



HAL
open science

STATISTICAL INFERENCE FOR THE MULTIDIMENSIONAL MIXED RASCH MODEL

Mohand Larbi Feddag

► **To cite this version:**

Mohand Larbi Feddag. STATISTICAL INFERENCE FOR THE MULTIDIMENSIONAL MIXED RASCH MODEL. *Communications in Statistics - Simulation and Computation*, 2008, 37 (09), pp.1732-1749. 10.1080/03610910802255832 . hal-00514327

HAL Id: hal-00514327

<https://hal.science/hal-00514327>

Submitted on 2 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



STATISTICAL INFERENCE FOR THE MULTIDIMENSIONAL MIXED RASCH MODEL

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2007-0146.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	25-Jan-2008
Complete List of Authors:	Feddag, Mohand; University of Warwick, Statistics Department
Keywords:	Approximate AIC and BIC; Fixed effects; GEE approach;; Generalized linear, Quality of life; Random effects; Rasch model.
Abstract:	Inference in Generalized linear mixed models with multivariate random effects is often made cumbersome by the high-dimensional intractable integrals involved in the marginal likelihood. This article presents an inferential methodology based on the GEE approach. This method involves the approximations of the marginal likelihood and joint moments of the variables. It is also proposed an approximate Akaike and Bayesian information criteria based on the approximate marginal likelihood using the estimation of the parameters by the GEE approach. The different results are illustrated with a simulation study and with an analysis of real data from health-related quality of life.
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>Feddag_CIS_08.tex</p>	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



For Peer Review Only

STATISTICAL INFERENCE FOR THE MULTIDIMENSIONAL MIXED RASCH MODEL

Mohand L. Feddag

Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail address: m-l.feddag@warwick.ac.uk

ABSTRACT

Inference in Generalized linear mixed models with multivariate random effects is often made cumbersome by the high-dimensional intractable integrals involved in the marginal likelihood. This article presents an inferential methodology based on the GEE approach. This method involves the approximations of the marginal likelihood and joint moments of the variables. It is also proposed an approximate Akaike and Bayesian information criterions based on the approximate marginal likelihood using the estimation of the parameters by the GEE approach. The different results are illustrated with a simulation study and with an analysis of real data from health-related quality of life.

Key words: Approximate AIC and BIC; Fixed effects; GEE approach; Generalized linear mixed model; HR-QoL; Longitudinal data; Random effects; Rasch model.

1. INTRODUCTION

Generalized linear mixed models (GLMMs) are extensions of generalized linear models (GLMs) which accommodate correlated and overdispersed data by adding random effects to the linear predictor. Their broad applications are useful in various disciplines, such as the analysis of clustered data including longitudinal data or repeated measures. Such generalized linear mixed models are also increasingly used in various fields where subjective variables (called latent traits) need to be measured using questionnaires with **polytomous** items. This is usual in health sciences and clinical trials, where those subjective variables might be for **example depressive symptoms or quality of life**. In such fields, the Rasch model, the most popular IRT (Item Response Theory) model (Fischer and Molenaar [14], and Baker and Kim [4]) is very often used. We consider the Rasch model with **random person effects**.

1 This sometimes called the stochastic subject view. We call it the mixed Rasch model and is
2 different from the mixture Rasch models described by Rost and Davier (Fischer and Molenaar
3 [14], chapter 14). The fixed effects parameters and the random effects of this model which
4 belongs to the family of logistic linear mixed models (Diggle et al. [8], and De Boeck and
5 Wilson [9]), are called the difficulty or item parameters and latent traits or person parameters
6 respectively (Fischer and Molenaar [14], p. 39-51). The estimation of the fixed effects and
7 the variance components is needed in this type of model.
8

9
10 In the case of quality of life in clinical trials, the appearance of repeated measures arises
11 when the same set of items is presented to a sample of persons at different time points/under
12 different conditions. More complex mixed Rasch models have received attention in the
13 literature. Multidimensional mixed Rasch models with item parameters considered equal
14 across time points are analyzed by Embretson [10], Hoijtink [17], Adams et al. [1] and [2]
15 and Rijmen [25]. A longitudinal IRT model with item parameters varying over time has
16 been proposed by te Marvelde et al. [27], where the main goal of the authors was to test if
17 the change in these parameters was or not significant.
18

19
20 Our work is motivated by the real data in quality of life, which are from the emotional
21 behavior subscale of the Sickness Impact Profile (SIP) questionnaire. This subscale is admin-
22 istered to the same patients at three different time points. The interest is in the study of the
23 variation of the item difficulties, the quality of life in time and the possible correlation be-
24 tween the three latent traits. The responses of our model are two-way correlated. Firstly, at
25 a given point of time, the binary responses of a single individual are correlated and secondly,
26 when repeated, they become also longitudinally correlated. To model both correlations, we
27 have used a multidimensional mixed Rasch model with item parameters assumed to vary
28 across time points. This is an extension of the one considered by Feddag and Mesbah [12].
29

30
31 A problem inherent in GLMMs is that the marginal likelihood function obtained after
32 integrating over the multidimensional random effects always involves intractable integrals.
33 Therefore, several approximation methods have been proposed in the literature. One ap-
34 proach consists of the approximation of the integral using either numerical integration tech-
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

niques (Rigdon and Tusutakawa [24], Hedeker and Gibbons [16], Pinheiro and Bates [21]), or Markov chain Monte Carlo techniques (Zeger and Karim [32], Gilks and Wild [15] and Wakefield et al. [28]). It is known that these methods become more and more difficult to use as the number of random effects increases. As an alternative to numerical integration, two different first order Taylor series expansion methods were used to approximate the marginal likelihood and/or the joint moments. These methods are based on a series expansion of the conditional mean or the likelihood about 0 or about the current estimates of the random effects. The fixed effects parameters are generally estimated by the generalized estimating equations (denoted by GEE) approach, defined by Liang and Zeger [18] as an extension of quasi-likelihood approach [29] and the variance components by the restricted maximum likelihood (REML) (Anderson and Bancroft [3]). These approaches include the method of Zeger et al. [31], the penalized quasi-likelihood and the marginal quasi-likelihood methods defined by Breslow and Clayton [6] and Breslow and Lin [7].

The application of GEE to GLMMs requires the first and second order marginal moments of the responses to estimate the fixed effects; the third and fourth joint moments of the responses are required to estimate the variance components. All these joint moments do not have a closed form under GLMMs and **must be approximated**. Feddag et al. [11] have proposed a GEE approach to estimate simultaneously the fixed effects and the variance of the univariate random effects in the logistic mixed models. These equations are based on the empirical pairwise covariances which are used by Prentice [22] and Prentice and Zhao [23] to estimate the correlation parameters in GLMs for binary data. The joint moments are obtained using the Sutradhar and Rao [26] approximations. This approach has been extended to the longitudinal mixed Rasch model (Feddag and Mesbah [12]) where the fixed effects parameters are assumed **invariant** in time.

This paper is thus an extension of the GEE method proposed by Feddag and Mesbah [12] for the longitudinal mixed Rasch model. The work considered here differs from the previous one in several aspects. We have generalized approximations of the marginal likelihood and joint moments proposed in the longitudinal mixed Rasch model to the multidimensional

1 mixed Rasch model, which assume that the fixed effects parameters vary across time points.
2 We also propose approximations to the Akaike and Bayesian information criterions. These
3 are based on the approximate marginal likelihood using the estimation of the parameters by
4 the GEE approach. The above criterions are used to compare the general structure of the
5 covariance matrix of the random effects of our model with two other matrices used commonly
6 in the literature. These two criterions are compared to the Wald test using the real data.
7
8
9
10
11

12 The specific outline of the paper is as follow. We present in Section 2 the studied model.
13 In Section 3 we give theoretical results of approximation of the marginal likelihood, then
14 we derive the approximate joint moments of the variables. The estimating equations for
15 the fixed effects parameters and the covariance matrix of random effects are constructed in
16 Section 4. We present thereafter some simulation results in order to illustrate our method.
17 The approximate Akaike information criterion (AIC) and approximate Bayesian information
18 criterion (BIC) are presented in Section 5. The various results are illustrated with a simula-
19 tion study and with an analysis of real data from a heath-related quality of life. A discussion
20 is presented in Section 6.
21
22
23
24
25
26
27
28
29

