01/07/2008

Document Version
Publisher's PDF, also known as Version of record

Log-linear Rasch-type Models for Repeated Categorical Data with a Psychobiological Application

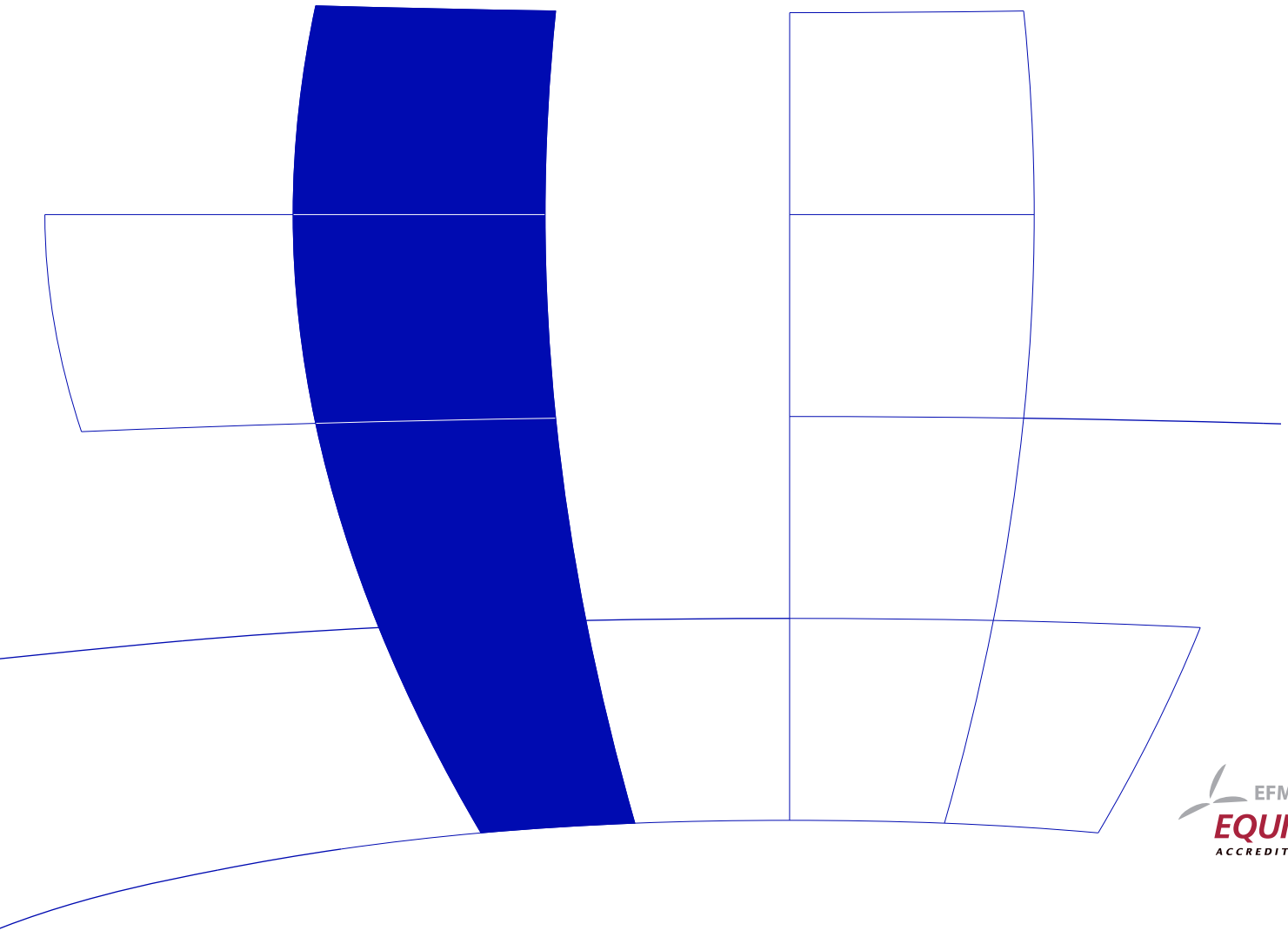
Reinhold Hatzinger, Walter Katzenbeisser

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 69
July 2008

http://statmath.wu-wien.ac.at/



Log-linear Rasch-type Models for Repeated Categorical Data with a Psychobiological Application

Reinhold Hatzinger Walter Katzenbeisser

Abstract

The purpose of this paper is to generalize regression models for repeated categorical data based on maximizing a conditional likelihood. Some existing methods, such as those proposed by Duncan (1985), Fischer (1989), and Agresti (1993, and 1997) are special cases of this latent variable approach, used to account for dependencies in clustered observations. The generalization concerns the incorporation of rather general data structures such as subject-specific time-dependent covariates, a variable number of observations per subject and time periods of arbitrary length in order to evaluate treatment effects on a categorical response variable via a linear parameterization. The response may be polytomous, ordinal or dichotomous. The main tool is the log-linear representation of appropriately parameterized Rasch-type models, which can be fitted using standard software, e.g., R. The proposed method is applied to data from a psychiatric study on the evaluation of psychobiological variables in the therapy of depression. The effects of plasma levels of the antidepressant drug Clomipramine and neuroendocrinological variables on the presence or absence of anxiety symptoms in 45 female patients are analyzed. The individual measurements of the time dependent variables were recorded on 2 to 11 occasions. The findings show that certain combinations of the variables investigated are favorable for the treatment outcome.

Keywords. Latent variables, Rasch model, time-dependent covariates, conditional maximum likelihood, log-linear models, quasi-symmetry, R.

1 Introduction

In this article we describe the analysis of data from a psychobiological study where neuroendocrine and pharmacological effects on anxiety symptoms in the therapy of depression are to be evaluated. The response is the presence or absence of anxiety symptoms collected at irregular times points during the therapy. Based on a logistic latent variable approach (Rasch, 1960), we introduce a log-linear regression model for repeated categorical data which allows for subject-specific time-constant and time-dependent covariates, a variable number of observations per subject and time periods of arbitrary length. In Section 2 general aspects of psychobiological research in the context of depression theory and the aims and hypotheses of the study are presented. Section 3 refers to categorical latent variable models and reformulates these ideas to incorporate rather flexible data structures using a log-linear representation. Section 4 shows how this model is specified for the data from the psychobiological study, the results are given in Section 5. In Section 6 we discuss several alternative models for repeated categorical data. Finally, in an Appendix we generalize our method to ordered and unordered polytomous responses.

2 A Psychobiological Study on Depression

2.1 General

For the last decades, antidepressant drugs have been successfully used in treatment and prophylaxis of patients with depressive syndromes. It is still a topic of research, however, to identify conditions under which patients are likely to benefit from antidepressant treatment. Two major pathways in psychobiological research, i.e., Psychopharmacology and Psychoneuroendocrinology, focus on the exploration of a biological substrate in "functional" psychiatric disorders.

One aim of pharmacological considerations is to establish effective drug treatment strategies. Whereas pharmacokinetic investigations mainly concentrate on dose - plasma level relations pharmacodynamic research is concerned with the effects of drug administration. In line with basic pharmacological concepts a dose (or plasma level) - therapeutic response relation can be expected for antidepressants, however, may not necessarily be disclosed in every study. A traditional way to demonstrate such a relationship is the determination of the drug in the plasma and its association to clinical syndromes. A considerable number of drug level studies have been reported in the psychiatric literature but the results do not show a clear picture, particularly in the case of antidepressants. Whereas some authors found linear associations others detected U-shaped relations (very low and very high plasma levels are less favorable than medium levels) between blood concentrations of a certain drug and clinical outcome (for a review see Peet and Coppen, 1979). The establishment of a systematic dose-clinical response relation, however, is a necessary requirement to support the assumption that the observable efficacy of a drug is based on its pharmacological properties.

