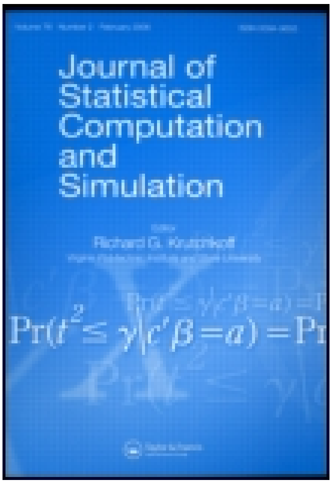

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Pairwise- and marginal-likelihood estimation for the mixed Rasch model with binary data

M.-L. Feddag [a] , J.-B. Hardouin [b] & V. Sebille [b]

[a] Southampton Statistical Sciences Research Institute , University of Southampton , Southampton , SO17 1BJ , UK

[b] EA 4275 'Biostatistique, Recherche Clinique et Mesures Subjectives en Sante', Faculté de Pharmacie , Université de Nantes , Nantes , France

Published online: 26 Sep 2011.

Journal of Statistical Computation and Simulation
Vol. 82, No. 3, March 2012, 419–430

Pairwise- and marginal-likelihood estimation for the mixed Rasch model with binary data

M.-L. Feddag[a†]*, J.-B. Hardouin[b] and V. Sebille[b]

[a] *Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, UK;* [b] *EA 4275 'Biostatistique, Recherche Clinique et Mesures Subjectives en Sante', Faculté de Pharmacie, Université de Nantes, Nantes, France*

(*Received 7 August 2008; final version received 5 November 2010*)

A marginal–pairwise-likelihood estimation approach is examined in the mixed Rasch model with the binary response and logit link. This method belonging to the broad class of composite likelihood provides estimators with desirable asymptotic properties such as consistency and asymptotic normality. We study the performance of the proposed methodology when the random effect distribution is misspecified. A simulation study was conducted to compare this approach with the maximum marginal likelihood. The different results are also illustrated with an analysis of the real data set from a quality-of-life study.

Keywords: binary data; composite likelihood; fixed effects; generalized linear mixed model; marginal–pairwise likelihood; marginal likelihood; quality-of-life; random effects; Rasch model; variance component

1. Introduction

Item response theory (IRT) models (see [1,2, pp. 3–14]) are increasingly used in various fields, where subjective variables need to be measured using questionnaires with dichotomous or polytomous items. They are sometimes used in health sciences and clinical trials, where these subjective variables could be pain, depression or quality-of-life. Other examples come from marketing where satisfaction or attitudes need to be well-measured and educational-testing services where well-calibrated examinations need to be produced. One of the most popular IRT models is the Rasch model [3]. For a single administration of a questionnaire, there are in fact two Rasch models: one with fixed individual parameters and another one with the random effect. The first model belongs to the family of generalized linear models (GLMs) and the second one to the generalized linear mixed models (GLMMs). When the primary interest is the population or to compare treatment groups, we consider the Rasch model with random effects. The fixed-effects parameters and the random effects of this model are, respectively, called item difficulty parameters and latent trait (see [2, pp. 8–9]). It is well known that estimating these parameters by the method of maximum likelihood faces computational difficulties.

*Corresponding author. Email: Mohand-Larbi.Feddag@univ-nantes.fr
†Current affiliation: EA 4275 'Biostatistique, Recherche Clinique et Mesures Subjectives en Sante', Faculté de Pharmacie, Université de Nantes, Nantes, France.

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2010.538691
http://www.tandfonline.com

Parameter estimation in GLMMs has been tackled in several ways but rarely from a general point of view. Thus, in order to evaluate the likelihood, integral calculus is required which is not analytically feasible. Therefore, several kinds of approximations are considered. One approach consists in numerical approximations of the integral by the Gaussian quadrature [4]. The useful approach is the maximum of the marginal likelihood (MML). Feddag *et al.* [5] have proposed a generalized estimating equation (GEE) approach by combining the work of Prentice [6], and Prentice and Zhao [7] and the approximations of Sutradhar and Rao [8]. As all the methods are based on the Taylor approximations (e.g. for the penalized quasi-likelihood (PQL) approach see [9,10]), the main disadvantage of these approaches is their validity only for the small variance component. Lee and Nelder [11] introduced the h-likelihood for inference in hierarchical GLMs, which can be viewed as an extended likelihood. It avoids the computational difficulties in the calculation of the maximum likelihood for complex models [12, chap. 6, pp. 173–199]. This approach has some bias for binary data, but it has been improved by a modification such as the second-order correction method to hierarchical-likelihood (HL) and denoted by HL(2). In the simulation study of Lee *et al.* [13], it is shown that PQL and HL may have a large bias in estimating the fixed parties, but both seems to have non-ignorable bias in estimating the variance component. By using the second-order HL (HL(2)), we can further reduce the bias of HL, which then becomes almost identical to MML.

Alternatively to these classical approaches, we propose the pairwise-likelihood (denoted by PL) method to estimate simultaneously the fixed-effect parameter and the variance component of the mixed Rasch model. In contrast to the approximate methods, this approach does not have any restriction on the variance component. The main advantage of this method is that it belongs to the broad class of pseudo-likelihood, first proposed by Besag [14] and also termed composite likelihood by Lindsay [15]. The motivation behind this class is to replace the likelihood by a function that is easier to evaluate, and hence to maximize. The function in question is a product of conditional or marginal densities. Thus, the main feature of a pseudo-likelihood function is that it is composed of pieces of likelihoods that can be exploited to prove general results about the consistency and asymptotic normality of pseudo-likelihood estimators.

Examples of PL estimation to multivariate correlated binary data were proposed by LeCessie and Van Houwelingen [16], Kuk and Nott [17], Renard *et al.* [18], Cox and Reid [19] and Bellio and Varin [20].