30 2. THE MATHEMATICAL SETTING OF THE MODEL

31 In our study, we consider binary responses of a questionnaire which is administrated to the
32 same subjects at various occasions t , $t = 1, \dots, T$. In such case, the responses are correlated
33 in two ways. Firstly, at a given time point, the binary responses of a single individual
34 are correlated and secondly, when repeated, they become also longitudinally correlated.
35 Recently, Feddag and Mesbah [12] have considered a longitudinal mixed Rasch model where
36 the fixed effects parameters are supposed to be invariant in time. The model considered in
37 this paper is a generalization of the preceding model, taking into account the variation of
38 the item parameters over time.
39
40
41
42
43
44
45
46

47 From now on, we will consider a sample of N independent $(TJ \times 1)$ random multivariate
48 binary observations $Y_i = (Y_i^1, \dots, Y_i^{T'})'$, $i = 1, \dots, N$. The vector $Y_i^t = (Y_{i1}^t, \dots, Y_{ij}^t)$ is
49 the response variable of individual i to the questionnaire at time t and Y_{ij}^t is the binary
50 variable response of individual i to item j at time t ($t = 1, \dots, T$). Let $Y = (Y_1, \dots, Y_N)$
51
52
53
54
55
56
57
58
59
60

be the vector of the variables, b_{it} the random effect associated with subject i at time t and $b_i = (b_{i1}, \dots, b_{iT})'$ be the multidimensional random effect for subject i . We denote by y a response variable to the variable Y . The multidimensional mixed Rasch model satisfies the following assumptions:

- Given the random effect $b_i, i = 1, \dots, N$, we have

$$P(Y_i^1 = y_i^1, \dots, Y_i^T = y_i^T | b_i, \beta) = \prod_{t=1}^T \prod_{j=1}^J P(Y_{ij}^t = y_{ij}^t | b_{it}, \beta_j^t), \quad (1)$$

where β_j^t and b_{it} are the fixed effects parameter associated to item j and the random effects associated to the subject i at time t , respectively.

- For all $i, j, t; i = 1, \dots, N, j = 1, \dots, J, t = 1, \dots, T$, the probability distribution of the random variable Y_{ij}^t is given by

$$P(Y_{ij}^t = y_{ij}^t | b_{it}, \beta_j^t) = \frac{\exp\{(b_{it} - \beta_j^t)y_{ij}^t\}}{1 + \exp(b_{it} - \beta_j^t)}. \quad (2)$$

- The random effects b_1, \dots, b_N , are independent and identically normally distributed with mean vector $\mu = (\mu_t)_{t=1, \dots, T}$ and covariance matrix $\Sigma = (\sigma_{jl})_{j, l=1, \dots, T}$.

This mixed Rasch model is a particular case of the GLMMs (see McCullagh and Nelder [19]) with the link and variance functions respectively defined by

$$h(\pi) = \ln(\pi/(1 - \pi)), \quad v(\pi) = \pi(1 - \pi).$$

It is clear that the model as formulated above is not identifiable, so some suitable restrictions will have to be imposed. If the mean μ is set to be 0, the model is identifiable without any constraints on the item parameters. Elsewhere, it is usual for each time point t , to impose the constraint $\sum_{j=1}^J (\mu_t + \beta_j^t) = 0$. This restriction leads to the relation between the mean of the random effects and the difficulty parameters which is given by $\mu_t = -\sum_{j=1}^J \beta_j^t / J$.

From now on, we suppose the mean $\mu = 0$, and we are interested in estimating the fixed effects parameters $\beta^t = (\beta_1^t, \dots, \beta_J^t)'$, $t = 1, \dots, T$, the vector of the variances of the random effects

$\alpha = (\sigma_{11}, \sigma_{22}, \dots, \sigma_{TT})'$ and the vector of their covariances $\gamma = (\sigma_{12}, \dots, \sigma_{1T}, \sigma_{23}, \dots, \sigma_{T-1,T})'$. Let $\beta = (\beta^1, \dots, \beta^T)$ be the fixed effects parameter associated to this model. The marginal likelihood of y is then given by

$$L(\beta, \alpha, \gamma | y) = \prod_{i=1}^N \int_{\mathbb{R}^T} \prod_{t=1}^T \prod_{j=1}^J \frac{\exp\{(b_{it} - \beta_j^t)y_{ij}^t\}}{1 + \exp(b_{it} - \beta_j^t)} \varphi(b_i, \alpha, \gamma) db_i, \quad (3)$$

where

$$\varphi(b_i, \alpha, \gamma) = \frac{1}{(2\pi)^{T/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} b_i' \Sigma^{-1} b_i\right\},$$

is the multivariate normal distribution with mean vector 0 and covariance matrix $\Sigma(\alpha, \gamma)$.

The maximization of this marginal likelihood is computationally difficult and needs numerical evaluation of high dimensional integrals.

3. APPROXIMATIONS OF MARGINAL LIKELIHOOD AND JOINT MOMENTS

The results of this section are very similar to those given in Feddag and Mesbah [12]. The variation of the fixed effects parameters β^t across the time $t = 1, \dots, T$, leads to different approximations given in Feddag and Mesbah [12]. We shall first give the approximation of the marginal likelihood given by Eq. (3), then we derive the joint moments up to order four of the observed variable y_i which are required later on, in the estimating equations.

As in Feddag and Mesbah [12], we assume that the joint moments from order six of the random effects b_i satisfies the following condition:

$$E(\|b_i\|^r) = o(f_r(\alpha, \gamma)) \quad \text{for all } r \geq 6, \quad (4)$$

where f_r is a function of r and of the parameters (α, γ) .

Proposition 1 *Under assumption (4), the marginal likelihood $L(\beta, \sigma^2 | y)$ defined by Eq. (3), is given by*

$$L(\beta, \alpha, \gamma | y) = L^*(\beta, \alpha, \gamma | y) + o(f_4(\alpha, \gamma)), \quad (5)$$

where

$$L^*(\beta, \alpha, \gamma | y) = \prod_{i=1}^N (1 + P_T(y_i, \beta, \alpha, \gamma)) \prod_{t=1}^T \prod_{j=1}^J g_{ij}^*(y_{ij}^t; \beta_j^t), \quad (6)$$

and

$$P_T(y_i, \beta, \alpha, \gamma) = \int_{\mathbb{R}^T} \sum_{t=1}^T \sum_{\substack{1 \leq t_1 < t_2 < \dots < t_k \leq T \\ \sum_j t_j = t}} \left[\prod_{l=1}^k h_i^{t_l}(y_i, b_i, \beta) \right] \varphi(b_i, \alpha, \gamma) db_i. \tag{7}$$

The quantities involved in the Eqs. (6) and (7) are defined below:

$$\begin{aligned} h_i^t(y_i, b_i, \beta) &= A_i^t b_{it} + \frac{1}{2} R_i^t b_{it}^2 + \frac{1}{6} P_i^t b_{it}^3 + \frac{1}{24} Q_i^t b_{it}^4, \\ g_{ij}^*(y_{ij}^t; \beta_j^t) &= \exp \left\{ -y_{ij}^t \beta_j^t - \ln(1 + e^{-\beta_j^t}) \right\}, \quad a_{i,tj} = \ln \{ 1 + \exp(b_{it} - \beta_j^t) \}, \\ A_i^t &= \sum_{j=1}^J A_{i,j}^t, \quad B^t = \sum_{j=1}^J B_j^t, \quad C^t = \sum_{j=1}^J C_j^t, \quad F^t = \sum_{j=1}^J F_j^t, \\ A_{i,j}^t &= y_{ij}^t - a_{tj}^{(1)}, \quad B_j^t = a_{tj}^{(2)}, \quad C_j^t = a_{tj}^{(3)}, \quad F_j^t = a_{tj}^{(4)}, \\ R_{i,j}^t &= (A_{i,j}^t)^2 - B_j^t, \quad P_{i,j}^t = (A_{i,j}^t)^3 - 3A_{i,j}^t B_j^t - C_j^t, \\ Q_{i,j}^t &= (A_{i,j}^t)^4 - 6(A_{i,j}^t)^2 B_j^t - 4A_{i,j}^t C_j^t + 3(B_j^t)^2 - F_j^t, \end{aligned}$$

and $a_{ij}^{(p)}$ is the p -order derivative of $a_{i,tj}$ at $b_{it} = 0$, $p = 1, \dots, 4$.

The proof of this proposition is omitted since it is similar to the one given in Theorem 1 of Feddag and Mesbah [12].

The likelihood approximation given by Eq. (5) is of order $o(f_4(\alpha, \gamma))$. The same order of error is involved in the maximum of the approximate likelihood estimators and in the following approximations.

We use the above proposition to derive the joint density within and between times which is the subject of the following corollary.