Additionally, it seems reasonable to account for the complexity of the system by investigating further psychobiological mechanisms that may presumably be involved. Psychoneuroendocrinological research indicates that the perturbations in hormone levels may reflect certain important central nervous system processes. In particular, investigations of the pituitary-thyroid subsystem revealed possible associations between thyroid dysfunction and psychiatric conditions (for a review see Prange and Loosen, 1982). One of the hormones secreted by hypothalamus – thyrotropin-releasing hormone (TRH), which regulates the thyroid subsystem – releases thyroid-stimulating hormone (TSH, or thyrotropin) from the pituitary. This physiologically occuring bioregulation provides a basis for a clinical test, the so-called TRH test. Intravenous administration of a supramaximal dose of TRH elevates the plasma level of TSH by more than 5 μU/ml (micro units per milliliter) over baseline within 45 minutes. Such a response is assumed to be "normal" whereas a response of less than 5 μU/ml is called a "blunted" response. Using this test, a high rate of blunted TSH responses in patients with primary depressive disorder can be demonstrated. Only few studies have raised questions of concern for therapy. In two studies a blunted TSH response at admission predicted a favourable outcome to antidepressants. Interestingly, the blunted TSH response becomes "corrected" with treatment in patients who show clinical recovery. It has been observed that a normalization ("disblunting") of a blunted TSH response during therapy may predict successful outcome with antidepressants (for a review see Langer et al., 1989).

When combining these results with the pharmacological assumption of a drug level - clinical response relationship, a hypothesis for the mechanism of drug action can be formulated, stating that a blunted TSH response before treatment indicates a certain psychobiological state that facilitates pharmacological efficacy. Clinical recovery is associated with a normalization of the TSH response (disblunting) during drug treatment, provided that the drug plasma level is within a certain therapeutic range. The reason for the assumption of a therapeutic plasma concentration range, i.e., a U-shaped rather than a linear relation, is based on the consideration that very high dosages of any psychopharmacological substances are unfavourable due to toxic or paradox effects, or at least severe side effects.

2.2 The Study

In this paper we consider data from a clinical study conducted between 1984 and 1987 at the Psychobiological Research Ward, Department of Psychiatry, University of Vienna. The sample consisted of 45 female patients with a depressive syndrome who were treated with the antidepressant *Clomipramine*. Except for establishing steady state conditions (constant dosage within 3 days before taking the blood sample 12 hours after the last drug administration) no attempt was made to standardize the dosage of Clomipramine for two reasons: the study should reflect a natural clinical situation, where the dosage usually depends on the assessment of the individual patient's symptoms, and individual dosages were expected to produce the necessary variation of plasma levels required in the statistical analysis. TRH tests were performed prior to therapy and (roughly) in weekly intervals during therapy. Blood levels were determined at the same times as well as psychiatric variables, in particular anxiety as one of the major symptoms of depression. The therapeutic plasma level range was defined to be between 25 and 150 ng/ml. The data are given in Appendix B.

The number of observation periods was from two up to eleven, i.e. there was a varying number of records among the patients due to clinical conditions. Besides the dichotomous response variable: anxiety (coded as 1 for symptoms present) at each observation *time* `T` (days on treatment) three variables, all of them defined as two-level factors, were included into the analysis: `P`, the TSH (Thyroid stimulating hormone) response to TRH (Thyroid releasing hormone) *prior* to therapy (coded as 1 for a 'blunted' response), `D`, the TSH response *during* therapy (coded as 1 for a 'nonblunted' response indicating normalized psychobiological stress) and `C`, the Clomipramine plasma level (coded as 1 for a level within the hypothesized therapeutic range). Whereas `P` is a constant covariate, the other two variables are time dependent. The study tried to answer the following questions: (1) Is a general trend to recovery from anxiety symptoms observable? (2) Are plasma levels within the therapeutic window of 25-150ng/ml favourable to clinical outcome? (3) Is a blunted TSH response before treatment and a normalization of the TSH response during treatment associated with clinical recovery? (4) Are there interactions between the pharmacological and the neuroendocrine variables with respect to therapeutic outcome?

3 Latent Variable Models for Discrete Responses

A main feature of longitudinal data analysis is the representation of data as clusters. This concerns mainly observational units from whom data have been repeatedly collected. Within-subject effects can be of crucial importance and, consequently, have to be modelled in such a way that they can be separated from the treatment effects, that are of interest. Latent variable models allow for this segregation. Lazarsfeld (1950) introduced the concept of latent structures and distinguished between manifest observations and unobservable latent traits. Covariation among observations should only be due to their common dependence on parameters characterizing the latent trait. These subject-specific effects that account for the dependencies amongst the responses may be dealt with in two different ways. One is, to treat them as random effects, assuming some distribution for them, and integrate the likelihood with respect to that distribution. The resulting marginal likelihood is then maximized to obtain estimates for the parameters of interest. A second way is to treat the subject effects as nuisance parameters and condition on the sufficient statistics for them. The resulting conditional likelihood depends only on the parameters that are the objective of the analysis. Often, such models are computationally complex, but it will be shown that conditional ML estimates have representations as usual ML estimates for certain log-linear models.

After a short review of the Rasch model and its linearized version this section will give a general formulation of models for repeated binary data accounting for time dependent covariates that may be specific to subjects.

3.1 The Rasch Model and a Linearized Version

Given a sample of subjects $i,\ i = 1, \ldots, n$, who have responded to items $j,\ j = 1, \ldots, J$, so that the observations y_{ij} can be regarded as realizations of a Bernoulli variable Y_{ij} (coded by 1 or 0, respectively) Rasch (1960) proposed the now well-known model

$$P(Y_{ij} = y_{ij}|\xi_i) = \frac{\exp(y_{ij}(\xi_i - \lambda_j))}{1 + \exp(\xi_i - \lambda_j)} \ . \tag{1}$$

Thus the probability of "correctly" responding to item j is dependent on the location ξ_i of subject i on the latent continuum ξ ("ability") and λ_j describes the "difficulty" of item j. A main assumption of latent variable models is that given ξ_i, the responses on separate items by the same subject are independent. The observations y_{ij} provide the $(n \times J)$–data matrix $\mathbf{Y}$ with marginal scores $r_i = \sum_j y_{ij}$ and $s_j = \sum_i y_{ij}$.

To avoid the "incidental parameter " problem discussed by Neyman and Scott (1948) estimation usually follows the conditional maximum likelihood (CML) approach. The main idea of the CML–method is to condition on the sufficient statistics $R_i = \sum_j Y_{ij}$ for the parameters ξ_i. Then the parameters ξ_i do not occur in the CML–equations. The kernel of the conditional likelihood is given by

$$L_c = \exp(-\sum_j \lambda_j s_j) / \prod_r \gamma(r; \lambda_1, \ldots, \lambda_J)^{n_r} \ , \tag{2}$$

where $\gamma(r; \lambda_1, \ldots, \lambda_J) = \sum_{A_r} \exp(-\sum_j \lambda_j y_{ij})$, A_r denotes the set of all possible response patterns with marginal r, i.e., $A_r = \{(x_1, \ldots, x_J) | \sum x_j = r, x_j = 0, 1\}$, and n_r is the number of subjects with $R_i = r,\ r = 0, \ldots, J$. A review on estimation theory for the Rasch model was given by Lindsay, Clogg and Grego (1991), and Fischer and Molenaar (1995).

Several authors (see e.g. Fienberg, 1981; or Tjur, 1982) showed that the Rasch model corresponds to a log-linear model and thus can be fitted using standard software (e.g., GLIM, see Hatzinger, 1989). The log-linear formulation of the Rasch model based on the conditional approach is obtained by using

$$P(Y_{i1} = y_{i1}, \ldots, Y_{iJ} = y_{iJ}|R_i = r) = \frac{\exp(-\sum_j \lambda_j y_{ij})}{\gamma(r; \lambda_1, \ldots, \lambda_J)} \ .$$

Since this probability is equal for all subjects with response pattern $\mathbf{y} = (y_1, \ldots, y_J)$ given marginal score r we have

$$\ln E(n_r) = \sigma_r - \sum_j \lambda_j x_j \ , \tag{3}$$

where the x_j's are dummy variables representing the row of the design matrix corresponding to response pattern $\mathbf{y}$,

$$\sigma_r = \ln \left(\frac{n_r}{\gamma(r; \lambda_1, \ldots, \lambda_J)} \right) \ ,$$

and n_r is the number of subjects with response pattern $\mathbf{y}$, which has marginal score r. Equation (3) is a log-linear model describing quasi-independence structures in contingency tables with structural zeros, a class of models first introduced by Goodman (1968). Fienberg (1981) related the log-linear representation of (1) to quasi-symmetry.