Considering the Rasch models with fixed effects, there are previous works based on the conditional PL, in which the person parameters are eliminated. Zwinderman [21] has proposed a pairwise parameter estimation in Rasch models, where the item parameters are estimated by the PL of all the pairs of item responses given the latent trait. The obtained estimators are consistent and similar in efficiency to the conditional maximum-likelihood (CML) and MML estimators. For the Rasch model with ordered response categories, Andrich and Luo [22] proposed a pairwise conditional algorithm with the use of the principal components. This method of estimation has two main advantageous properties: the consistency of the parameters and its adaptation for missing data.

However, the PL estimation has not yet been studied in the framework of Rasch models with random effects. We evaluated the potential of using the PL approach in the context of mixed Rasch models and compared its performance with the MML approach which is more traditionally used. We point out that this proposed marginal approach is different from the one proposed by Andrich and Luo [22], which is based on the conditional PL. This approach is more interesting for the complex models, which are faced with numerical problems. For example, in longitudinal studies as proposed by Feddag and Bacci [23] or in GLMs with crossed random effects [20], the PL reduces the high-dimensional integral involved in the marginal likelihood.

The specific outline of the paper is as follows. In Section 2, we present the model considered with the logit link. In Section 3, we define the PL where the marginal pairwise probabilities are detailed. In Section 4, we present some simulation results in which the pairwise approach is compared with the MML, one obtained by the use of STATA Rasch test command [24] and the Proc NLMIXED of the SAS software [25]. A simulation study is also presented for the misspecification of the random effect. The emotional behaviour subscale of the Sickness impact profile questionnaire [26] is also used. A discussion is finally presented in Section 5.

2. Rasch model

Let $Y = \{Y_{ij},\ i = 1, \dots, N;\ j = 1, \dots, J\}$ be a set of binary variable, where Y_{ij} is the response of individual i to item j, coded by 0 and 1. Consider $U_i,\ i = 1, \dots, n$, the random effect (person parameter) associated with the individuals. We denote by y_{ij} and u_i the realizations of Y_{ij} and U_i, respectively.

The mixed Rasch model satisfies the following assumptions which are common to the IRT models.

- Given the random effect u_i, the variables $Y_{i1}, \dots, Y_{iJ}$ are independent with probability given by

$$p_{ij} = \Pr(Y_{ij} = 1 \mid u_i, \beta_j) = \frac{\exp[(u_i - \beta_j)]}{1 + \exp(u_i - \beta_j)},$$

  where β_j is the item difficulty (fixed-effects parameter) associated with the item j.
- The linear predictor η_{ij} is given by

$$\eta_{ij} = g(\Pr(Y_{ij} = 1 \mid u_i, \beta_j)) = u_i - \beta_j,$$

  where g is the link function, which could be the logit defined by $\operatorname{logit}(\pi) = \ln(\pi/(1-\pi))$ or the probit Φ^{-1}, where Φ is the CDF of the normal distribution.
- The random effects $u_i,\ i = 1, \dots, N$, are mutually independent and identically normally distributed with mean 0 and variance σ^2.

We are interested in estimating the parameters $\theta = (\beta, \sigma)$, where $\beta = (\beta_1, \dots, \beta_J)$. The marginal likelihood is given by

$$L(\theta; y) = \prod_{i=1}^{N} \int_{\mathbb{R}} \prod_{j=1}^{J} p_{ij}^{y_{ij}} (1 - p_{ij})^{1-y_{ij}} \varphi(u_i)\, \mathrm{d}u_i, \tag{1}$$

where $\varphi(\cdot)$ is the distribution of the normal variable with mean 0 and variance σ^2.

As an alternative to the MML or the GEE approach (see Feddag *et al.* [5]), we propose the PL method. We give below the definition of this approach.

3. PL for the Rasch model

An alternative to the marginal likelihood is the pseudo-likelihood obtained by replacing the joint likelihood by any product of the conditional or marginal densities. In this work, our attention is restricted to products of marginal PLs. The basic contribution of the ith individual to the log PL denoted by $l_{2,i}(,)$ is given by

$$l_{2,i}(\theta; y_i) = \sum_{j=1}^{J} \sum_{j^\star > j} \log P(Y_{ij} = y_{ij}, Y_{ij^\star} = y_{ij^\star}; \theta), \tag{2}$$

where

$$P(Y_{ij} = y_{ij}, Y_{ij^\star} = y_{ij^\star}; \theta) = \int P(Y_{ij} = y_{ij} \mid u_i) P(Y_{ij^\star} = y_{ij^\star} \mid u_i) \varphi(u_i)\, \mathrm{d}u_i.$$

For each individual i, we have $J * (J-1)/2$ pairs of marginal probability. These marginal–pairwise probabilities are straightforward to calculate in terms of univariate and bivariate probit for the multilevel probit models as suggested by Renard *et al.* [18].

For the logit version of the model, the marginal–pairwise probabilities are not straightforward. These are derived from the expression of the logistic distribution function as a normal scale mixture (see [27,28]), which is given by

$$F(t) = \frac{e^t}{1+e^t} \simeq \sum_{i=1}^{k} p_{k,i} \Phi(t S_{k,i}),$$

where $(S_{k,i}, p_{k,i})$ are tabled for $k = 1, \ldots, 8$ and Φ denotes the standardized normal distribution.