Corollary 1 Under condition (4), we have the following assertions:

1. The joint density of $Y_{i,m}^t = (Y_{ij_1}^t, \dots, Y_{ij_m}^t)$, $i = 1, \dots, N$, $m = 1, \dots, J$, $t = 1, \dots, T$ is approximated by

$$L_{i,(m)}^*(y_{i,m}^t; \beta, \alpha, \gamma) = \prod_{l=1}^m g_{ij_l}^*(y_{ij_l}^t; \beta_{j_l}^t) \left[1 + \frac{1}{2} \sigma_{tt} R_{i,j_1 \dots j_m}^t + \frac{1}{8} \sigma_{tt}^2 Q_{i,j_1 \dots j_m}^t \right]. \tag{8}$$

2. The density of $(Y_{i,m}^t, Y_{i,s}^h) = (Y_{ij_1}^t, \dots, Y_{ij_m}^t, Y_{ij_1}^h, \dots, Y_{ij_s}^h)$, $i = 1, \dots, N$, $m, s = 1, \dots, J$, $1 < t \neq h < T$, is approximated by

$$\begin{aligned}
 L_{i,(m,s)}^* (y_{i,m}^t, y_{i,s}^h; \beta, \alpha, \gamma) &= \prod_{l=1}^m g_{ij_l}^*(y_{ij_l}^t; \beta_{j_l}^t) \prod_{l=1}^s g_{ij_l}^*(y_{ij_l}^h; \beta_{j_l}^h) \\
 &\times \left[1 + \frac{1}{2} (\sigma_{tt} R_{i,j_1 \dots j_m}^t + \sigma_{hh} R_{i,j_1 \dots j_s}^h) + \sigma_{th} A_{i,j_1 \dots j_m}^t A_{i,j_1 \dots j_s}^h \right. \\
 &+ \frac{1}{2} (A_{i,j_1 \dots j_m}^t P_{i,j_1 \dots j_s}^h \sigma_{th} \sigma_{hh} + P_{i,j_1 \dots j_m}^t A_{i,j_1 \dots j_s}^h \sigma_{tt} \sigma_{th}) \\
 &\left. + \frac{1}{4} R_{i,j_1 \dots j_m}^t R_{i,j_1 \dots j_s}^h (\sigma_{tt} \sigma_{hh} + 2\sigma_{th}^2) + \frac{1}{8} (\sigma_{tt}^2 Q_{i,j_1 \dots j_m}^t + \sigma_{hh}^2 Q_{i,j_1 \dots j_s}^h) \right].
 \end{aligned} \tag{9}$$

The quantities involved in Eqs. (8) and (9) are defined in Proposition 1 and below:

$$\begin{aligned}
 A_{i,j_1 \dots j_m}^t &= \sum_{l=1}^{j_m} A_{i,j_l}^t, \quad B_{j_1 \dots j_m}^t = \sum_{l=1}^{j_m} B_{j_l}^t, \quad C_{j_1 \dots j_m}^t = \sum_{l=1}^{j_m} C_{j_l}^t, \quad F_{j_1 \dots j_m}^t = \sum_{l=1}^{j_m} F_{j_l}^t, \\
 R_{i,j_1 \dots j_m}^t &= (A_{i,j_1 \dots j_m}^t)^2 - B_{j_1 \dots j_m}^t, \quad P_{j_1 \dots j_m}^t = (A_{j_1 \dots j_m}^t)^3 - 3A_{j_1 \dots j_m}^t B_{j_1 \dots j_m}^t - C_{j_1 \dots j_m}^t, \\
 Q_{i,j_1 \dots j_m}^t &= (A_{i,j_1 \dots j_m}^t)^4 - 6(A_{i,j_1 \dots j_m}^t)^2 B_{j_1 \dots j_m}^t - 4A_{i,j_1 \dots j_m}^t C_{j_1 \dots j_m}^t + 3(B_{j_1 \dots j_m}^t)^2 - F_{j_1 \dots j_m}^t.
 \end{aligned}$$

The proof of this corollary is similar to the one given in Theorem 2 of Feddag and Mesbah [12].

This Corollary 1 is used to derive the joint moments up to order two which are given as follows.

Corollary 2 Under assumption (4), we have the approximations of the joint moments which are given by:

1. The marginal mean of Y_{ij}^t is given by

$$\begin{aligned}
 E(Y_{ij}^t) = \mu_j^t &= \frac{1}{1 + e^{\beta_j^t}} + \frac{\sigma_{tt} e^{\beta_j^t} (e^{\beta_j^t} - 1)}{2 (1 + e^{\beta_j^t})^3} \\
 &+ \frac{\sigma_{tt}^2 e^{\beta_j^t} (e^{3\beta_j^t} - 11e^{2\beta_j^t} + 11e^{\beta_j^t} - 1)}{8 (1 + e^{\beta_j^t})^5}.
 \end{aligned} \tag{10}$$

2. The covariance (within time) of (Y_{ij}^t, Y_{ik}^t) is given by

$$\begin{aligned} \text{Cov}(Y_{ij}^t, Y_{ik}^t) &= \frac{\sigma_{tt}}{2} \frac{1}{(1 + e^{\beta_j^t})^2 (1 + e^{\beta_k^t})^2} \times \left[2e^{\beta_j^t} e^{\beta_k^t} + \right. \\ &\quad \left. \sigma_{tt} \frac{(e^{3\beta_j^t} - 4e^{2\beta_j^t} + e^{\beta_j^t})e^{\beta_k^t}}{(1 + e^{\beta_j^t})^2} + \sigma_{tt} \frac{e^{\beta_j^t}(e^{\beta_j^t} - 1)e^{\beta_k^t}(e^{\beta_k^t} - 1)}{(1 + e^{\beta_j^t})(1 + e^{\beta_k^t})} \right. \\ &\quad \left. + \sigma_{tt} \frac{e^{\beta_j^t}(e^{3\beta_k^t} - 4e^{2\beta_k^t} + e^{\beta_k^t})}{(1 + e^{\beta_k^t})^2} \right]. \end{aligned} \tag{11}$$

3. The covariance (between times) of (Y_{ij}^t, Y_{ik}^h) is given by

$$\begin{aligned} \text{Cov}(Y_{ij}^t, Y_{ik}^h) &= \frac{\sigma_{th}}{2} \frac{1}{(1 + e^{\beta_j^t})^2 (1 + e^{\beta_k^h})^2} \times \left[2e^{\beta_j^t} e^{\beta_k^h} + \right. \\ &\quad \left. \sigma_{th} \frac{(e^{3\beta_j^t} - 4e^{2\beta_j^t} + e^{\beta_j^t})e^{\beta_k^h}}{(1 + e^{\beta_j^t})^2} + \sigma_{th} \frac{e^{\beta_j^t}(e^{\beta_j^t} - 1)e^{\beta_k^h}(e^{\beta_k^h} - 1)}{(1 + e^{\beta_j^t})(1 + e^{\beta_k^h})} \right. \\ &\quad \left. + \sigma_{hh} \frac{e^{\beta_j^t}(e^{3\beta_k^h} - 4e^{2\beta_k^h} + e^{\beta_k^h})}{(1 + e^{\beta_k^h})^2} \right]. \end{aligned} \tag{12}$$

4. ESTIMATION OF THE PARAMETERS

4.1. GENERALIZED ESTIMATING EQUATIONS

The estimating equations of the parameters of this model differ from those of Feddag and Mesbah [12] only on the level of the dimension of the parameter β . In this particular case, we note that this dimension is equal to TJ , whilst in the classical longitudinal case it is equal to J .

Consider $y_i = (y_i^1, \dots, y_i^T)'$, $i = 1, \dots, N$, to be the outcomes of the model defined in Section 2. The estimating equations for the parameters (β, α, γ) are based on the observations y_i , $i = 1, \dots, N$ and on their empirical covariances within and between times.