One of the several extensions of the Rasch-model (RM) was introduced by Fischer (1974), where he incorporated a linear structure into (1). Fischer termed this model Linear Logistic Test Model (LLTM). Assuming the Rasch model holds, the item parameters λ_j can be reparameterized by

$$\lambda_j = \sum_p u_{jp}\eta_p \quad \text{for } j = 1, \ldots, J, \quad p = 1, \ldots, P, \quad P < J. \tag{4}$$

The η_p's are called effect parameters describing certain characteristics of items and the u_{jp} are the corresponding covariates. Inserting (4) into (3) provides the model

$$\ln E(n_{\mathbf{y}}) = \sigma_r - \sum_p \eta_p \sum_j u_{jp} x_j \ . \tag{5}$$

Like the log-linear Rasch model this linearized version is a generalized linear model and can be fitted using standard software.

3.2 A Latent Variable Model for Repeated Dichotomous Responses

Two kinds of latent variable models can be distinguished that take repeated categorical responses into account. The first class of models incorporates time effects into the log-linear Rasch model, by alternative parameterization of the item parameters. Duncan (1985) used the log-linear representation of a linearized version of the RM to analyze data from a three-wave panel study with dichotomous responses at each wave. The main idea was to use the correspondence between the responses of subjects of a questionnaire consisting of J different items that measure the same trait and the classification of subjects according to the same question at J different times. This equivalence follows from the unidimensionality assumption of the RM. To take serial dependencies into account, Duncan introduced parameters into his model defining a Markov type structure. Agresti (1993) applied this approach to the analysis of data from an orthogonal Latin square cross-over design for three different treatments. Following Kenward and Jones (1991), he explicitely incorporates a parameterization describing the six possible treatment groups. Time dependences, such as period and carry-over effects are modelled as treatment-by-period interactions. However, common to these two approaches is that the length of the time periods between observations is not considered.

The second class consists of models that have been developed in the psychometrical context and have particularly been designed to provide means for the analysis of item response data, observed repeatedly. Fischer (1989), for example, suggested a model for dichotomous longitudinal data. This model extends the LLTM to designs with arbitrary number of time points and allows for using different sets of items, possibly presented at different occasions. The logistic formulation of these models allows to deal with a rather large number of items, but suffers from computational complexity due to the conditional ML estimation and, moreover, in practical application special purpose software is required.

Both types of models, though suggested from different points of view, have in common that conditional likelihood methods are used to eliminate nuisance parameters, which usually are subject effects. The aim of this section is to generalize these ideas by suggesting a model for dichotomous responses that combines the advantages of both approaches: log-linear representation, inclusion of time periods of arbitrary length, variable number of observations per subject and subject-specific covariates. A generalization to models with polytomous responses will be given in Appendix A.

Let $\mathbf{y}_i^*$ be the response pattern for subject i, $i = 1, \ldots, n$, responding to J_i items at T_i times, i.e., $\mathbf{y}_i^* = (y_{i11} \ldots, y_{ijt}, \ldots, y_{iJ_iT_i})$ with $y_{ijt} = 1$ if subject i responds positively to item j at time t (0, otherwise) and score $r_i = \sum_{j,t} y_{ijt}$. By crossclassifying the possible reponses of subject i we can construct a $2^{J_i \cdot T_i}$ contingency table. Let $\psi_{\mathbf{y}_i^*}$ be the probability for subject i to enter cell $\mathbf{y}_i^*$ in the corresponding contingency table. Then a log-linear model that allows for subject-specific effects is

$$\ln \psi_{\mathbf{y}_i^*} = \sum_j \sum_t \tau_{ijt} x_{ijt} - \rho_{r_i} \ , \tag{6}$$

where the x's are again dummy variables corresponding to the rows of the design matrix. The parameter ρ_{r_i} is a normalizing constant describing nuisance subject effects and corresponds to the symmetry parameters in the equivalent quasi-symmetry log-linear model (cf. Agresti, 1993). As usual, the parameters τ can be used to represent various interesting odds. Consider, e.g., the case $J_i = T_i = 2$, the log odds of response $(00, 10)$ in favor of $(01, 00)$ are given by

$$\ln \frac{\psi_{(00,10)}}{\psi_{(01,00)}} = \tau_{i12} - \tau_{i21} \, .$$

The special case $J_i = J$ and $T_i = T$, for all subjects i, is considered in Agresti (1997), where he formulates a Rasch type model using marginal ML. Under certain specifications, the nonparametric treatment of the random effects implies a multivariate log-linear model that conincides with model (6).

To incorporate concomitant information, the parameters τ_{ijt} may be linearly reparameterized as

$$\tau_{ijt} = \sum_{p=1}^{P} u_{itp} \eta_p \ , \quad j = 1, \ldots, J_i \ , \quad t = 1, \ldots, T_i \ . \tag{7}$$

The covariates u_{itp} may be quantitative as well as categorical and are assumed to be equal for all items j. We adopt the psychometrical concept of unidimensionality here saying that all items j measure the same latent variable or, more formally, all τ_{ijt}'s reflect changes in location along one latent dimension. A generalization would require to consider models with parameters τ_{ijkt} (possibly reparametrized by $\tau_{ijkt} = \sum_k \sum_p u_{iktp} \eta_{kp}$), where $j = 1, \ldots, J_k$, and $k = 1, \ldots, K$. The index k can then be assumed to describe K different latent dimensions. For example, suppose the two latent dimensions *Side effect* and *Disease Status* are measured by one item each, i.e., $K = 2$ and $J_k = 1$. The parameters η_{1p} and η_{2p} could then be the effect of the pth treatment on the side effects and on the disease status, respectively. However, we will not further consider this generalization here.

The parameters η in (7) reflect the effects of change in the covariates on the log odds. For the example above, this is

$$\tau_{i12} - \tau_{i21} = \sum_p (u_{i2p} - u_{i1p}) \eta_p \, .$$

By introducing the linear reparameterization (7) model (6) becomes

$$\ln \psi_{\mathbf{y}_i^* | \mathbf{u}_i} = \sum_p u_{ip}^* \eta_p - \rho_{r_i} \ , \tag{8}$$

where $\psi_{\mathbf{y}_i^* | \mathbf{u}_i}$ is the probability, that subject i with covariate vector $\mathbf{u}_i$ enters cell $\mathbf{y}_i^*$ in the $2^{J_i \cdot T_i}$ contingency table. The sufficient statistics for the parameters η_p are $\sum_j \sum_t u_{itp} Y_{ijt}$. The u_{ip}^*s are thus the number of times subject i has responded positively to items j, $j = 1, \ldots, J_i$, multiplied by the value of the covariate for η_p at times t, and ρ_{r_i} is again a normalizing constant.

4 Fitting the Model to the Anxiety Data

Rewriting model (6) in matrix notation gives the (conventional) log-linear model for subject i

$$\ln \boldsymbol{\psi}_i^* = \boldsymbol{A}_i\boldsymbol{\theta}_i = \mathbf{1}\alpha + \boldsymbol{X}_i\boldsymbol{\tau}_i - \boldsymbol{R}_i\boldsymbol{\rho}_i \ , \tag{9}$$

where $\boldsymbol{\psi}_i^* = (\psi^*_{i,0\ldots0}, \psi^*_{i,0\ldots01}, \ldots, \psi^*_{i,1\ldots1})'$, $\boldsymbol{A}_i = (\mathbf{1}, \boldsymbol{X}_i, \boldsymbol{R}_i)$ is an appropriately chosen design matrix, and $\boldsymbol{\theta}_i$ is a vector of parameters, $\boldsymbol{\theta}_i = (\alpha, \tau_{i11}, \ldots, \tau_{i,J_i,T_i}, -\rho_{i0}, \ldots, -\rho_{i,J_i\cdot T_i})'$. Since the ρ's are obviously overparameterized we impose the restriction $\rho_{i,J_i\cdot T_i} = 0$, for all i. Since we set the first column of $\boldsymbol{A}_i$ to $\mathbf{1}$, we have to additionally restrict $\rho_{i0} = 0$. For example, the restricted design matrix for patient 20 (see Table 1) can be read off as

$$\boldsymbol{A}_i = \begin{pmatrix} 1&0&0&0&0&0 \\ 1&0&0&1&1&0 \\ 1&0&1&0&1&0 \\ 1&0&1&1&0&1 \\ 1&1&0&0&1&0 \\ 1&1&0&1&0&1 \\ 1&1&1&0&0&1 \\ 1&1&1&1&0&0 \end{pmatrix}$$

i.e.