Then the two first marginal probabilities are approximated as follows:

$$P(Y_{ij} = 1; \theta) = \sum_{l=1}^{k} p_{k,l} \Phi\left(\frac{-\beta_j S_{k,l}}{\sqrt{1 + \sigma^2 S_{k,l}^2}} \right), \tag{3}$$

$$P(Y_{ij} = 1, Y_{ij^\star} = 1; \theta) = \sum_{l=1}^{k} \sum_{h=1}^{k} p_{k,l} p_{k,h} \Phi_2\left(\frac{-\beta_j S_{k,l}}{\sqrt{1 + \sigma^2 S_{k,l}^2}}, \frac{-\beta_{j^\star} S_{k,h}}{\sqrt{1 + \sigma^2 S_{k,h}^2}}; \rho(k,h) \right), \tag{4}$$

where the function $\Phi_2(x, y; \rho)$ denotes the standardized bivariate normal distribution function with correlation coefficient ρ, and $\rho(k,h)$ is given by

$$\rho(k,h) = \frac{S_{k,l} S_{k,h} \sigma^2}{\sqrt{1 + \sigma^2 S_{k,l}^2} \sqrt{1 + \sigma^2 S_{k,h}^2}}.$$

The three marginal probabilities related to the combinations $(1,0)$, $(0,1)$, $(0,0)$ are straightforward and are obtained using the above probabilities given by Equations (3) and (4).

The maximum pairwise-likelihood (MPL) estimators $\hat{\theta}_{\mathrm{MPL}} = (\hat{\beta}_{\mathrm{MPL}}, \hat{\sigma}_{\mathrm{MPL}})$ are obtained by maximizing the whole log PL function given by

$$l_2(\theta; y) = \sum_{i=1}^{N} l_{2,i}(\theta; y_i) = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{j^\star > j} \log P(Y_{ij} = y_{ij}, Y_{ij^\star} = y_{ij^\star}; \theta). \tag{5}$$

It follows from the standard theory of estimating equations (see [15,19,20]) that as the number of individuals N increases, $\hat{\theta}_{\mathrm{MPL}}$ is asymptotically normal with mean θ and asymptotic covariance matrix $H^{-1}KH^{-1}$, where

$$H = \mathbb{E}\left(-\sum_{i=1}^{N} \frac{\partial^2 l_{2,i}}{\partial\theta\partial\theta^T}\right),$$

and

$$K = \mathbb{E}\left(\sum_{i=1}^{N} \frac{\partial l_{2,i}}{\partial\theta}\frac{\partial l_{2,i}}{\partial\theta^T}\right).$$

To obtain sample estimates of the standard errors (s.e.) of $\hat{\theta}_{\mathrm{MPL}}$, we estimate H and K by

$$\begin{aligned}\hat{H} &= -\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{j^\star>j} \frac{\partial \log P(y_{ij}, y_{ij^\star})}{\partial\theta}\frac{\partial \log P(y_{ij}, y_{ij^\star})}{\partial\theta^T}, \\ \hat{K} &= \sum_{i=1}^{N} \frac{\partial l_{2,i}(\hat{\theta})}{\partial\theta}\frac{\partial l_{2,i}(\hat{\theta})}{\partial\theta^T}.\end{aligned} \qquad (6)$$

4. Illustrations

The PL approach is illustrated by a simulation study to evaluate its performance and is compared with the MML obtained by the use of the STATA software [29] (denoted by $\mathrm{MML_{Stata}}$) and by Proc NLMIXED of the SAS software (denoted by $\mathrm{MML_{SAS}}$). Thereafter, a simulation study is presented to investigate how the misspecification of the distribution of the random effects will affect the fixed-effects parameter. On the real data, it is further compared with the MML approach obtained by the software RSP (see [30]) and with the GEE method (see [5]).

4.1. *Simulation study*

A simulation study was conducted to evaluate the performance of the PL estimator. This method is compared with the MML approach with the use of the Stata *raschtest* command proposed by Hardouin [24], where the integral involved in the ML given by Equation (1) is approximated by Gauss–Hermite quadrature methods (denoted $\mathrm{MML_{Stata}}$) and with Proc NLMIXED for the Rasch model of the SAS software [25] (denoted by $\mathrm{MML_{SAS}}$).

The Stata software uses the Gauss–Hermite quadrature with 12 points and $\mathrm{MML_{SAS}}$ uses the adaptive Gauss–Hermite quadrature, where the number of points is selected adaptively. The calculation of the maximum for the three approaches is obtained by the direct maximization algorithm: the simplex algorithm [31] for the PL, Newton–Raphson for $\mathrm{MML_{Stata}}$ and quasi-Newton for Proc NLMIXED.

The parameters considered for this study are as follows:

- Two sample sizes: $N = 100,\ 300$
- three different sets of item difficulty
  - (i) $J = 2, \beta = (-1, 0.5)$
  - (ii) $J = 4, \beta = (-1, -0.2, 0.5, 1)$
  - (iii) $J = 9, \beta = (-3, -2, -1.5, -1, 0.5, 1, 1.5, 2, 3)$
- $\sigma = 0.5,\ 1,\ 2$

The mean and the standard deviation (s.d.) of the estimates values obtained for each size based on 500 replications are given in Tables 1–5.

Table 1. Parameter estimates for $\beta = (-1, 0.5)$ and σ, and their s.d. in parenthesis for $J = 2$ for both sizes, with the normal random effect.