These empirical covariances are respectively given by the vectors $S'_i = (S_i^t)_{t=1, \dots, T}$ and $W'_i = (W_i^{lh})_{1 \leq l < h \leq T}$ where S_i^t is a $\frac{J(J-1)}{2} \times 1$ vector of empirical pairwise covariances (within times) associated with y_i^t defined by

$$S_i^t = (S_{i,jk}^t)_{1 \leq j < k \leq J} \quad \text{where} \quad S_{i,jk}^t = (y_{ij}^t - \mu_j^t)(y_{ik}^t - \mu_k^t), \tag{13}$$

and W_i^{lh} is a $J^2 \times 1$ vector of empirical pairwise covariances associated with (y_i^l, y_i^h) defined by

$$W_i^{lh} = (W_{i,km}^{lh})_{1 \leq k, m \leq J} \quad \text{where} \quad W_{i,km}^{lh} = (y_{ik}^l - \mu_k^l)(y_{im}^h - \mu_m^h). \quad (14)$$

The mean τ and the covariance matrix V of the vectors $\xi_i = (y_i^l, S_i', W_i')'$, $i = 1, \dots, N$, are given by

$$\tau = (\mu, \eta, \nu) \quad \text{and} \quad V = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix},$$

respectively, where

$$\begin{aligned} \mu &= E(Y_i) = (\mu^t)_{t=1, \dots, T}, \quad \eta = E(S_i) = (\eta^t)_{t=1, \dots, T}, \quad \nu = E(W_i) = (\nu^{lh})_{1 \leq l < h \leq T}, \\ \mu^t &= E(Y_i^t) = (\mu_j^t)_{j=1, \dots, J}, \quad \eta^t = (\eta_{jk}^t)_{1 \leq j < k \leq J}, \quad \nu^{lh} = (\nu_{jk}^{lh})_{\substack{1 \leq j \leq J \\ 1 \leq k \leq J}}, \\ V_{11} &= \text{Var}(Y_i), \quad V_{22} = \text{V}(S_i), \quad V_{33} = \text{Var}(W_i), \\ V_{12} &= \text{Cov}(Y_i, S_i), \quad V_{13} = \text{Cov}(Y_i, W_i), \quad V_{23} = \text{Cov}(S_i, W_i). \end{aligned}$$

All above quantities are computed using the marginal joint moments up to order four of the variables. The quantities μ_j^l , η_{jk}^l , ν_{jk}^{lh} are given by Eqs. (10), (11) and (12) respectively. The coefficients of the matrix V are computed using the joint moments of Y_i up to order four. The elements of V_{11} are obtained from Eqs. (11) and (12). The computation of V_{12} , V_{13} , V_{21} and V_{31} requires the joint moments up to order three and the other matrices require the joint moments up to order four.

The parameters (β, α, γ) will be estimated using the solution of the following generalized estimating equations

$$U(\beta, \alpha, \gamma) = D'V^{-1} \sum_{i=1}^N (\xi_i - \tau) = 0_{TJ + \frac{T(T+1)}{2}}, \quad (15)$$

where

$$D = \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix},$$

with

$$D_{11} = \frac{\partial \mu}{\partial \beta}, \quad D_{12} = \frac{\partial \mu}{\partial \alpha}, \quad D_{13} = \frac{\partial \mu}{\partial \gamma}, \quad D_{21} = \frac{\partial \eta}{\partial \beta},$$

$$D_{22} = \frac{\partial \eta}{\partial \alpha}, \quad D_{23} = \frac{\partial \eta}{\partial \gamma}, \quad D_{31} = \frac{\partial \nu}{\partial \beta}, \quad D_{32} = \frac{\partial \nu}{\partial \alpha}, \quad D_{33} = \frac{\partial \nu}{\partial \gamma},$$

and 0_p is the vector of zeros of order p . These derivatives are easily computed using Eqs. (10), (11) and (12).

Let $(\hat{\beta}, \hat{\alpha}, \hat{\gamma})$ be the root of the estimating equations defined by Eq. (15). Following a similar reasoning as in [12], it is easily seen that $N^{1/2} \left\{ (\hat{\beta} - \beta)', (\hat{\alpha} - \alpha)', (\hat{\gamma} - \gamma)' \right\}'$ is asymptotically multivariate normal with zero mean and covariance W . This covariance may be consistently estimated by

$$\hat{W} = \lim_{N \rightarrow \infty} N \left(\hat{A}_1^{-1} \hat{A}_2 \hat{A}_1^{-1} \right), \quad (16)$$

where

$$A_1 = N (D'V^{-1}D), \quad A_2 = D'V^{-1} \left(\sum_{i=1}^N (\xi_i - \tau)(\xi_i - \tau)' \right) V^{-1}D,$$

and \hat{A}_1, \hat{A}_2 are the values of A_1 and A_2 at $(\hat{\beta}, \hat{\alpha}, \hat{\gamma})$ respectively. The computation of $(\hat{\beta}, \hat{\alpha}, \hat{\gamma})$ is achieved by the Fisher-scoring algorithm.

4.2. SIMULATION STUDY

We focus here on the sample performance for the estimation of the fixed effects parameters β and the variance components α and γ by the GEE approach. We have considered two sets of the items difficulties: the first is arbitrary and the second is related to the parameter estimates given by this approach to the real data which is presented in the next section. We note that these two sets of parameters are changing over time.

We performed 500 replications of the mixed Rasch model where the parameters are fixed as follows:

- $T = 3$
- two sample sizes $N = 100, 300$ and two covariance matrices defined by:

$$\Sigma_1 = \begin{pmatrix} 0.3 & 0.1 & 0.2 \\ 0.1 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.7 & 0.4 & 0.5 \\ 0.4 & 0.8 & 0.6 \\ 0.5 & 0.6 & 0.9 \end{pmatrix}$$

- two different sets of difficulty parameters. The first one with $J = 5$, $\beta^1 = (-2.5, -1.5, -1, 0, 0.5)$, $\beta^2 = (-2.5, -1, 0.5, 1, 2.5)$ and $\beta^3 = (1.5, -0.5, 0.5, 1, 2.5)$.

The second set with 6 items is related to the parameter estimates provided by this approach on the real data, which is given by: $\beta^1 = (-1.09, -0.53, -0.32, -1.01, 0.66, -2.71)$, $\beta^2 = (1.79, 1.35, 1.05, 0.46, 1.43, 1.32)$ and $\beta^3 = (2.47, 2.54, 1.88, 1.18, 2.05, 2.58)$.

We denote by $\alpha_1 = (0.3, 0.4, 0.5)$ and $\gamma_1 = (0.1, 0.2, 0.3)$ the parameters of Σ_1 and by $\alpha_2 = (0.7, 0.8, .9)$ and $\gamma_2 = (0.4, 0.5, 0.6)$ the parameters of Σ_2 . The software package Splus is used for the simulations.

The simulation results for the classical mixed Rasch model (Feddag et al. [11]) show that the specification of the third and fourth joint moments of the variables do not improve the estimates of different parameters. As in [12], all elements containing 3rd and 4th order joint moments of the matrix V are set to equal zero. The different matrices V_{jl} , $j, l = 1, 2, 3$, are defined as follows:

- V_{11} a matrix whose elements are obtained using Eqs. (10), (11) and (12).
- V_{12} a matrix with the property that $Cov(Y_{ij}^h, S_{i,kl}^h) = Cov(Y_{ij}^h, S_{i,kl}^t) = 0$ for $j \neq k, l$ and $h \neq t$.
- V_{13} , where $Cov(Y_{ij}^t, W_{i,kl}^{sh}) = 0$, ($t \neq s \neq h$), $Cov(Y_{ij}^t, W_{i,kl}^{th}) = Cov(Y_{ij}^t, W_{i,kl}^{ht}) = 0$, ($j \neq k, l$).
- V_{22} , V_{33} are diagonal and $V_{23} = 0$.

Tables 1,2,3 and 4 display the mean value and the standard deviation for each the above cases, based on 500 data sets. From Tables 1 and 2, which correspond to the first set of difficulty parameters, we note that all the estimates present a small bias. This bias tends

Table 1: Mean values and standard deviation in parentheses of the estimates for the parameters (denoted by p) $\beta^1 = (-2.5, -1.5, -1, 0, 0.5)$, $\beta^2 = (-2.5, -1, 0.5, 1, 2.5)$, $\beta^3 = (1.5, -0.5, 0.5, 1, 2.5)$, $\alpha_1 = (0.3, 0.4, 0.5)$, $\gamma_1 = (0.1, 0.2, 0.3)$.