response pattern	Y	1	X0	X1	X2	R1	R2
000	0	1	0	0	0	0	0
001	0	1	0	0	1	1	0
010	0	1	0	1	0	1	0
011	0	1	0	1	1	0	1
100	0	1	1	0	0	1	0
101	1	1	1	0	1	0	1
110	0	1	1	1	0	0	1
111	0	1	1	1	1	0	0

where Y in the right table denotes the response vector. To obtain the design matrix under model (8) we have to transform (9) into

$$\ln \boldsymbol{\psi}_i^* = \boldsymbol{B}_i\boldsymbol{\lambda}_i = \mathbf{1}\alpha + \boldsymbol{U}_i^*\boldsymbol{\eta} - \boldsymbol{R}_i\boldsymbol{\rho}_i \ , \tag{10}$$

by using

$$\boldsymbol{U}_i^* = \boldsymbol{X}_i(\boldsymbol{U}_i \otimes \mathbf{1}_{J_i}), \tag{11}$$

where $\boldsymbol{U}_i = (u_{itp})$ is the $(T_i \times P)$ matrix of covariates and, possibly, interactions between covariates and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_P)'$. Care has to be taken when specifying the design matrix. As in conditional analysis there is no information about coefficients of covariates which do not vary over time (see, e.g., Diggle et al., 1994, p.178) and, accordingly, only differences between the τ's can be estimated. Failure of the estimation procedure occurs if $u_{ipt} = u_{ipt'}$ for all i or $u_{ipt} = u_{i'pt}$ for all t, since

$$\tau_{it} - \tau_{it'} = \sum_p (u_{ipt} - u_{ipt'})\eta_p = 0$$

and

$$\tau_{it} - \tau_{i't} = \sum_p (u_{ipt} - u_{i'pt})\eta_p = 0 \ . \tag{12}$$

A possible solution to the estimation problem of constant variates (such as gender) is to multiply the corresponding u_{ip} by t (or a function of t). A specific example and, moreover, a description how to include interaction terms into the model will be given in Section 5.

An example for a design matrix $\boldsymbol{B}_i$ (including main effects for the time varying covariates T, C, and D) for patient 20, e.g., is given as

$$\boldsymbol{B}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 25 & 1 & 1 & 1 & 0 \\ 1 & 18 & 1 & 1 & 1 & 0 \\ 1 & 43 & 2 & 2 & 0 & 1 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 1 & 28 & 1 & 1 & 0 & 1 \\ 1 & 21 & 1 & 1 & 0 & 1 \\ 1 & 46 & 2 & 2 & 0 & 0 \end{pmatrix} \qquad \text{i.e.}$$

response pattern	Y	1	T	C	D	R1	R2
000	0	1	0	0	0	0	0
001	0	1	25	1	1	1	0
010	0	1	18	1	1	1	0
011	0	1	43	2	2	0	1
100	0	1	3	0	0	1	0
101	1	1	28	1	1	0	1
110	0	1	21	1	1	0	1
111	0	1	46	2	2	0	0

In computer packages such as R (2008) the covariate vectors T, C, and D for patient 20 could be generated by appropriate matrix commands or something like:

```
T <- X0*3 + X1*18 + X2*25      observation days: 3, 18, 25
C <- X0*0 + X1*1 + X2*1        Clomipramine: 0, 1, 1
D <- X0*0 + X1*1 + X2*1        TSH (during): 0, 1, 1
```

The log-linear model (10) for all subjects can be written as

$$\ln \boldsymbol{\psi}^* = \mathbf{1}\alpha + \boldsymbol{U}^*\boldsymbol{\eta} - \boldsymbol{R}^*\boldsymbol{\rho} \;, \tag{13}$$

where $\ln \boldsymbol{\psi}^* = (\ln \boldsymbol{\psi}_1^{*'}, \ldots, \ln \boldsymbol{\psi}_n^{*'})'$, $\boldsymbol{\eta}$ is defined as above, $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \ldots, \boldsymbol{\rho}_n')'$,

$$\boldsymbol{U}^* = \begin{pmatrix} \boldsymbol{U}_1^* \\ \vdots \\ \boldsymbol{U}_n^* \end{pmatrix} \;, \quad \text{and} \quad \boldsymbol{R}^* = \operatorname{diag}(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n) \;.$$

The complete table for all subjects can be graphically represented as given in Figure 1.

Regarding (13), it is implicit in the log-linear versions of the model presented, that a $2^{J_i \cdot T_i}$ table has to be specified for each subject i, where there is only one entry corresponding to his/her response pattern. If subject i' has covariates equal to subject i a second entry occurs in the the ith table and no separate table is constructed for i'.

5 Results

Different specifications of the design matrix $\boldsymbol{U}^*$ correspond to different hypotheses and allow for model selection using likelihood ratio tests (difference of deviances) in a hierarchical sequence of fitted models. The maximal model contains all higher order interaction terms and can be written as

$$\begin{aligned} &\mathtt{T}*\mathtt{P}*\mathtt{D}*\mathtt{C} = \\ &\mathtt{T}+\mathtt{P}+\mathtt{D}+\mathtt{C}+\mathtt{T}:\mathtt{P}+\mathtt{T}:\mathtt{D}+\mathtt{T}:\mathtt{C}+\mathtt{P}:\mathtt{D}+\mathtt{P}:\mathtt{C}+\mathtt{D}:\mathtt{C}+ \\ &\mathtt{T}:\mathtt{P}:\mathtt{D}+\mathtt{T}:\mathtt{P}:\mathtt{C}+\mathtt{T}:\mathtt{D}:\mathtt{C}+\mathtt{P}:\mathtt{D}:\mathtt{C}+\mathtt{T}:\mathtt{P}:\mathtt{D}:\mathtt{C} \end{aligned} \tag{14}$$

according to the notation of Wilkinson and Rogers (1973) and as used in R. Please note that any interaction terms have to be included in $\boldsymbol{U}$ and calculated prior to applying the reparameterization (11). For example, the term C:D for patient 20 can be specified by