		$N = 100$			$N = 300$		
True σ	Approach	β_1	β_2	σ	β_1	β_2	σ
0.5	PL	−1.060	0.529	0.546	−1.024	0.506	0.469
		(0.267)	(0.247)	(0.455)	(0.149)	(0.128)	(0.308)
	MML_{Stata}	−1.060	0.530	0.548	−1.024	0.516	0.470
		(0.268)	(0.247)	(0.454)	(0.149)	(0.129)	(0.307)
	$MML_{SAS}^{\star\star}$	−1.001	0.535	0.746	−1.050	0.503	0.593
		(0.267)	(0.254)	(0.350)	(0.150)	(0.136)	(0.223)
1	PL	−1.037	0.515	1.030	−1.006	0.500	1.004
		(0.316)	(0.268)	(0.496)	(0.176)	(0.145)	(0.266)
	MML_{Stata}	−1.037	0.515	1.030	−1.006	0.501	1.004
		(0.316)	(0.269)	(0.496)	(0.177)	(0.146)	(0.266)
	$MML_{SAS}^{\star\star}$	−1.036	0.513	1.068	−1.002	0.499	0.993
		(0.304)	(0.265)	(0.389)	(0.173)	(0.144)	(0.242)
2	PL	−1.050	0.544	2.124	−1.019	0.504	2.039
		(0.370)	(0.354)	(0.629)	(0.223)	(0.215)	(0.312)
	MML_{Stata}	−1.051	0.545	2.127	−0.998	0.520	2.013
		(0.373)	(0.355)	(0.634)	(0.163)	(0.155)	(0.251)
	MML_{SAS}	−0.997	0.526	1.894	−0.997	0.491	1.849
		(0.336)	(0.330)	(0.445)	(0.173)	(0.144)	(0.235)

Table 2. Parameter estimates for $\beta = (-1, -0.2, 0.5, 1)$ and σ, and their s.d. in parenthesis for $J = 4$ and $N = 100$, with the normal random effect.

True σ	Approach	β_1	β_2	β_3	β_4	σ
0.5	PL	−1.031	−0.197	0.515	1.032	0.498
		(0.248)	(0.222)	(0.219)	(0.248)	(0.268)
	MML_{Stata}	−1.030	−0.197	0.515	1.031	0.449
		(0.248)	(0.222)	(0.219)	(0.248)	(0.256)
	MML_{SAS}	−1.050	−0.204	0.527	1.050	0.524
		(0.249)	(0.215)	(0.277)	(0.255)	(0.195)
1	PL	−0.989	−0.182	0.516	0.999	1.007
		(0.272)	(0.255)	(0.242)	(0.272)	(0.217)
	MML_{Stata}	−1.002	−0.220	0.509	1.008	1.010
		(0.276)	(0.250)	(0.256)	(0.266)	(0.227)
	MML_{SAS}	−1.002	−0.220	0.509	1.008	1.010
		(0.277)	(0.251)	(0.257)	(0.266)	(0.227)
2	PL	−1.014	−0.213	0.537	1.046	2.042
		(0.350)	(0.330)	(0.342)	(0.337)	(0.321)
	MML_{Stata}	−1.012	−0.212	0.539	1.046	2.040
		(0.347)	(0.328)	(0.341)	(0.336)	(0.320)
	MML_{SAS}	−1.001	−0.213	0.537	1.049	2.017
		(0.347)	(0.328)	(0.340)	(0.336)	(0.304)

Table 1 corresponding to $J = 2$ shows that the estimates given by the three approaches for the size $N = 100$ are biased with considerable s.d. We note that for $N = 300$, the estimates are better: the bias and the s.d. decrease. We point out that there is a difference between MML_{SAS} and the two other approaches for $\sigma = 0.5,\ 1$ for both sizes. The large bias for the estimate of $\sigma = 0.5$ with MML_{SAS} is mainly caused by the rate of the convergent data sets, which is equal to 70% for $N = 100$ and 77.2% for $N = 300$. The bias and the number of non-convergent replications decrease considerably for $\sigma = 1$ for both sizes. In fact, we have 468 convergent replications with $N = 100$ and 498 for $N = 300$. It could also deal with the quasi-Newton algorithm used in the

Table 3. Parameter estimates for $\beta = (-1, -0.2, 0.5, 1)$ and σ, and their s.d. in parenthesis for $J = 4$ and $N = 300$, with the normal random effect.

True σ	Approach	β_1	β_2	β_3	β_4	σ
0.5	PL	−1.014	−0.203	0.503	1.007	0.487
		(0.136)	(0.123)	(0.131)	(0.142)	(0.187)
	$\text{MML}_{\text{Stata}}$	−1.014	−0.203	0.503	1.006	0.495
		(0.136)	(0.123)	(0.131)	(0.142)	(0.163)
	MML_{SAS}	−1.016	−0.205	0.504	1.008	0.506
		(0.137)	(0.124)	(0.131)	(0.143)	(0.148)
1	PL	−1.000	−0.204	0.497	1.003	1.006
		(0.151)	(0.134)	(0.134)	(0.159)	(0.128)
	$\text{MML}_{\text{Stata}}$	−1.000	−0.204	0.497	1.002	1.006
		(0.150)	(0.134)	(0.134)	(0.159)	(0.128)
	MML_{SAS}	−1.000	−0.204	0.498	1.000	1.005
		(0.153)	(0.134)	(0.134)	(0.160)	(0.128)
2	PL	−1.003	−0.194	0.515	1.011	2.016
		(0.212)	(0.195)	(0.194)	(0.198)	(0.178)
	$\text{MML}_{\text{Stata}}$	−1.002	−0.195	0.516	1.011	2.014
		(0.210)	(0.196)	(0.194)	(0.198)	(0.178)
	MML_{SAS}	−1.001	−0.195	0.514	1.009	1.995
		(0.210)	(0.195)	(0.194)	(0.198)	(0.170)

Table 4. Parameter estimates for $\beta = (-3, -2, -1.5, -1, 0.5, 1, 1.5, 2, 3)$ and σ, and their s.d. in parenthesis for $J = 9$, $N = 100$, with the normal random effect.