N	p						
100	β^1	-2.564 (.416)	-1.544 (.304)	-1.028 (.250)	0.005 (.219)	0.499 (.228)	
	β^2	-2.604 (.413)	-1.004 (.257)	0.504 (.238)	1.045 (.254)	2.573 (.408)	
	β^3	1.523 (.285)	-0.506 (.231)	0.507 (.227)	0.989 (.239)	2.579 (.415)	
	α_1	0.307 (.270)	0.463 (.328)	0.518 (.301)			
	γ_1	0.102 (.168)	0.207 (.167)	0.315 (.192)			
300	β^1	-2.519 (.231)	-1.506 (.155)	-0.994 (.139)	-0.002 (.124)	0.497 (.116)	
	β^2	-2.526 (.212)	-1.000 (.147)	0.488 (.133)	0.999 (.142)	2.513 (.230)	
	β^3	1.508 (.159)	-0.495 (.125)	0.496 (.130)	0.999 (.144)	2.535 (.215)	
	α_1	0.315 (.147)	0.409 (.181)	0.493 (.170)			
	γ_1	0.101 (.101)	0.202 (.096)	0.302 (.106)			

Table 2: Mean values and standard deviation in parentheses of the estimates for the parameters (denoted by p) $\beta^1 = (-2.5, -1.5, -1, 0, 0.5)$, $\beta^2 = (-2.5, -1, 0.5, 1, 2.5)$, $\beta^3 = (1.5, -0.5, 0.5, 1, 2.5)$, $\alpha_2 = (0.7, 0.8, 0.9)$, $\gamma_2 = (0.4, 0.5, 0.6)$.

N	p										
100	β^1	-2.549	(.393)	-1.527	(.323)	-1.006	(.259)	0.000	(.221)	0.494	(.240)
	β^2	-2.561	(.374)	-0.991	(.257)	0.492	(.244)	1.000	(.261)	2.550	(.420)
	β^3	1.508	(.310)	-0.503	(.243)	0.494	(.235)	0.964	(.243)	2.530	(.390)
	α_2	0.706	(.340)	0.788	(.369)	0.835	(.333)				
	γ_2	0.405	(.217)	0.529	(.239)	0.598	(.242)				
300	β^1	-2.518	(.236)	-1.493	(.164)	-0.980	(.145)	0.002	(.136)	0.499	(.131)
	β^2	-2.515	(.212)	-0.986	(.150)	0.481	(.139)	0.985	(.146)	2.516	(.236)
	β^3	1.486	(.161)	-0.482	(.135)	0.486	(.134)	0.979	(.153)	2.526	(.215)
	α_2	0.699	(.193)	0.791	(.215)	0.846	(.195)				
	γ_2	0.407	(.128)	0.508	(.127)	0.599	(.130)				

to be negligible when the size increases. As expected, we also note that the standard errors for almost all the estimates decrease when N increases. For the second set of the difficulty parameters, the results displayed in Tables 3 and 4 are very similar to the first case. We observe small bias for both difficulty parameters and variance component estimates. We note as well an improvement in the estimates for the size $N = 300$. This simulation study shows that this approach performs better for large values of N .

In terms of running time (s denotes second), this approach is very fast. For the first example, in average, for each data set, it takes 0.74 s to converge with $N = 100$ and 1.04 s with $N = 300$. For the second case, the convergence is reached in 0.92 s with $N = 100$ and 1.15 s with $N = 300$.

5. APPROXIMATE AKAIKE AND BAYESIAN INFORMATION CRITERIONS

It may be of interest to compare the general structure of the variance of the random effects considered in the model with alternative structures, using selection model **criteria** and tests. The well known AIC and BIC criteria based on the marginal likelihood appear to be complicated since the maximum of the approximate likelihood for this model is prone to numerical problems. This is not the case for the univariate random effects, where the approximate likelihood method is **comparable** to the GEE approach (see Feddag and Mesbah [13]). **In the following, we propose two alternative criteria, namely the approximate AIC and the approximate BIC.** These two criteria based on the approximate marginal likelihood using the estimation of the parameters by the GEE approach, will be illustrated by a simulation study and compared to the Wald test using real data.

5.1. DEFINITIONS

Definition 1 Let $\ell^* = \ln L^*(\hat{\beta}, \hat{\alpha}, \hat{\gamma} | y)$ the logarithm of the approximate likelihood given by the expression (6) where $(\hat{\beta}, \hat{\alpha}, \hat{\gamma})$ are the estimates of the parameters (β, α, γ) by the GEE method. The Akaike Information Criterion (denoted by AIC) is defined as follows:

$$AIC = -2\ell^* + 2p,$$

Table 3: Mean values and standard deviation in parentheses of the estimates for the parameters $\beta^1 = (-1.09, -0.53, -0.32, -1.01, 0.66, -2.71)$, $\beta^2 = (1.79, 1.35, 1.05, 0.46, 1.43, 1.32)$, $\beta^3 = (2.47, 2.54, 1.88, 1.18, 2.05, 2.58)$, $\alpha_1 = (0.3, 0.4, 0.5)$, $\gamma_1 = (0.1, 0.2, 0.3)$.

N	Parameter							
100	β^1	-1.101	-0.524	-0.327	-1.035	0.660	-2.760	
		(.250)	(.212)	(.216)	(.244)	(.230)	(.442)	
	β^2	1.792	1.378	1.055	0.456	1.439	1.341	
		(.298)	(.280)	(.246)	(.226)	(.281)	(.284)	
	β^3	2.530	2.602	1.918	1.168	2.082	2.641	
		(.405)	(.414)	(.335)	(.255)	(.355)	(.415)	
	α_1	0.301	0.399	0.493				
		(.203)	(.215)	(.304)				
	γ_1	0.103	0.203	0.294				
		(.131)	(.168)	(.168)				
	300	β^1	-1.089	-0.531	-0.323	-1.007	0.661	-2.732
			(0.141)	(.130)	(.124)	(.140)	(.132)	(.239)
β^2		1.795	1.351	1.053	0.460	1.427	1.330	
		(.164)	(.152)	(.139)	(.138)	(.152)	(.159)	
β^3		2.483	2.547	1.874	1.178	2.047	2.581	
		(.223)	(.234)	(.187)	(.149)	(.181)	(.223)	
α_1		0.313	0.392	0.483				
		(.118)	(.117)	(.176)				
γ_1		0.101	0.196	0.289				
		(.077)	(.097)	(.099)				

Table 4: Mean values and standard deviation in parentheses of the estimates for the parameters $\beta^1 = (-1.09, -0.53, -0.32, -1.01, 0.66, -2.71)$, $\beta^2 = (1.79, 1.35, 1.05, 0.46, 1.43, 1.32)$, $\beta^3 = (2.47, 2.54, 1.88, 1.18, 2.05, 2.58)$, $\alpha_2 = (0.7, 0.8, 0.9)$, $\gamma_2 = (0.4, 0.5, 0.6)$.

N	Parameter							
100	β^1	-1.091	-0.529	-0.312	-1.001	0.636	-2.767	
		(.270)	(.236)	(.233)	(.242)	(.240)	(.436)	
	β^2	1.765	1.340	1.043	0.437	1.450	1.317	
		(.300)	(.290)	(.264)	(.235)	(.278)	(.273)	
	β^3	2.480	2.597	1.907	1.144	2.061	2.609	
		(.393)	(.433)	(.331)	(.261)	(.328)	(.419)	
	α_2	0.703	0.745	0.808				
		(.274)	(.239)	(.328)				
	γ_2	0.429	0.494	0.539				
		(.188)	(.210)	(.208)				
	300	β^1	-1.077	-0.521	-0.322	-1.001	0.661	-2.749
			(.147)	(.138)	(.132)	(.139)	(.146)	(.230)
β^2		1.772	1.329	1.024	0.445	1.408	1.294	
		(.167)	(.159)	(.142)	(.132)	(.164)	(.153)	
β^3		2.437	2.541	1.856	1.151	2.035	2.564	
		(.212)	(.230)	(.192)	(.159)	(.175)	(.228)	
α_2		0.718	0.732	0.801				
		(.175)	(.142)	(.177)				
γ_2		0.407	0.489	0.546				
		(.106)	(.126)	(.109)				

the Bayesian information criterion (BIC) is defined by

$$BIC = -2\ell^* + p \ln(N),$$

where p is the number of parameters in the model and N is the number of individuals.

We will illustrate these criteria with a simulation study and with real data from quality of life. We achieve this by comparing the general structure of the covariance matrix of the random effects considered in our model with two other structures classically used in the literature. These different matrices are denoted as follows:

Type 1. The matrix Σ is completely unspecified with a general structure. It corresponds to the one considered in the model.

Type 2. Exchangeable matrix where all off-diagonal elements are supposed to be equal. In other words: $\sigma_{12} = \sigma_{13} = \sigma_{23}$, while for $j = 1, 2, 3$, σ_{jj} remains unspecified.

Type 3. Diagonal matrix defined by $\Sigma = \text{diag}\{\sigma_{jj}\}_{j=1,2,3}$.

The Wald test used to compare these three models is defined as follows.