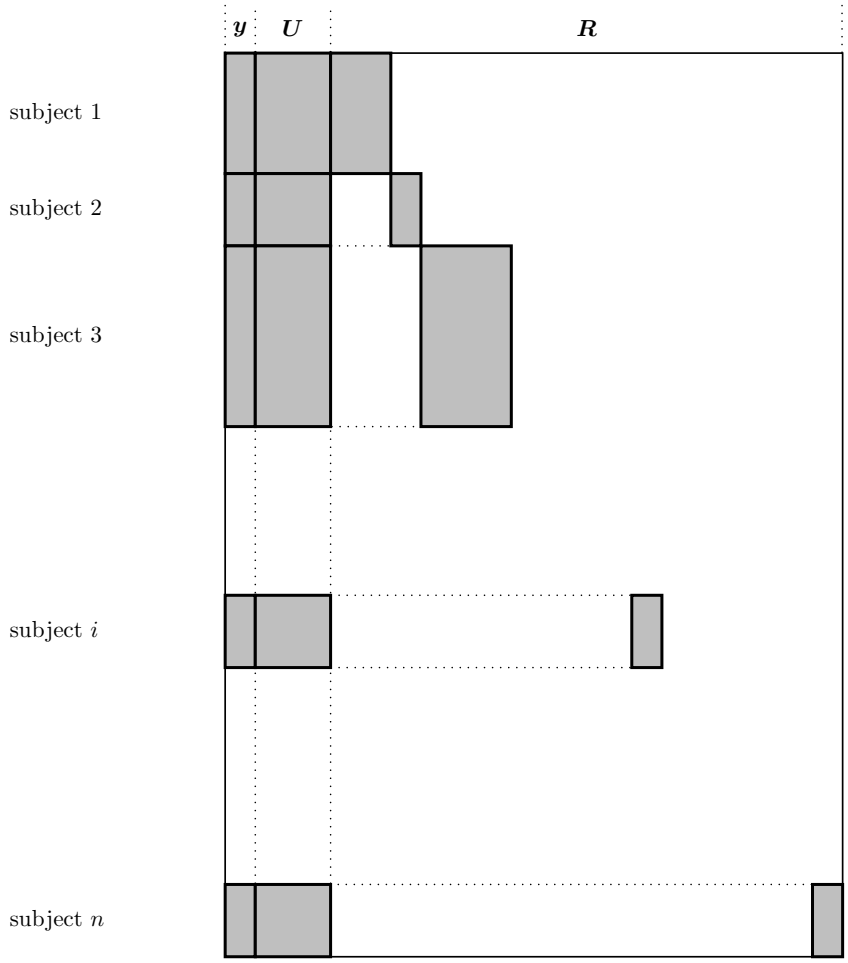


Figure 1: Graphical Representation of Model (13). The responses are represented by $\boldsymbol{y}$, $\boldsymbol{U}$ are the covariate matrices, and $\boldsymbol{R}$ are the dummies for the nuisance parameters, which take the value zero everywhere except for the individual submatrices.

`DC <- X0*0*0 + X1*1*1 + X2*1*1` Clomipramine: 0, 1, 1 – TSH (during): 0, 1, 1

Following the considerations concerning time-constant covariates in Section 4 all terms in (14) containing `P` must be $\texttt{T:P} \stackrel{def}{=} \texttt{TP}$ and may be set up (again for patient 20) by

`TP <- X0*33*1 + X1*18*1 + X2*25*1` observation days: 3, 18, 25 – TSH (prior): 1

This new covariate TP acts as a main effect and therefore, all terms `T:P` in (14) have to be replaced and the interaction terms adjusted correspondingly. The maximal model is thus

$$\texttt{1} + \texttt{T} + \texttt{TP} + \texttt{D} + \texttt{C} + \texttt{TD} + \texttt{TC} + \texttt{TPD} + \texttt{TPC} + \texttt{DC} + \texttt{TDC} + \texttt{TPDC}$$

Unfortunately, usual procedures for model selection based on goodness-of-fit statistics fail. Due to the large number of zero cells the number of dfs is inflated and cannot be used relative to the general

likelihood ratio or to the Pearson statistic, which might be far from the asymptotic χ^2-distribution. However, the sample size in sparse tables is often sufficiently large to use the likelihood ratio test for comparing two nested models. The null distribution of these statistics converge to its limiting χ^2-distribution more quickly than the overall goodness-of-fit statistics since they depend on the data only through the sufficient statistics for the marginals rather than through individual cell frequencies, and, as for most log-linear models the expected values refer to marginal tables. Thus, the χ^2 approximation is likely to be adequate for the model-based statistics in most cases (cf. Haberman, 1977b).

To find a suitable model a backward elimination strategy outlined, e.g., in Christensen (1990, p. 128ff) was applied. The first step is to find an appropriate initial model from where the backward selection process can be started. A suitable choice is a model that contains all effects of a particular level s, i.e., the smallest all s factor model that fits the data. Accordingly, LR-tests (differences of deviances) between the models containing all interaction effects of order s and the models consisting of all $s-1$ interaction terms were used. The results are given in Table 1.

Table 1: Initial Model Selection.

model	s	Δ deviance	Δ df
maximal model	4	–	–
third order interaction	3	1.220	1
all second order interaction	2	1.232	1
all main effects	1	15.387	5

Since the increase of the deviance between the model containing the third order interaction and all second order interaction terms is less than the critical χ^2 value but is significant ($\alpha = 0.05$ is used throughout) for the difference between the all two-factors model and the all main effects model a well fitting representation of the data should contain at least one second order interaction effect.

The second step of the selection procedure was to start with the all two-factor interaction model as the initial model and to reduce it by deleting the interaction term with the smallest increase in deviance. The selection procedure based on the reduced model was then continued analogously until achieving with a suitable minimal model. Table 2 gives the results of the reduction procedure, i.e., the deleted second order effects and the corresponding deviance increases ($df = 1$ throughout).

Table 2: Results of successively eliminating second-order terms.

deleted term	deleted in step	Δ deviance
TPD	1	0.012
TD	2	0.130
TC	3	2.650
TPC	4	2.458

NOTE: The Δ deviance entries are computed with respect to the model including the deleted term.

The only remaining important 2-factor effect is `C.D`. A last step is to test the effects of `TP` and, finally, of `T`. The corresponding LR-tests yield 2.396, and 9.348 respectively, both with $df = 1$. Thus the term `TP` but not `T` can be removed from the model. Accordingly, a suitable representation of the data is the model `T+D+C+DC` . The parameter estimates are given in Table 3 (the estimates for the grand mean and the nuisance parameters are omitted).

The interpretation is as follows: (1) There is a slight general trend towards recovery which might be ascribed to different unspecific influences of the therapeutic setting. The effect of the covariate time `T` on the odds in favour of recovery can be estimated by $\exp\{-(\sum_t T_t)\hat{\eta}_{\mathtt{T}}\}$, which is greater

Table 3: Parameter estimates of the final model

parameter	estimate	standard error
T	-.0347	.01236
D	1.3126	.82749
C	.5535	.64646
DC	-2.9072	.93113

than one due to $\hat{\eta}_{\mathtt{T}} < 0$. (2) There is a favourable interaction effect between Clomipramine levels (within the therapeutic range) combined with nonblunted TSH responses during therapy. This interaction effect DC on the odds in favour of recovery can be estimated by

$$\exp\{-\#[\textit{nonblunted response} \text{ and } \textit{therapeutic Clomipramine levels }]\hat{\eta}_{\mathtt{DC}}\}\,.$$

Having once estimated the parameters η, various interesting log odds can be calculated. For instance, consider a fictitious patient which is observed at 5 different times. Consider further the log odds for the best case (showing never anxiety symptoms) compared to the worst case (showing always anxiety symptoms), i.e.

$$\ln \frac{\psi_{(00000)}}{\psi_{(11111)}}\,.$$

Then, apart from additive constants (time effects and nuisance parameters) the combined effect of the covariates D and C on this log-odds can be seen from Figure 2.

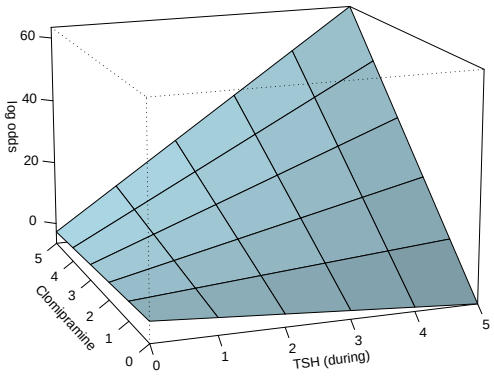


Figure 2: Estimated log-odds (best vs. worst case) for different values of Clomipramine therapeutic levels and normalized TSH responses.

To summarize the results of the study: there is some evidence for a general trend towards recovery. The hypothesis concerning TSH-responses for the pretherapeutic assessment (cf. Section 2.2) is not supported. Finally, there is a clear indication for interactions favourable to clinical outcome between the pharmacological variable (plasma levels within the therapeutic range of 25-150ng/ml) and the neuroendocrine variable (TSH-normalization during treatment).

6 Discussion

Several other methods for the analysis of repeated categorical data have been reported in the literature. Liang and Zeger (1993) discuss three different types of models – marginal, observation-driven (transitional) and random effects models – for correlated data that address the dependence between the responses. In marginal models the marginal expectation (or 'population average') is the average response over the population of individuals (or clusters) with a common value of a covariate. In observation-driven model the conditional distribution of the actual response given the entire past is modelled as a function of the explanatory variables and explicitely as a function of the past responses themselves. The main characteristic of random effects models is the assumption that parameters vary from cluster to cluster and thus reflect natural heterogeneity due to unmeasured factors. The random effects model proves to be especially useful in situations where the main scientific interest concerns 'subject–specific' rather than 'population–averaged' effects. Whereas parameters in population-averaged models describe differences in marginal distributions of the repeated response in terms of marginal probabilities, parameters in subject-specific models describe differences in a way that directly incorporates the dependence of repeated responses into the joint distribution.