True σ	Approach	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	σ
0.5	PL	−3.082	−2.033	−1.539	−1.004	0.520	1.030	1.553	2.085	3.109	0.465
		(0.526)	(0.321)	(0.292)	(0.257)	(0.224)	(0.236)	(0.279)	(0.337)	(0.495)	(0.235)
	$\text{MML}_{\text{Stata}}$	−3.095	−2.074	−1.534	−1.015	0.516	1.014	1.494	2.035	3.225	0.490
		(0.528)	(0.370)	(0.269)	(0.251)	(0.221)	(0.231)	(0.271)	(0.337)	(0.525)	(0.182)
	MML_{SAS}	−3.100	−2.075	−1.536	−1.017	0.516	1.015	1.498	2.037	3.103	0.508
		(0.482)	(0.305)	(0.292)	(0.251)	(0.269)	(0.273)	(0.301)	(0.341)	(0.498)	(0.158)
1	PL	−3.081	−2.030	−1.506	−0.999	0.525	1.015	1.528	2.056	3.096	0.997
		(0.502)	(0.310)	(0.304)	(0.251)	(0.271)	(0.277)	(0.300)	(0.345)	(0.499)	(0.148)
	$\text{MML}_{\text{Stata}}$	−3.049	−1.984	−1.519	−1.017	0.509	1.018	1.519	2.037	3.177	1.003
		(0.483)	(0.308)	(0.292)	(0.248)	(0.268)	(0.272)	(0.302)	(0.341)	(0.498)	(0.160)
	MML_{SAS}	−3.045	−1.982	−1.520	−1.015	0.510	1.016	1.519	2.036	3.106	1.002
		(0.482)	(0.305)	(0.292)	(0.251)	(0.269)	(0.273)	(0.301)	(0.341)	(0.498)	(0.160)
2	PL	−3.029	−2.024	−1.512	−0.999	0.542	1.037	1.561	2.058	3.088	2.033
		(0.466)	(0.392)	(0.378)	(0.337)	(0.349)	(0.357)	(0.393)	(0.375)	(0.454)	(0.243)
	$\text{MML}_{\text{Stata}}$	−3.069	−2.004	−1.519	−0.990	0.506	1.013	1.536	2.036	3.058	2.035
		(0.462)	(0.370)	(0.357)	(0.334)	(0.347)	(0.363)	(0.386)	(0.372)	(0.449)	(0.240)
	MML_{SAS}	3.068	−2.004	−1.519	−0.990	0.506	1.012	1.536	2.036	3.058	2.030
		(0.461)	(0.369)	(0.356)	(0.333)	(0.347)	(0.363)	(0.386)	(0.371)	(0.448)	(0.239)

optimization of this approach. For the last value of σ, there is no significant difference between the three approaches.

Table 2 corresponding to $J = 4$ and $N = 100$ shows that the bias for all the parameter estimates given by the three approaches are very small and close to 0. Their s.d. also are small and the s.d. of the items parameter seem to increase with σ. The results are very similar for the two methods. Regarding the bias and the s.d., these results show clearly that the PL approach is as good as the MML one. In time, running the MML approach seems a bit faster than the PL one. This is caused by the approximation in the normal scale mixture. With $N = 300$ as shown in Table 3, the results

Table 5. Parameter estimates for $\beta = (-3, -2, -1.5, -1, 0.5, 1, 1.5, 2, 3)$ and σ, and their s.d. in parenthesis for $J = 9$, $N = 300$, with the normal random effect.

True σ	Approach	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	σ
0.5	PL	−3.041	−1.978	−1.507	−0.999	0.525	1.027	1.531	2.039	3.043	0.522
		(0.301)	(0.187)	(0.161)	(0.146)	(0.131)	(0.143)	(0.158)	(0.190)	(0.294)	(0.113)
	MML_{Stata}	−3.028	−2.009	−1.514	−1.008	0.503	1.007	1.504	2.010	3.031	0.488
		(0.256)	(0.182)	(0.158)	(0.144)	(0.123)	(0.146)	(0.161)	(0.185)	(0.290)	(0.110)
	MML_{SAS}	−3.028	−2.009	−1.514	−1.008	0.503	1.007	1.504	2.011	3.031	0.489
		(0.256)	(0.182)	(0.158)	(0.144)	(0.123)	(0.146)	(0.161)	(0.185)	(0.290)	(0.105)
1	PL	−3.020	−1.981	−1.493	−0.991	0.515	1.019	1.516	2.022	3.013	0.994
		(0.261)	(0.199)	(0.172)	(0.155)	(0.146)	(0.148)	(0.164)	(0.182)	(0.277)	(0.092)
	MML_{Stata}	−2.997	−2.004	−1.502	−1.004	0.507	1.014	1.507	2.012	3.033	0.998
		(0.257)	(0.200)	(0.162)	(0.156)	(0.145)	(0.146)	(0.166)	(0.178)	(0.272)	(0.087)
	MML_{SAS}	−2.997	−2.004	−1.501	−1.004	0.507	1.013	1.507	2.011	3.033	0.997
		(0.257)	(0.200)	(0.162)	(0.156)	(0.145)	(0.146)	(0.166)	(0.178)	(0.272)	(0.087)
2	PL	−3.028	−1.995	−1.491	−0.975	0.512	1.0180	1.526	2.009	3.044	2.003
		(0.254)	(0.217)	(0.212)	(0.209)	(0.205)	(0.199)	(0.208)	(0.206)	(0.279)	(0.132)
	MML_{Stata}	−3.030	−2.009	−1.533	−1.014	0.497	1.011	1.500	2.004	3.015	2.015
		(0.255)	(0.214)	(0.208)	(0.209)	(0.202)	(0.197)	(0.206)	(0.219)	(0.267)	(0.131)
	MML_{SAS}	−3.030	−2.008	−1.532	−1.013	0.497	1.011	1.500	2.004	3.015	2.011
		(0.254)	(0.214)	(0.208)	(0.208)	(0.202)	(0.197)	(0.206)	(0.219)	(0.267)	(0.132)

are the same for the two methods and are better as expected: the bias is smaller (negligible) and the s.d. decreases hence providing better precision of the estimates.