Definition 2 let $\gamma = (\sigma_{jl})_{1 \leq j < l \leq 3} = (\gamma_1, \gamma_2)$, $\gamma_2 \in \mathbb{R}^m$ ($m = 1, 2$) be the parameter of the model. Consider the test: $H_0 : \gamma_2 = 0$ against $H_1 : \gamma_2 \neq 0$.

Under H_0 , the Wald test statistic S_m is given by

$$S_m = \hat{\gamma}_1 [\hat{W}_{\hat{\gamma}_1}/N]^{-1} \hat{\gamma}_1$$

and has an asymptotic (as $N \rightarrow \infty$) χ^2 distribution with $(3 - m)$ degrees of freedom (χ^2_{3-m}), where $\hat{W}_{\hat{\gamma}_1}$ is obtained using formula (16).

5.2. SIMULATION STUDY

A simulation study was conducted to evaluate the performance of the approximate AIC and BIC criteria and to detect differences between the three covariance structures defined above. The parameters considered in this simulation for the true model are the same as

Table 5: Approximate AIC and BIC in the simulation study.

N		True model					
		Type 1		Type 2		Type 3	
		AIC	BIC	AIC	BIC	AIC	BIC
100	Type 1	50	49	48	41	39	34
	Type 2	48	49	49	56	17	15
	Type 3	2	2	3	3	44	51
300	Type 1	50	51	48	39	29	23
	Type 2	47	46	51	60	14	11
	Type 3	3	3	1	1	57	66

those given in Table 4: two different sizes $N = 100$ and 300 , difficulty parameters related to the estimates from the real data and Σ_2 as variance components. From each of the three true models, we simulate 100 data sets and we estimate the parameters under the three different structures. Then we evaluate the AIC and BIC and for each of these criteria we select the best one (smaller value) amongst the three values. Table 5 displays for each criterion the number of times (frequency) that each structure is the best. From this table, the results are similar for the two sizes. The frequencies for each criterion under the structures Type 1 and Type 2 are very close to each other and a larger one under structure Type 3. For the Type 3 structure as the true model, it is clearly shown that the two criteria with large frequencies perform better for this structure than other two. In conclusion, the two criteria are better under the true model even there are very close under the two first structures.

5.3. REAL DATA

We illustrate in this section the application of this methodology in the analysis of the data from a sickness impact profile (SIP) questionnaire (see Bergner et al. [5]). **Initially, this subscale is composed of 470 depressive patients who answered to 9 items. For the longitudinal study with three time points, there were only 131 individuals who respond to the**

Table 6: Description of the Emotional Behavior subscale of SIP.

1. I say how bad or useless I am, for example, that I am a burden on others
2. I laugh or cry suddenly
3. I often moan and groan in pain or discomfort
4. I act nervous or restless
5. I keep rubbing or holding areas of my body that hurt or are uncomfortable
6. I talk about the future in a hopeless way

six items given in Table 6. This subscale measures the deterioration of the health-related quality of life (HR-QoL).

We estimate the various parameters using the GEE approach, then we evaluate the approximate AIC and BIC criteria and the Wald tests to compare the three structures Types 1, 2 and 3. We report in Table 7 the difficulty parameters estimates and their standard deviation and in Table 8 the variance components estimates and their standard deviation. Table 9 displays the approximate AIC and BIC values and Wald tests under the three different structures. We note that the standard deviations are obtained by the estimate of the covariance matrix given by formula (16).

From Table 7, we note that the fixed effects parameters are different between time points. For each type of model, all the fixed effects parameters grow considerably in time. We denote by $\mu_t = (\mu_t^1, \mu_t^2, \mu_t^3)$ the mean of the latent trait at time point t ($t = 1, 2, 3$), where μ_t^i is the corresponding mean under the structure i ($i = 1, 2, 3$). Then using the relation $\mu_t = -\sum_{j=1}^J \beta_j^t / J$ between the mean of the latent trait and the estimates of the difficulty parameters, we deduce the different means which are given by $\mu_1 = (0.83, 0.84, 0.86)$, $\mu_2 = (-1.24, -1.21, -1.20)$ and $\mu_3 = (-2.12, -2.13, -2.13)$. The decreasing of the mean over time denotes clearly the deterioration of the quality of life of the patients over the time period. It should be noted that at each time point, the difference of the estimates of the fixed effects parameters between the three models is negligible. However, as shown in Table 8,

1 the estimates of the variance components differ from one model to another. If the estimates
2 of the variances (diagonal elements) are rather close, the estimates of the covariances are
3 very different from one model to another. It should be noted that the estimates of these
4 parameters for the models Type 1 and 2 are different from zero. This explains the correlation
5 between the three random effects (latent traits). For the general model (without constraints
6 on the variance components) the first and the second random effects are positively correlated
7 and the correlation between the third and the two first random effects are negative.
8

9 From Table 9, we note that the values of the AIC and BIC are smaller under the general
10 model (Type 1) relative to the other models. These criteria confirm the goodness of fit of
11 the suggested model compared to the two other models commonly used in the literature. We
12 note also that the largest values of these criteria are realized under the diagonal structure
13 of the covariance matrix of the random effects (Type 3). This reinforces the conclusion
14 that correlation exists between the three random effects. Table 9 lists the three Wald tests
15 considered where Type 2 and Type 3 models are compared with each other and with Type
16 1 model. These tests with significance level 0.05 are significant and reinforce the same
17 conclusion given by the two approximate AIC and BIC criteria.
18

31 6. DISCUSSION

32 In this paper, we used the GEE approach for the multidimensional mixed Rasch models.
33 Feddag et al. [11] have proposed estimating equations to estimate simultaneously the fixed
34 effects parameters and the variance of the random effects for the logistic mixed model. The
35 paper of Feddag and Mesbah [12] illustrates how the estimating equations proposed by Fed-
36 dag et al. [11], can be extended to the longitudinal mixed Rasch model with fixed effects
37 parameters being considered as invariant in time. The approximations of the marginal like-
38 lihood and the joint moments given in this work are generalizations of those of Feddag and
39 Mesbah [12]. The simulation results confirm the theoretical asymptotic properties of the
40 estimators and show the performance of this approach. The second main result is the pro-
41 posal of the Akaike and Bayesian information criteria which are based on the approximate
42 marginal likelihood combined with the estimate of the parameters by the GEE approach.
43
44
45
46
47
48
49
50
51
52
53
54

Table 7: Parameter estimates for β and their standard deviation in parenthesis to the SIP data.

Parameter	Model						
β^1	Type 1	-1.091	-0.531	-0.323	-1.012	0.667	-2.717
		(.231)	(.203)	(.203)	(.220)	(.214)	(.332)
	Type 2	-1.101	-0.540	-0.332	-1.023	0.660	-2.730
		(.233)	(.204)	(.202)	(.221)	(.213)	(.340)
	Type 3	-1.125	-0.560	-0.354	-1.046	0.641	-2.736
		(.233)	(.205)	(.201)	(.224)	(.211)	(.345)
β^2	Type 1	1.795	1.353	1.052	0.461	1.435	1.322
		(.300)	(.260)	(.227)	(.189)	(.265)	(.250)
	Type 2	1.754	1.321	1.025	0.438	1.402	1.290
		(.293)	(.252)	(.221)	(.188)	(.257)	(.243)
	Type 3	1.740	1.313	1.018	0.430	1.395	1.278
		(.290)	(.249)	(.220)	(.188)	(.255)	(.242)
β^3	Type 1	2.473	2.545	1.888	1.188	2.053	2.588
		(.347)	(.336)	(.359)	(.298)	(.378)	(.343)
	Type 2	2.497	2.569	1.898	1.177	2.048	2.618
		(.357)	(.347)	(.375)	(.302)	(.396)	(.349)
	Type 3	2.483	2.556	1.895	1.187	2.059	2.599
		(.350)	(.339)	(.367)	(.299)	(.385)	(.347)

Table 8: Parameter estimates for (α, γ) and their standard deviation in parenthesis to the SIP data.

Model	Parameters					
	α_1	α_2	α_3	γ_1	γ_2	γ_3
Type 1	0.871 (.283)	1.472 (.183)	1.812 (.258)	0.265 (.201)	-0.379 (.174)	-0.532 (.242)
Type 2	0.881 (.288)	1.457 (.184)	1.841 (.261)	-0.186 (.088)	-0.186 (.088)	-0.186 (.088)
Type 3	0.848 (.284)	1.447 (.184)	1.822 (.260)	* *	* *	* *

Table 9: Approximate AIC-BIC and Wald test to the SIP data.