Latent variable models are closely related to the random effects approach in the way that the subject random effect corresponds to the subject's location on the latent variable. In the conditional latent variable approach, however, it is not necessary to specify a distribution for the subject random effects. The analysis allows for distribution-free modelling and is equivalent to an extended form of marginal maximum likelihood methods (Kelderman, 1984), where no assumptions are made about the random effects distribution. Cressie and Holland (1982) discuss moment-inequalities in the RM. If CML estimates satisfy these contraints the conditional and the random effects methods yield the same results. The CML approach requires to restrict oneself to the usage of the logit link. The advantage of logistic models, however, is their equivalence to log-linear models which can easily be fitted using standard software, e.g., R.

References

Agresti, A. (1993). Distribution-free fitting of logit models with random effects for repeated categorical responses. *Statistics in Medicine*, **12**, 1969-1987

Agresti, A. (1997). A model for repeated measurements of a multivariate binary response. *Journal of the American Statistical Association* , **92**, 315-332

Andersen, E.B. (1973). Conditional inference and multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, **26**, 31-44

Christensen, R. (1990). *Log-Linear Models*. Springer-Verlag, New York

Cressie, N., Holland, P.W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, **48**, 129-141

Diggle, P.J., Liang, K-Y., Zeger, S.L. (1994). *Analysis of Longitudinal Data*, Clarendon Press, Oxford

Duncan, O.D. (1985). Some models of response uncertainty for panel analysis. *Social Science Research*, **14**, 126-141

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. 3rd Ed., John Wiley & Sons, New York

Fienberg, S.E. (1981). Recent advances in theory and methods for the analysis of categorical data: Making the link to statistical practice. Bull. Int. Stat. Inst., 43rd session

Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests* (Introduction to the theory of psychological tests). Bern: Huber

Fischer, G.H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika*, **54**, 599-624

Fischer, G.H., Molenaar, I.W. (1995). *Rasch models. Foundations, recent developments, and applications.* Springer-Verlag, New York

Friedl, H., Hatzinger, R. (1994). A Note on Generating the Factors for Symmetry Parameters in Loglinear Models. *GLIM-Newsletter*, **24**,33-36

Francis, B.J., Green, M., Payne, C. (eds) (1993). *The GLIM System. Release 4 Manual*, Clarendon Press, Oxford

Goodman, L.A. (1968). The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries. *Journal of the American Statistical Association* , **63**, 1091-1131

Hatzinger, R. (1989). The Rasch-model, some extensions and their relation to the class of generalized linear models. *Statistical modelling: Proceedings of GLIM89 and the 4th International Workshop on Statistical Modelling.* Lecture Notes in Statistics, **57**, Springer, Berlin

Kelderman, H. (1984). Loglinear Rasch Model Tests. *Psychometrika*, **49**, 223-245

Kenward, M.G., Jones, B. (1991). The analysis of categorical data from cross-over trials using a latent variable model. *Statistics in Medicine*, **10**, 1607-1619

Langer, G., Koinig, G., Schönbeck, G., Hatzinger, R. (1989). Neuroendocrine Factors in Antidepressant Drug Therapy. In: Lerer, B., Gershon, S. (eds.): *New Directions in Affective Disorders.* Springer-Verlag, New York

Lazarsfeld, P.F. (1950). The interpretation and computation of some latent structures. In Samuel S. Stouffer et al. (Eds.) *Measurement and prediction in World War II*, **4**, Princeton: Princeton University Press

Liang, K.Y., Zeger, S.L. (1993). Regression analysis for correlated data. *Annual Review of Public Health*, **14**, 43-68

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`.

Lindsay, B., Clogg, C.C., Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* , **86**, 96-107

Neyman, J., Scott, E.L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrika*, **16**, 1-32.

Peet, M., Coppen, A. (1979). The pharmacokinetics of antidepressant drugs: Relevance to their therapeutic effects. In: Paykel, E.S., Coppen, A. (eds): *Psychopharmacology of affective disorders.* Oxford University Press, New York - Toronto

Prange, A.J., Jr., Loosen, P.T. (1982). Hormone therapy in depressive diseases. In: Costa, E., Racagni, G. (eds): *Typical and atypical antidepressants: clinical practice.* Raven Press, New York

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Paedagogiske Institut

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics*, **9**, 23-30

Wilkinson, G.N., Rogers, C.E. (1973). Symbolic Description of Factorial Models for Analysis of Variance. *Applied Statistics*, **22**, 392–399

A Appendix: Extension to Polytomous Responses

Having once reformulated models (1) and (4) into log-linear form it seems natural to look for polytomous extensions in the sense that the response is not restricted to be binary but may be nominal or ordinal. The nominal model is simply introduced by adding an index m for an answer in one of M categories, i.e., $m = 1, \ldots, M$. The response pattern for subject i, responding to J_i items with M categories at T_i times can be written as a $(J_i T_i M \times 1)$ vector $\boldsymbol{y}_i = (\boldsymbol{y}'_{i11}, \ldots, \boldsymbol{y}'_{iJ_i1}, \ldots, \boldsymbol{y}'_{ijt}, \ldots, \boldsymbol{y}'_{i1T_i}, \ldots, \boldsymbol{y}'_{iJ_iT_i})'$, where the $(M \times 1)$ subvectors $\boldsymbol{y}_{ijt}$ are given as $\boldsymbol{y}_{ijt} = (y_{ijt1}, \ldots, y_{ijtm}, \ldots, y_{ijtM})'$, with $y_{ijtm} = 1$, if subject i responds to item j at time t in category m, and zero otherwise. Crossclassification of all possible response patterns for subject i yields a $M^{J_i \cdot T_i}$ contingency table. The log-linear formulation of the multinomial Rasch model, extended from (6), is

$$\ln \psi_i = \sum_j \sum_t \sum_m \tau_{ijtm} x_{ijtm} - \rho_{\boldsymbol{r}_i} \ ,$$

where ψ_i denotes the probability for subject i to enter cell $\mathbf{y}_i$, and the $x's$ are dummy variables corresponding to the rows of the design matrix. The parameters $\rho_{\boldsymbol{r}_i}$ with $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{iM})$ are normalizing constants and correspond to the symmetry parameters in the equivalent quasi-symmetry log-linear model. The elements r_{im} of the vector $\boldsymbol{r}_i$ denote the number of responses for subject i in category m. Note that there are k parameters $\rho_{\boldsymbol{r}_i}$, where k denotes the number of configurations of the form $\{(x_1, \ldots, x_M) | x_j \geq 0, \sum_j x_j = J_i T_i\}$, i.e., $k = \binom{J_i \cdot T_i + M - 1}{M-1}$, cf. Feller (1968, vol.1, p.38). Using matrix notation, the multinomial Rasch model is analogously to (9) given as

$$\ln \boldsymbol{\psi}_i = \boldsymbol{A}_i \boldsymbol{\theta}_i = \mathbf{1}\alpha + \boldsymbol{X}_i \boldsymbol{\tau}_i - \mathbf{R}_i \boldsymbol{\rho}_i \ , \tag{15}$$

where $\mathbf{X}_i$ is a $(M^{J_i \cdot T_i} \times M J_i T_i)$ matrix, $\boldsymbol{\tau}_i$ is a $(M J_i T_i \times 1)$ parameter vector, $\mathbf{R}_i$ is a $(M^{J_i \cdot T_i} \times k)$ matrix, and $\boldsymbol{\rho}_i$ is a $(k \times 1)$ vector containing the nuisance parameters ρ (a method for computing the design matrix $\mathbf{R}_i$ is given in Friedl and Hatzinger, 1994). Since (15) is again overparameterized, all parameters have to be suitably restricted (see the example below).