Finally, for the third set of item difficulty with $J = 9$, the results given in Tables 4 and 5 are similar to those corresponding to $J = 4$. For $N = 100$ as shown in Table 4, the estimates of all the parameters for the three approaches has a small bias with a considerable s.d. And as expected, the bias and the s.d. decrease for all the estimates with the size $N = 300$ as shown in Table 5.

4.2. *Sensitivity analysis: misspecification*

Nehaus and Hauck [32] examined the performance of the mixed-effects logistic regression analysis, when the distribution of the random effects is misspecified. In their simulation study, they have considered the Gamma and the t distribution. In our study, we consider only the Gamma distribution for the random effects with scale parameter α and shape $k = 1$: $U \sim \Gamma(1, \alpha)$ with density distribution given by $f(x) = (1/\alpha)\exp(-(x/\alpha)),\ x > 0, \alpha > 0$.

We denote the integrals involved in the marginal likelihood and PL by $I_i^1(h_1)$ and $I_i^2(h_2)$, respectively, given by

$$I_i^1(h_1) = \frac{1}{\alpha}\int_0^{+\infty} \frac{\exp(u_i - \beta_j)y_{ij}}{1 + \exp(u_i - \beta_j)} \exp\left(-\frac{u_i}{\alpha}\right) \mathrm{d}u_i$$

$$= \int_0^{+\infty} \frac{\exp(\alpha u_i - \beta_j)y_{ij}}{1 + \exp(\alpha u_i - \beta_j)} \exp(-u_i)\,\mathrm{d}u_i,$$

$$I_i^2(h_2) = \int_0^{+\infty} \frac{\exp(\alpha u_i - \beta_j)y_{ij}}{1 + \exp(\alpha u_i - \beta_j)} \frac{\exp(\alpha u_i - \beta_{j^\star)}y_{ij^\star}}{1 + \exp(\alpha u_i - \beta_{j^\star})} \exp(-u_i)\,\mathrm{d}u_i.$$

The integrals defined above has the form for $k = 1, 2$, $I_i^k(h_k) = \int_0^{+\infty} h_k(y_i \mid u_i, \beta, \alpha) \times \exp(-u_i)\,\mathrm{d}u_i$. Using the Laguerre quadrature, these integrals are approximated by

$$I_i^k(m) = \sum_{l=1}^{m} w_l h_k(y_i \mid z_l, \beta, \alpha),$$

where w_i and z_i, $i = 1, \ldots, m$, are Laguerre quadrature weights and nodes computed by the *Statmod* package of the R software [33].

A simulation study based on 500 replications was conducted to compare the PL and MML approaches on the Rasch model with the following parameters:

- Two sample sizes: $N = 100,\ 300$
- $J = 4$ items
- $\beta = (-1, -0.2, 0.5, 1)$
- $\alpha = 0.5,\ 1,\ 4$

We give in Tables 6 and 7 the mean and the s.d. of the estimates for the size 100 and 300, where the integrals are approximated by the Laguerre quadrature with 20 points.

For the size $N = 100$, as shown in Table 6, the bias and the s.d. for all the estimates given by both approaches are not considerable, except the s.d. of the estimate of α for the value 4, which is greater than 1. However, as expected the results for the size $N = 300$ given in Table 7, are better: the bias is smaller and the s.d. decreases. With regard to the s.d.: for all the parameters, s.d. for the MML approach are smaller than those for the PL method. That indicates as expected the loss of efficiency with the PL approach. According to this results, we can say that the misspecification

Table 6. Parameter estimates for $\beta = (-1, -0.2, 0.5, 1)$ and α, and their s.d. in parenthesis for $N = 100$, with the Gamma random effect.

True α	Approach	β_1	β_2	β_3	β_4	α
0.5	PL	−1.078	−0.251	0.460	0.953	0.465
		(0.369)	(0.340)	(0.361)	(0.399)	(0.335)
	MML	−1.079	−0.253	0.458	0.951	0.462
		(0.368)	(0.338)	(0.360)	(0.398)	(0.329)
1	PL	−1.066	−0.237	0.461	0.974	0.991
		(0.393)	(0.361)	(0.390)	(0.413)	(0.412)
	MML	−1.063	−0.234	0.465	0.978	0.996
		(0.395)	(0.359)	(0.388)	(0.413)	(0.409)
4	PL	−1.113	−0.229	0.484	1.011	4.167
		(0.574)	(0.492)	(0.512)	(0.554)	(1.383)
	MML	−1.100	−0.214	0.501	1.032	4.22
		(0.570)	(0.483)	(0.499)	(0.543)	(1.360)

Table 7. Parameter estimates for $\beta = (-1, -0.2, 0.5, 1)$ and α, and their s.d. in parenthesis for $N = 300$, with the Gamma random effect.

True α	Approach	β_1	β_2	β_3	β_4	α
0.5	PL	−1.049	−0.238	0.465	0.960	0.462
		(0.231)	(0.228)	(0.236)	(0.249)	(0.218)
	MML	−1.050	−0.239	0.464	0.959	0.461
		(0.225)	(0.219)	(0.228)	(0.242)	(0.210)
1	PL	−1.013	−0.199	0.501	0.997	1.006
		(0.221)	(0.202)	(0.214)	(0.232)	(0.235)
	MML	−1.011	−0.197	0.503	0.999	1.008
		(0.215)	(0.193)	(0.204)	(0.220)	(0.221)
4	PL	−1.015	−0.205	0.499	1.007	4.037
		(0.300)	(0.274)	(0.261)	(0.293)	(0.721)
	MML	−1.011	−0.201	0.503	1.012	4.050
		(0.295)	(0.271)	(0.257)	(0.284)	(0.696)

of the distribution of the random effects does not affect the item parameters neither in terms of bias nor in terms of s.d.