AIC and BIC				
	p	ℓ^*	AIC	BIC
Type 1	24	-858.29	1764.59	1833.60
Type 2	22	-936.32	1916.65	1979.91
Type 3	21	-968.42	1978.84	2039.22
Wald tests				
test	df	S_m	$\chi^2_{0.05}$	
Type 3 vs Type 2	1	4.461	3.84	
Type 2 vs Type 1	2	19.209	5.99	
Type 3 vs Type 1	3	8.009	7.81	

1 By simulation study and analysis of real data from the emotional behavior subscale, these
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

By simulation study and analysis of real data from the emotional behavior subscale, these criteria confirm the goodness of fit of the suggested model, compared to the two other models classically used in the literature. The Wald test for the variance components used for the real data, leads to the same conclusion as the two proposed approximate criteria. Further, we point out that Pan [20] introduced two extensions of such measures based on the quasilielihood. It would be interesting to further compare our current results with those of Pan. Finally, this work could be easily generalized to cope with scales containing several correlated subscales.

References

- [1] R. J. Adams, M. Wilson and M. Wu (1997). Multilevel Item Response Models: An Approach to Errors in Variables Regression, *Journal of Educational and Behavioral Statistics*, **22**, 47–76.
- [2] R. J. Adams, M. Wilson and W. C. Wang (1997). The Multidimensional Random Coefficients Multinomial Logit Model, *Applied Psychological Measurement*, **21**, 1–23.
- [3] N. E. Anderson and T. A. Bankroft(1952). statistical Theory in Research, *New York: Mc Graw-Hill*.
- [4] F. B. Baker and S. H. Kim (2004). Item Response Theory, Parameter Estimation Tehniques, *Marcel Dekker, Inc*.
- [5] M. Bergner et al (1973). The Sickness Impact Profile: Validation of a health status measure, *Medical care*, **14**, 56–67.
- [6] N. E. Breslow and D. G. Clayton (1993). Approximate inference in generalized linear mixed models, *J. Amer. Statist. Assoc.*, **88**, 9–25.
- [7] N. E. Breslow and X. Lin (1995). Bias correction in generalized linear models with a single component of dispersion, *Biometrika*, **82**, 81–92.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- [8] P. J. Diggle, K. Y. Liang and S. L. Zeger (1994). *Analysis of Longitudinal Data*, Oxford University Press.
- [9] P. De Boeck and M. Wilson (2004). *Explanatory Item Response Theory*, Springer.
- [10] S. E. Embretson (1991). A multidimensional latent trait model for measuring learning and change, *Psychometrika*, **56**, 491–515.
- [11] M. L. Feddag, I. Grama and M. Mesbah (2003). Generalized Estimating Equations for Mixed Logistic Models, *Communications in Statistics: Theory and Methods*, **32 N.4**, 851–874.
- [12] M. L. Feddag and M. Mesbah (2005). Generalized Estimating Equations for Longitudinal Mixed Rasch Model, *Journal of Statistical Planning and Inference*, **129**, 159–179.
- [13] M. L. Feddag and M. Mesbah (2006). Approximate estimation in generalized linear mixed models with applications to the Rasch model, *Computer and Mathematics with Applications*, **51**, 269–278.
- [14] G. H. Fischer and I. W. Molenaar (1995). *Rasch Models Foundations, recents developments and Applications*, Springer-Verlag.
- [15] W. R. Gilks and P. Wild (1992). Adaptive Rejection Sampling for Gibbs Sampling, *Applied Statistics*, **4**, 337–348.
- [16] D. Hedeker and R. D. Gibbons (1994). A Random Effects Ordinal Regression Model for Multilevel Analysis, *Biometrics*, **50**, 933–944.
- [17] H. Hoijsink (1995). Linear and Repeated Measures Models for the Person Parameter, In *Rasch Models Foundations, recents developments and Applications*, 203–214 Springer-Verlag.
- [18] K. Y. Liang and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

- 1 [19] P. McCullagh and J. A. Nelder (1989). Generalized Linear Models (Second edition),
2
3 *London: Chapman and Hall.*
4
- 5 [20] W. Pan (2001). Akaike's information criterion Bias in generalized estimating equations,
6
7 *Biometrics*, **57**, 120–125.
8
9
- 10 [21] J. C. Pinheiro and D. M. Bates (1995). Approximation to the Log-Likelihood Function in
11
12 the Nonlinear Mixed Effects Models, *Journal of Computational and Graphical Statistics*,
13
14 **4**, 12–35.
15
- 16 [22] R. L. Prentice (1988). Correlated Binary Regression with Covariates Specific to Each
17
18 Binary Observation, *Biometrics*, **44**, 1033–1048.
19
- 20 [23] R. L. Prentice and L. P. Zhao (1991). Estimating Equation for Parameters in Means
21
22 and Covariances of Multivariate Discrete and Continuous Responses, *Biometrics*, **47**,
23
24 825-839.
25
26
- 27 [24] S. E. Rigdon and R. K. Tsutakawa (1983). Parameter estimation in latent trait models,
28
29 *Psychometrika*, **48**, 567–574.
30
31
- 32 [25] F. Rijmen, P. De Boeck and H. L. J. Van Der Maas (2005). An IRT model with
33
34 parameter-Driven process for change *Psychometrika*, **70**, 651–669.
35
36
- 37 [26] B. C. Sutradhar and R. P. Rao (2001). On Marginal Quasi-Likelihood Inference in
38
39 Generalized Linear Mixed Models, *Journal of Multivariate Analysis*, **76**, 1–34.
40
41
- 42 [27] J. M. te Marvelde, C. A. W. Glass, G. V. Landeghem and J. V. Damme (2006). Ap-
43
44 plication of Multidimensional Item Response Theory Models to Longitudinal Data, *Ed-
45
46 ucational and Psychological Measurement*, **66**, 5–34.
47
- 48 [28] J. C. Wakefield, A. F. M. Smith, A. Racine-Poon and A. E. Gelfand (1994). Bayesian
49
50 Analysis of Linear and Non-Linear Population Models bu Using the Gibbs Sampler,
51
52 *Applied Statistics*, **43**, 201–221.
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- [29] R. W. M. Wedderburn (1974). Quasi-likelihood function, generalized linear models and the Gauss-Newton method, *Biometrika*, **48**, 439–447.
- [30] L. P. Zhao and R. L. Prentice (1990). Correlated Binary Regression using a quadratic exponential model, *Biometrika*, **77**, 642-648.
- [31] S. L. Zeger, K. Y. Liang and P. S. Albert (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach, *Biometrics*, **44**, 1049–1060.
- [32] S. L. Zeger and M. R. Karim (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach, *Journal of the American Statistical Association*, **86**, 79-86.

1 Report 1: To the Associate Editor

On Statistical inference for the multidimensional mixed Rasch model by M-L. Feddag

Many thanks for your valuable and interesting comments on my submitted paper. Please note that all changes made in the new version of the manuscript are highlighted in red colour. The answers to your comments are as follows.

1. **Page 5, after formula 3: with mean vector \mathbf{u} instead of with mean vector $\mathbf{0}$.**

It is corrected.

2. **We note that this dimension is equal TJ, whilst in the longitudinal case it is equal to J. Do you mean in the classical longitudinal case?**

Right, I mean the classical longitudinal case, like the one considered by Feddag and Mesbah (2005).

3. **The simulation study (section 4.2) are maid assuming that the mean of the latent parameter is zero every time, $\mathbf{u}(t)=\mathbf{0}$ (=sum of beta(t)), while your model is interesting mainly when $\mathbf{u}(t)$ is also changing with time. Why choosing for \mathbf{u} only an unchanged zero value?**

Please note that the item difficulties of the first set considered in the simulation study, has been changed (Tables 1 and 2). Regarding to the relation between the mean μ_t and the mean of the item parameter β^t , deduced from the identifiability constraint (see the last paragraph of page 5), we can consider that the mean is changing over time in both two cases.

More details about the identifiability and the relation between the mean μ_t and the mean of the item parameter β^t are as follows :

At each time point t , let $p_{ij}^t = Pr(Y_{ij}^t = 1 | b_{it}, \beta_j^t)$. The Rasch model defined in Section 2 can be expressed as

$$\text{logit}(p_{ij}^t) = b_{it} - \beta_j^t, \quad b_{it} \rightsquigarrow N(0, \sigma_t^2). \quad (1)$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

This model as defined by Eq.(1) does not have any identifiability problem (no constraints on the parameters β_j^t), and it is equivalent to the following one :

$$\text{logit}(p_{ij}^t) = b_{it}^* - (\mu_t + \beta_j^t), \quad b_{it}^* \sim N(\mu_t, \sigma_t^2). \quad (2)$$

To be identifiable, the model defined by Eq.(2) needs the following constraint

$$\sum_{j=1}^J (\mu_t + \beta_j^t) = 0.$$

This constraint leads to the relation between μ_t and β_j^t which is given by

$$\mu_t = -\frac{1}{J} \sum_{j=1}^J \beta_j^t.$$

From this relation, we can deduce the mean of the random effects of the studied model. That what I did for the real data in Section 5.3 (p. 20).