Consider the simple case $J_i = 1$, $T_i = 2$, and $M = 3$. Here, the $(3^2 \times 6)$ matrix $\mathbf{X}_i$ is given as

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{1}_3 & 0 & 0 & \mathbf{I}_3 \\ 0 & \mathbf{1}_3 & 0 & \mathbf{I}_3 \\ 0 & 0 & \mathbf{1}_3 & \mathbf{I}_3 \end{pmatrix} \ ,$$

and the (9×6) matrix $\mathbf{R}_i$ is given as

$$\mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \ .$$

Since the rank of the design matrix $\mathbf{A}_i$ is eight, we have to impose 5 restrictions. For example, we could restrict $\tau_{i111} = \tau_{i121} = \rho_{i200} = \rho_{i020} = \rho_{i002} = 0$. The parameters τ can again be interpreted as log-odds: consider for example the log-odds of the responses (100,001) as compared to (001,100) then

$$\ln \frac{\psi_{(100,001)}}{\psi_{(001,100)}} = \tau_{i123} - \tau_{i113} \ .$$

To incorporate concomitant information, the parameters τ_{ijtm} may again be linearly reparameterized as $\tau_{ijtm} = \sum_{p=1}^{P} u_{itp}\eta_{pm}$. The multinomial Rasch model can accordingly be reformulated as

$$\ln \psi_i = \sum_p \sum_m v_{ipm}\eta_{pm} - \rho_{\mathbf{r}_i} \,,$$

where $v_{ipm} = \sum_j \sum_t u_{itp}x_{ijtm}$ and the sufficient statistics for the parameters η_{pm} are $\sum_j \sum_t u_{itp}Y_{ijtm}$. In matrix notation, this model can analogously to (10) be rewritten as

$$\ln \boldsymbol{\psi}_i = \mathbf{B}_i\boldsymbol{\lambda}_i = \mathbf{1}\alpha + \mathbf{V}_i\boldsymbol{\eta} - \mathbf{R}_i\boldsymbol{\rho}_i \,,$$

where $\mathbf{V}_i = \mathbf{X}_i\mathbf{H}_i$ with a suitably chosen transformation matrix $\mathbf{H}_i$.

For the special case above with $J_i = 1$, $T_i = 2$, and $M = 3$, consider two subject-specific covariates whose values are contained in the (2×2) matrix $\mathbf{U}_i = (u_{itp})$. The matrix $\mathbf{H}_i$ is then given as

$$\mathbf{H}_i = (\mathbf{U}_i \otimes \mathbf{e}_1, \mathbf{U}_i \otimes \mathbf{e}_2, \mathbf{U}_i \otimes \mathbf{e}_3) \,,$$

where the vectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$ denote three-component unit-vectors. In general, for $J_i = 1$, T arbitrary and M categories, the transformation matrix $\mathbf{H}_i$ is given as

$$\mathbf{V}_i = \mathbf{X}_i\,(\mathbf{U}_i \otimes \mathbf{e}_1, \mathbf{U}_i \otimes \mathbf{e}_2, \ldots, \mathbf{U}_i \otimes \mathbf{e}_M) \,,$$

where the vectors $\mathbf{e}_j\,, j = 1, \ldots, M$ denote M-component unit-vectors.

The log-linear formulation of the multinomial Rasch model for all n subjects can analogously to (13) be written in matrix notation as

$$\ln \boldsymbol{\psi} = \mathbf{1}\alpha + \mathbf{V}\boldsymbol{\eta} - \mathbf{R}\boldsymbol{\rho} \,, \tag{16}$$

where $\ln \boldsymbol{\psi} = (\ln \boldsymbol{\psi}_1', \ln \boldsymbol{\psi}_2', \ldots, \ln \boldsymbol{\psi}_n')'$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_P)'$, and $\boldsymbol{\rho} = (\boldsymbol{\rho}_1{}', \boldsymbol{\rho}_2', \ldots, \boldsymbol{\rho}_n')'$. The matrices $\mathbf{V}$ and $\mathbf{R}$ are given as

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_n \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \operatorname{diag}(\mathbf{R}_1, \ldots, \mathbf{R}_n) \,.$$

The multinomial model can easily be specialized for ordinal responses. The basic idea is to reparameterize the parameters τ by a linear structure, i.e., $\tau_{ijtm} = \tau_{ijt}\phi_m + \kappa_m$, where ϕ_m is a scaling parameter and κ_m is a parameter that reflects the mth category. The ϕ's may in principle be estimated (see, e.g., Andersen, 1973), however, assuming the ϕ's to be known constants which reflect the ordering of the categories, e.g. $\phi_m = m$, an ordinal version of the multinomial model is obtained by

$$\ln \psi_i = \sum_j \sum_t \tau_{ijt} \sum_m m\, x_{ijtm} + \sum_m \kappa_m \sum_j \sum_t x_{ijtm} - \rho_{r_i} \,,$$

where the sufficient statistics for the τ_{ijt}'s and for the κ's are $\sum_m mY_{ijtm}$ and $\sum_j \sum_t Y_{ijtm}$, respectively . It should be noted that the ρ_{r_i}'s are not the symmetry parameters as the $\rho_{\mathbf{r}_i}$'s in the nominal case, but have the same value for all responses with equal marginal score r. In matrix notation, this model can be written as

$$\ln \boldsymbol{\psi}_i = \mathbf{A}_i\boldsymbol{\theta}_i = \mathbf{1}\alpha + \mathbf{X}_i\boldsymbol{\tau}_i + \mathbf{K}_i\boldsymbol{\kappa}_i - \mathbf{R}_i\boldsymbol{\rho}_i \,,$$

where $\mathbf{X}_i$ is a $(M^{J_i \cdot T_i} \times J_iT_i)$ matrix, $\boldsymbol{\tau}_i$ is a $(J_iT_i \times 1)$ vector containing the parameters τ. The matrices $\mathbf{K}_i$ and $\mathbf{R}_i$ are of order $(M^{J_i \cdot T_i} \times M)$ and $(M^{J_i \cdot T_i} \times J_iT_i(M-1)+1)$, respectively, and finally, $\boldsymbol{\kappa}_i$, and $\boldsymbol{\rho}_i$ are parameter vectors.

Consider the example $J_i = 1$, $T_i = 2$, and $M = 3$. The matrices $\mathbf{X}_i$, $\mathbf{K}_i$, and $\mathbf{R}_i$ are then given as

$$\mathbf{X}_i = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 3 & 1 \\ 3 & 2 \\ 3 & 3 \end{pmatrix} , \; \mathbf{K}_i = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} , \; \mathbf{R}_i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

Inserting again the linear reparameterization $\tau_{ijt} = \sum_p u_{itp}\eta_p$, the linearized ordinal model can be written as

$$\ln \psi_i = \sum_p v_{ip}\eta_p + \sum_m \kappa_m \sum_j \sum_t x_{ijtm} - \rho_{r_i}\,, \tag{17}$$

where $v_{ip} = \sum_j \sum_t u_{itp} \sum_m m\, x_{ijtm}$. The sufficient statistics for the parameters η_p are $\sum_j \sum_t u_{itp} \sum_m m\, Y_{ijtm}$. In matrix notation model (17) can be rewritten as

$$\ln \psi_i = \mathbf{B}_i\boldsymbol{\lambda}_i = \mathbf{1}\alpha + \mathbf{V}_i\boldsymbol{\eta} + \mathbf{K}_i\boldsymbol{\kappa}_i - \mathbf{R}_i\boldsymbol{\rho}_i\,.$$

For the special case $J_i = 1$, $T_i = 2$, $M = 3$ and two subject-specific covariates, the design matrix for the η's is

$$\mathbf{V}_i = \mathbf{X}_i\mathbf{U}_i\,,$$

where $\mathbf{U}_i = (u_{itp})$ is a (2×2) matrix containing the covariate values.