Further simulations not reported in this study dealing with the centred gamma random effects show that the estimates present a small bias for small variance component α and wrong estimates for large α.

4.3. *Real data*

In this section, we illustrate the application of these two approaches to the analysis of real data from a quality-of-life study. Further, we have used the MML approach (denoted by MML_{em}), which uses the expectation–maximization (EM) algorithm, where the estimates are obtained by the software program RSP (see [30]) and the GEE approach (see [5]). The sample is composed of 470 depressive patients who answered the French version of the emotional behaviour subscale of the Sickness Impact Profile (SIP) questionnaire (see Bergner *et al.* [26] for the international version). This questionnaire includes 12 dimensions (subscales), each one relating to a particular aspect of quality-of-life. The items and their frequencies of responses are presented in Feddag *et al.* [5].

Table 8 presents the estimation of the difficulty parameter β and the variance components σ^2, and their s.e. The standard error is estimated using the sandwich estimator for the GEE approach and expression (6) for the PL one.

From Table 3, we note that the estimates obtained under the PL, MML, MML_{em} and GEE approaches for the model are similar for both the item difficulty and the variance component parameters. In fact, the large difference between the item difficulty parameters is very small and the variance component estimates are equal. The s.d. of the estimates are very close for the four approaches.

It is clear that for all the estimation methods, the most difficult item is number 4 (with the largest estimation) and the easiest is item 7 (with the smallest estimation). The estimate of β_3 is close to 0, which means that item 3 has been positively responded by approximately 50% of individuals. This rate is confirmed in the table presented in Feddag *et al.* [5], where the positive response to item 3 is equal to 48%. We point out that this estimated value does not affect the measurement of the latent trait.

Instead of the residual analysis for the validation of the model which is not easy for this model, we have used the R_{1m} test proposed by Glas [34]. It compares the observed frequency of the positive and negative response for each item in different groups of individuals as a function of the scores. This test is available in Stata and RSP softwares. Under the hypothesis H_0: 'Good adequacy to the model', the statistic test following an asymptotic χ^2 distribution with $df = 55$ are given as: 60.075 with $p = 0.297$ for the MML_{em}, 60.079 with $p = 0.2969$ for the $\text{MML}_{\text{Stata}}$, and 60.028 with $p = 0.289$ for the GEE. All the tests are not significant, which means that we do not reject the adequacy.

Table 8. Parameter estimates and their s.e. in parenthesis for the emotional behaviour subscale of SIP data.

Approach	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	σ^2
PL	−0.714	−0.274	0.097	2.034	−1.268	0.843	−1.414	−1.375	−0.608	0.672
	(0.128)	(0.109)	(0.105)	(0.177)	(0.154)	(0.122)	(0.171)	(0.166)	(0.120)	(0.172)
MML_{em}	−0.683	−0.273	0.099	2.037	−1.270	0.846	−1.415	−1.375	−0.608	0.678
	(0.099)	(0.095)	(0.095)	(0.131)	(0.107)	(0.101)	(0.111)	(0.110)	(0.098)	(0.103)
$\text{MML}_{\text{Stata}}$	−0.712	−0.273	0.099	2.034	−1.268	0.844	−1.413	−1.373	−0.607	0.671
	(0.110)	(0.106)	(0.105)	(0.145)	(0.120)	(0.112)	(0.124)	(0.123)	(0.109)	(0.096)
GEE	−0.702	−0.246	0.128	2.028	−1.221	0.863	−1.411	−1.371	−0.570	0.674
	(0.108)	(0.104)	(0.103)	(0.154)	(0.120)	(0.113)	(0.123)	(0.122)	(0.107)	(0.102)

5. Discussion

This paper describes a PL approach to estimate simultaneously the item difficulty parameters and the variance component of the mixed Rasch model with the logit link. This method which belongs to the broad class of pseudo-likelihood is developed to solve the estimation problems of the Rasch model, even if many have already been solved. In fact, the CML and the marginal maximum-likelihood methods are available in many computer programs, for example, RSP [30] or the Rasch test command [24]. A simulation study is conducted to analyse the sample performance of the PL approach and to compare it with the MML one. In term of the bias and s.d., the results show that the proposed method is as good as the MML approach. This work could be easily generalized to cope with the Rasch model which includes continuous covariates [35]. However, the following need further investigation: first, the extension of the approach to the model with polychotomous items. Finally, it would be interesting to study the performance of this proposed approach to the unbalanced data structure. For this case, the weighted PL as defined by Kuk and Nott [17] and Joe and Lee [36] will be worthy to study its performance.

We have examined the performance of the two approaches when the distribution of the random effects is misspecified. The simulation study shows that the item difficulty parameters are quite robust to the considered misspecification.

Acknowledgements

The authors would like to thank the two referees and the associate editor for their valuable comments that led to substantive improvements of the paper. The first author gratefully acknowledges support from ESRC grant RES-576-25-5020, 'Developing Statistical Modelling in the Social Sciences', held jointly at the Universities of Lancaster and Warwick (UK).