1 Report 2: To the first referee

On Statistical inference for the multidimensional mixed Rasch model by M-L. Feddag

Many thanks for your valuable and interesting comments on my submitted paper. Please note that all changes made in the new version of the manuscript are highlighted in red colour. The answers to your comments are as follows.

1. General comments

Some modifications in the introduction has been made, where the formulas is removed. Further, three main references on this kind of models has been added.

2. Specific comments

All the suggested changes were taken into account in the new version (highlighted in red colour). The formulae and eqn were all replaced by Eq. (for one) or Eqs. (two or more).

3. Simulation study

Please note that the item parameters in the first example of the simulation study has been changed (Tables 1-2). These parameters are changing over time, so the example is adaptable to the model considered.

1 Report 3: To the second referee

On Statistical inference for the multidimensional mixed Rasch model by M-L. Feddag

Many thanks for your valuable and interesting comments on my submitted paper. Please note that all changes made in the new version of the manuscript are highlighted in red colour. My answers to your comments are as follows.

1. **First, the author proposes a development of a longitudinal model already proposed in the reference [11]. In this new development, the author suggests to have different values for the items parameters at each passation of the questionnaire. The main problem is this one. Indeed, in Psychometrics, the parameters which characterize the items are commonly considered as constant, because there is no reason to consider that the properties of a given item change over time. This consideration is presented by the author as the main improvement of this paper compared to the reference [11], but the author don't made none justification of its choice !**

This is an interesting question. I recognize that the real data considered are not suitable in Psychometrics point of view, but the mathematical modelling which considers the variation of the items parameters, permits the deduction of the mean of the latent traits (see details to the answers for your second comments).

The main idea of this model with item parameters varying over time is its applications for certain scales with items changing over time. These scales could be related to learning or reading ability.

More, the author named its model a multidimensional Rasch model, but it can be considered only as a longitudinal IRT model. In IRT, a multidimensional model allows considering several latent traits (or a multivariate latent trait). In this paper, there is not several latent traits, but a unic latent trait measured several times, so the model is not "multidimensional". The Rasch model is a specific IRT model wich has several specific properties. One of these properties is the specific objectivity which implies that the individuals can be compared independently of the characteristics of the questionnaire (=items parameters). A consequence of this property is that the items parameters are the same for several samples

of individuals (and so, constant over time). The main references in the framework of the longitudinal IRT model always considered the items parameters as fixed in the time... In the present paper, the author don't justify at all its choice to allow the items parameters to change over time, and this point is the main drawback of these paper.

All the cited references are added in the introduction with some details on the paper of te Marvelde et al.. Even if this model seems to be problematic in psychometric point of view, it is more general in mathematical setting than the classical one. The variation over time of the item parameters, provides the estimation of the mean of the latent traits (see the following comments). Further, this model could be generalized to scales containing several correlated subscales, say Q and each one with J_q items ($q = 1, \dots, Q$). In this case, the time T will be replaced by Q and the term multidimensional as you have suggested, is more suitable. This last comment is added in the Discussion (section 6).

2. **A second point concerns the parametrization of the models.** Generally, in longitudinal model, we estimate the difference on the latent trait over time. In the mixed IRT models, it is frequent to estimate the mean of the latent trait. To allow the model to be identified, a constraint is made on the items parameters. In the paper, the author put the identifiability constraints on the mean of the latent trait at each passation of the questionnaire ($\mu_t = 0$), and the results are more difficult to interpret.

I have mentioned in the last paragraph of page 5, the relation between the mean μ_t and the mean of the item parameter β^t , deduced from the identifiability constraint. This relation highlighted below, permits the deduction of the estimate of the mean μ_t from the estimate of β^t .

The details of this relation are as follows :

At each time point t , denote by $p_{ij}^t = Pr(Y_{ij}^t = 1 | b_{it}, \beta_j^t)$. The Rasch model defined in Section 2 can be expressed as

$$\text{logit}(p_{ij}^t) = b_{it} - \beta_j^t, \quad b_{it} \rightsquigarrow N(0, \sigma_t^2). \quad (1)$$

Hence, this model given by Eq.(1) does not have any identifiability problem (no constraints on the parameters β_j^t). It is equivalent to the

following one :

$$\text{logit}(p_{ij}^t) = b_{it}^* - (\mu_t + \beta_j^t), \quad b_{it}^* \sim N(\mu_t, \sigma_t^2). \quad (2)$$

To be identifiable, the above model (Eq.(2)) needs the following constraint

$$\sum_{j=1}^J (\mu_t + \beta_j^t) = 0.$$

This constraint leads to the relation between μ_t and β_j^t which is given by

$$\mu_t = -\frac{1}{J} \sum_{j=1}^J \beta_j^t.$$

From this relation, we can deduce the mean of the random effects of the studied model. What we made for the real data in Section 5.3 (p. 20).

3. **A third point concerns the use of the GEE as method of estimation of the parameters. A short simulation study is realized and the author shows that the estimations are unbiased. Nevertheless I found this simulation very short : only two cases are treated - a first with parameters arbitrary chosen by the authors (and constant over time with a different parametrization of the model), and another based on estimation obtained in a real case. The author doesn't compare its results with the MML method which is the reference in this field. Advantages and drawbacks of the GEE are not presented : time ? bias ? asymptotic properties ? Robustness of the GEE approach when the Taylor approximation of the Likelihood is bad (for example for a large variance of the latent trait) ? This last point is very problematic in Psychometrics, because large values of the variance of the latent trait are frequent. These remarks concern too the analysis of the approximation of the AIC and BIC with the GEE approach : are the approximations of these criteria enough precise to made this approach useless in practice in a large set of situations.**

As, I have mentioned before, the item difficulties of the first example in the simulation study has changed. According to the previous answer, we can deduce the estimate of the mean of the random effects from the item parameters estimates.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For the mixed Rasch model with univariate random effects, the GEE was compared to the MML (by the use of RSP software) and to the CML (see Feddag et al., [12]). The results show the performance of the GEE approach comparatively to the two methods for a small variance components. For the multivariate case, it provides good properties (consistency and asymptotic properties) and it performs well only for a small variance components. In terms of running time, this approach is very fast. The running time for each case is added in the simulation study (end of the subsection).

Unfortunately, for a large variance components this approach is not working well and sometimes it confronted to convergence problems. Due to the Taylor approximation which is bad for large values. Regarding the large estimates of the variance of the second and third latent traits (Table 8), I suspect a small bias in the estimates for the real data. The proposed approximate AIC and BIC are useful only for a data sets with a small variation.

4. **Last, the example presented by the author is not very informative: The author uses a specific dimension of the Sickness Impact Profile (SIP) with 3 repeated measures over time. First, it is curious to use the Emotional Behaviour scale, because generally, the emotional dimensions are the less stable dimensions of the HR-Qol scales (often these dimensions are multi-dimensional ones and a poor fit to unidimensional models). More, the SIP contains 9 items in the Emotional Behaviour scale, and the author reports only 6 items (why the 3 others items have been omitted ?). Finally, the author finds different values of the items parameters over the 3 passations, and there is no justification or explanation of what it can be do of this information !! The treatment opered by the author to deduce the mean of the latent trait at each passation of the questionnaire is unjustified and certainly is illusive as soon as the items parameters are not considered as constant over time.**

In fact, for the univariate case (see Feddag et al., 2003 [11]), this subscale is composed of 470 individuals who answers to 9 items. However, for the longitudinal case, there were few patients who answered to the 9 items, so not enough information for the analysis. There were only 131 who answered to 6 items during three times.

About the deduction of the mean of the latent trait, please see the

second comments given above.

5. Other minor points

- The three errors in page 1 has been corrected and the email address appeared in pages 2 and 5 is removed.
- Simulation Study: More comments on the second set of parameters are added (lines 17-19, page 11). The first set of item parameters has been changed to another one which are varying over time.
- The missing component in β^1 is added (second example of the simulation study).
- Page 20: More additional information on the SIP subscale is added in this page with red colour.