For all n subjects the ordinal model can analogously to (16) be written as

$$\ln \boldsymbol{\psi} = \mathbf{1}\alpha + \mathbf{V}\boldsymbol{\eta} + \mathbf{K}\boldsymbol{\kappa} - \mathbf{R}\boldsymbol{\rho}\,,$$

where $\ln \boldsymbol{\psi} = (\ln \boldsymbol{\psi}_1', \ln \boldsymbol{\psi}_2', \ldots, \ln \boldsymbol{\psi}_n')'$, $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_P)'$, $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1,' \boldsymbol{\kappa}_2', \ldots, \boldsymbol{\kappa}_n')$ and $\boldsymbol{\rho} = (\boldsymbol{\rho}_1,' \boldsymbol{\rho}_2', \ldots, \boldsymbol{\rho}_n')$. The matrices $\mathbf{V}$, $\mathbf{K}$, and $\mathbf{R}$ are given by

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_n \end{pmatrix}\,, \quad \mathbf{K} = \operatorname{diag}(\mathbf{K}_1, \ldots, \mathbf{K}_n)\,, \quad \text{and} \quad \mathbf{R} = \operatorname{diag}(\mathbf{R}_1, \ldots, \mathbf{R}_n)\,.$$

B Appendix: Data from the Study on Anxiety Symptoms

1 1 1 0 1 0	7 0 48 0 0 0	14 1 21 0 1 0	21 1 1 0 1 0	29 1 37 1 1 0	38 1 3 0 0 1
1 1 2 0 1 0		14 0 29 0 1 0	21 1 8 1 1 0	29 1 44 1 1 1	38 0 13 1 0 1
1 1 6 0 1 0	8 1 19 1 1 1	14 0 36 1 1 1	21 0 16 1 1 0		38 0 20 1 0 1
1 0 8 0 1 0	8 0 27 1 1 1		21 1 20 1 1 0	30 1 3 0 0 1	38 1 27 0 0 1
1 1 11 0 1 0	8 1 34 1 1 1	15 1 1 0 0 1	21 1 24 1 1 0	30 1 6 0 0 1	38 1 34 0 0 1
1 1 18 1 1 0	8 1 59 0 1 1	15 1 16 0 0 1	21 1 27 0 1 0	30 1 14 0 0 0	
1 0 24 1 1 0	8 1 85 0 1 0	15 1 22 1 0 1	21 0 31 0 1 0	30 0 21 1 0 1	39 1 0 0 1 0
	8 1 91 0 1 1	15 1 35 1 0 1		30 0 31 1 0 1	39 1 3 0 1 0
2 1 0 0 0 1	8 1 102 0 1 1	15 1 37 1 0 1	22 1 0 0 0 1	30 0 33 0 0 1	39 1 10 0 1 0
2 0 13 0 0 1	8 1 123 0 1 1	15 1 44 0 0 1	22 0 33 0 0 1	30 0 42 1 0 1	39 0 21 0 1 0
2 1 21 1 0 1	8 1 133 0 1 1	15 1 51 1 0 1	22 1 41 0 0 1	30 0 56 1 0 1	
2 1 29 0 0 1	8 1 140 0 1 1	15 0 58 1 0 1			40 1 1 0 0 1
2 0 40 1 0 1			23 0 1 0 1 0	31 1 0 0 0 1	40 0 8 0 0 1
	9 1 0 0 0 1	16 1 0 0 0 1	23 0 19 1 1 0	31 1 16 0 0 1	40 0 11 0 0 1
3 1 1 0 0 1	9 1 3 0 0 1	16 1 1 0 0 1	23 1 29 1 1 0	31 1 23 1 0 1	40 0 15 1 0 1
3 1 16 1 0 1	9 1 3 0 0 1	16 1 5 0 0 1	23 0 36 1 1 0		40 0 21 1 0 1
3 1 23 1 0 1	9 1 12 0 0 0	16 0 16 1 0 1		32 1 3 0 0 1	40 0 29 1 0 1
	9 1 19 1 0 0	16 1 19 1 0 1	24 0 1 0 1 0	32 1 17 1 0 1	
4 1 0 0 1 0	9 1 27 1 0 1	16 1 21 1 0 1	24 1 16 0 1 0	32 1 17 1 0 1	41 0 1 0 1 0
4 1 1 0 1 0	9 1 35 1 0 1	16 0 26 1 0 1	24 1 23 1 1 0	32 1 31 1 0 1	41 0 2 0 1 0
4 0 12 0 1 0		16 1 35 1 0 1	24 1 29 1 1 0	32 1 31 1 0 1	41 0 11 1 1 1
4 0 19 1 1 0	10 1 0 0 0 1	16 0 49 1 0 1	24 1 35 1 1 0	32 1 45 1 0 1	
4 0 27 0 1 0	10 0 1 0 0 1		24 1 44 1 1 0	32 1 45 1 0 1	42 1 0 0 1 0
	10 0 16 1 0 1	17 1 1 0 0 1	24 1 49 1 1 0		42 1 3 0 1 0
5 1 1 0 0 1		17 0 15 1 0 1		33 1 0 0 0 1	42 0 9 0 1 0
5 0 13 0 0 1	11 1 3 0 0 0	17 1 27 1 0 1	25 1 3 0 0 1	33 1 2 0 0 1	42 1 16 1 1 0
5 0 20 1 0 1	11 0 6 0 0 0	17 0 34 1 0 1	25 1 11 0 0 1	33 1 10 0 0 0	42 1 23 0 1 0
5 0 34 1 0 1	11 0 10 0 0 0	17 0 41 0 0 1	25 1 17 1 0 1	33 1 17 1 0 0	42 0 32 0 1 0
5 0 40 1 0 0	11 1 17 1 0 0	17 1 43 0 0 1		33 0 24 1 0 1	
5 0 47 0 0 0	11 1 24 1 0 0	17 1 50 0 0 1	26 0 0 0 1 0	33 1 31 0 0 1	43 1 2 0 1 0
5 0 54 1 0 0	11 1 32 0 0 1	17 1 62 0 0 1	26 0 9 0 1 0	33 1 37 1 0 0	43 1 16 0 1 0
5 0 57 1 0 0	11 0 34 0 0 1	17 0 76 0 0 1	26 0 17 0 1 0		43 1 24 1 1 0
5 0 61 1 0 0	11 1 39 1 0 1	17 0 83 0 0 1		34 1 0 0 0 1	43 1 31 1 1 0
5 0 68 1 0 1	11 1 46 1 0 1		27 1 3 0 1 0	34 1 3 0 0 1	43 1 37 1 1 0
	11 1 52 0 0 1	18 1 9 0 0 1	27 0 8 1 1 1	34 1 11 0 0 1	43 0 42 1 1 0
6 1 1 0 0 0	11 0 67 0 0 1	18 1 15 0 0 1		34 0 19 1 0 1	
6 1 9 1 0 1		18 1 22 1 0 1	28 1 0 0 1 0		44 1 0 0 0 1
6 1 29 0 0 1	12 0 0 0 0 0	18 1 29 1 0 1	28 1 13 0 1 0	35 1 0 0 0 1	44 0 5 0 0 1
6 1 36 0 0 1	12 1 27 1 0 1	18 1 43 1 0 1	28 1 21 1 1 0	35 1 7 0 0 1	44 1 12 0 0 1
6 1 49 1 0 1	12 0 33 1 0 1		28 1 28 1 1 0	35 1 12 0 0 1	44 0 19 1 0 1
6 1 56 1 0 1		19 1 7 0 0 1	28 1 35 1 1 0	35 1 27 1 0 1	44 0 28 0 0 1
	13 1 4 0 1 0	19 0 13 0 0 1	28 1 42 1 1 0		44 1 35 0 0 1
7 0 0 0 0 1	13 0 13 1 1 0	19 0 20 0 0 1	28 0 64 1 1 1	36 1 3 0 1 0	
7 1 2 0 0 1	13 0 13 1 1 0			36 1 7 0 1 0	45 1 0 0 0 1
7 0 15 0 0 0	13 0 18 1 1 0	20 1 3 0 1 0	29 1 1 0 1 0	36 1 17 1 1 1	45 1 14 1 0 1
7 0 92 0 0 0		20 0 18 1 1 1	29 1 16 1 1 1		45 1 21 1 0 1
7 0 34 0 0 0	14 1 4 0 1 0	20 1 25 1 1 1	29 1 23 1 1 1	37 1 1 0 1 0	45 0 28 1 0 1
7 0 41 1 0 0	14 1 15 0 1 0		29 1 28 0 1 0	37 1 12 1 1 1	

Variables: Patient ID, anxiety symptom, days on treatment, therapeutic Clomipramine level, TSH admission, TSH during treatment