References

[1] F.B. Baker, *Item Response Theory. Parameter Estimation Techniques*, Marcel Dekker, New York, 1992.
[2] G.H. Fischer and I.W. Molenaar, *Rasch Models Foundations, Recent Developments and Applications*, Springer-Verlag, New York, 1995.
[3] G. Rasch, *On General Laws and the Meaning of Measurement in Psychology*, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, IV, University of California Press, Berkeley, CA, 1961.
[4] S.E. Rigdon and R.K. Tsutakawa, *Parameter estimation in latent trait models*, Psychometrika 48 (1983), pp. 567–574.
[5] M.L. Feddag, I. Grama, and M. Mesbah, *Generalized estimating equations for mixed logistic models*, Commun. Stat. Theory Methods 32(4) (2003), pp. 851–874.
[6] R.L. Prentice, *Correlated binary regression with covariates specific to each binary observation*, Biometrics 44 (1988), pp. 1033–1048.
[7] R.L. Prentice and L.P. Zhao, *Estimating equation for parameters in means and covariances of multivariate discrete and continuous responses*, Biometrics 47 (1991), pp. 825–839.
[8] B.C. Sutradhar and R.P. Rao, *On marginal quasi-likelihood inference in generalized linear mixed models*, J. Multivariate Anal. 76 (2001), pp. 1–34.
[9] N.E. Breslow and D.G. Clayton, *Approximate inference in generalized linear mixed models*, J. Amer. Statist. Assoc. 88 (1993), pp. 9–25.
[10] N.E. Breslow and X. Lin, *Bias correction in generalized linear models with a single component of dispersion*, Biometrika 82 (1995), pp. 81–92.
[11] Y. Lee and J.A. Nelder, *Hierarchical generalised linear models (with discussion)*, J. R. Statist. Soc. B 58 (1996), pp. 619–656.
[12] Y. Lee, J.A. Nelder, and Y. Pawitan, *Generalized Linear Models with Random Effects. Unified Analysis via H-likelihood*, Chapman & Hall/CRC, Raton, FL, 1996.
[13] Y. Lee, J.A. Nelder, and M. Noh, *H-likelihood: Problems and solutions*, Stat. Comput. 17 (2007), pp. 49–55.
[14] J.E. Besag, *Statistical analysis of non lattice data*, Statistician 24(2) (1975), pp. 179–195.
[15] B.G. Lindsay, *Composite likelihood methods*, in *Statistical Inference from Stochastic Processes Contemporary Mathematics 80*, N.U. Prablu, ed., American Mathematical Society, Providence, RI, 1988, pp. 221–39.
[16] S. LeCessie and J.C. van Houwelingen, *Logistic regression for correlated binary data*, Appl. Stat. 43 (1994), pp. 95–108.
[17] Y.C. Kuk and D.N. Nott, *A pairwise likelihood approach to analyzing correlated binary data*, Stat. Probab. Lett. 47 (2000), pp. 329–335.

[18] D. Renard, G. Geert, and H. Geys, *A pairwise likelihood approach to estimation in multilevel probit models*, Comput. Stat. Data Anal. 44 (2004), pp. 649–667.
[19] D.R. Cox and N. Reid, *A note on pseudolikelihood constructed from marginal densities*, Biometrika 91 (2004), pp. 729–737.
[20] R. Bellio and C. Varin, *A pairwise likelihood approach to generalized linear models with crossed effects*, Statist. Model. 5 (2005), pp. 217–227.
[21] A.H. Zwinderman, *Pairwise parameter estimation in the Rasch models*, Appl. Psychol. Meas. 19(4) (1995), pp. 369–375.
[22] D. Andrich and G. Luo, *Conditional pairwise estimation in the Rasch model for ordered response categories using principal components*, J. Appl. Meas. 4(3) (2003), pp. 205–221.
[23] M.L. Feddag and S. Bacci, *Pairwise likelihood for the longitudinal mixed Rasch model*, Comput. Stat. Data Anal. 53 (2009), pp. 1027–1037.
[24] J.B. Hardouin, *Rasch analysis: Estimation and tests with raschtest*, Stata J. 7(1) (2007), pp. 1–23.
[25] J.B. Hardouin and M. Mesbah, *The SAS macro-program: AnaQol to estimate the parameters of IRT models*, Commun. Stat. Simul. Comput. 36(2) (2009), pp. 437–453.
[26] M. Bergner *et al.*, *The sickness impact profile: Validation of a health status measure*, Med. Care 14 (1973), pp. 56–67.
[27] J.F. Monahan and L.A. Stefanski, *Normal scale mixture approximations to $F^\star(z)$ and computation for the logistic-normal integral*, In *Handbook of the Logistic Distribution*, N. Balakrishnan, ed., Marcel Dekker, New York, 1989, pp. 529–540.
[28] M.L. Drum and P. McCullagh, *REML estimation with exact covariance in the logistic mixed models*, Biometrics 49 (1993), pp. 677–689.
[29] *STATA Base Reference Manual G-M Release*, Vol. 2, A Stata Press Publication, STATA Corporation, College Station, TX, 2003, pp. 305–337.
[30] C.A.W. Glas and J.L. Elis, *RSP: Rasch Scaling Program*, iec ProGAMMA, Groningen, The Netherlands, 1993.
[31] J.A. Nelder and R. Mead, *A simplex method for function minimization*, Comput. J. 7 (1965), pp. 308–313.
[32] J.M. Neuhaus and W.W. Hauck, *The effects of mixture distribution misspecification when fitting mixed-effects logistic models*, Biometrika 79 (1992), pp. 755–762.
[33] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0. Available at http://www.R-project.org/.
[34] C.A.W. Glas, *The derivation of some tests for the Rasch model from the multinomial distribution*, Psychometrika 53(4) (1988), pp. 525–546.
[35] W. Kang, M.S. Lee, and Y. Lee, *HGLM versus conditional estimators for the analysis of clustered binary data*, Stat. Med. 24 (2005), pp. 741–752.
[36] H. Joe and Y. Lee, *On weighting of bivariate margins in pairwise likelihood*, J. Multivariate Anal. 100 (2009), pp. 670–685.