Generalized estimating equations for longitudinal mixed Rasch model

Mohand L. Feddag\textsuperscript{a,}\textsuperscript{*}, Mounir Mesbah\textsuperscript{b}

\textsuperscript{a}Laboratory SABRES, University of South Brittany, Rue Yves Mainguy, 56000 Vannes, France
\textsuperscript{b}UFR SSI, University of South Brittany, Rue Yves Mainguy, 56000 Vannes, France

Available online 21 August 2004

Abstract
In this paper, the problem of estimating the fixed effects parameters and the variance of the random effects (variance components) in longitudinal mixed Rasch model is considered. It is well known that estimating these parameters by the method of maximum likelihood faces computational difficulties. As an alternative, we propose the generalized estimating equations approach. Approximations of the joint moments of the variables are proposed. The estimators obtained are consistent and asymptotically normal. We illustrate the usefulness of the method with simulations and with an analysis of real data from quality of life.
© 2004 Elsevier B.V. All rights reserved.

Keywords: IRT models; Rasch model; Generalized linear mixed model; Fixed effects; Random effects; Generalized estimating equations; Longitudinal data; Fisher-scoring algorithm; Quality of life

1. Introduction

Item response theory (IRT) models (Fischer and Molenaar, 1995) are increasingly used in various fields where subjective variables need to be measured using questionnaires with polychotomous items. This is usual 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 exams need to be produced. One of the most popular IRT models is the Rasch model (Fischer and Molenaar, 1995). For a single administration of

\textsuperscript{*} Corresponding author.
E-mail address: mohand-larbi.feddag@univ-ubs.fr (M.L. Feddag).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.06.045
a questionnaire, there are in fact two Rasch models: one with fixed individual parameters and another one with random effect. The first model belongs to the family of Generalized linear models (GLMs) and the second to the Generalized linear mixed models (GLMMs) (McCullagh and Nelder, 1989). When the primary interest is the population or to compare treatment groups, we consider Rasch model with random effects. The fixed effects parameters and the random effects of this model are, respectively, called difficulty parameters and latent traits (see Fischer and Molenaar, 1995). It’s well known that estimating these parameters by the method of maximum likelihood faces computational difficulties.

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 of numerical approximations of the integral by Gaussian quadrature (Rigdon and Tsutakawa, 1983). As an alternative, we propose the GEE approach defined by Liang and Zeger (1986) as an extension of Quasi-likelihood approach (Wedderburn, 1974). They introduced a class of estimating equations that give consistent estimates of regression parameters and of their variances in the class of GLMs for longitudinal data. The correlation parameter is considered as nuisance in their method. Several modifications and extensions of the GEE approach have been studied. Prentice (1988) extends this approach to correlated binary data by specifying second generalized estimating equations, based on the empirical pairwise covariances that permit the estimation of the correlation parameters. Zhao and Prentice (1990), and Prentice and Zhao (1991) have generalized this work to the correlated binary regression using a quadratic exponential model. The later joint estimating equations are referred to as GEE2 (GEE of order two). For the GLMMs, Zeger et al. (1988) have used GEE approach to estimate only the regression parameter through the use of approximations to the mean and the covariance matrix while the variance components is estimated empirically. Breslow and Clayton (1993), and Breslow and Lin (1995) have used the penalized quasi-likelihood approach and the marginal quasi-likelihood for inferences in GLMMs.

The application of GEE to GLMMs requires the first and second-order marginal moments of the responses to estimate the fixed effects and the third and fourth joint moments of the responses to estimate the variance components. All these joint moments do not have a closed form under GLMMs and their approximations are necessary. Sutradhar and Rao (2001) have used approximations of the joint moments up to order four to estimate separately the regression parameters and the variance of the univariate random effects. Feddag et al. (2003) have proposed second estimating equations to estimate the variance of random effects in the logistic mixed model. These equations are based on the empirical pairwise covariances which are used by Prentice (1988) and Prentice and Zhao (1991) to estimate the correlation parameters in GLMs for binary data. The Sutradhar and Rao approximations are used to estimate simultaneously the fixed effects and the variance components parameters.

In the case of quality of life in clinical trials, the same questionnaire is usually administered to the same subjects at various occasions, \( t = 1, \ldots, T \). In such case, the random effects is multivariate and it seems to be well suited to estimate the correlation existing among the different random effects. Recently, Sutradhar and Sinha (2002) have considered correlated random effect based longitudinal mixed (CRBLM) model to binary responses. They are interested only on the estimates of the fixed effects and variance component parameters, the
longitudinal correlation is considered as nuisance. To circumvent the computation of the 3rd and 4th joint moments, they estimate the fixed effects parameters by the GEE2 approach, the longitudinal correlation by the method of moments and the variance component by the pseudo-likelihood approach. The estimation of the later parameter is compared through a simulation study with GEE2 approach. The responses of our model are two-way correlated. First, at a given point of time, the binary responses of a single individual are correlated and second, when they are repeated, they, also become longitudinally correlated. To model both correlations, we have used a GLMM with multivariate random effects. The primary interest is to estimate the fixed effects and the covariance matrix of the random effects parameters. Estimating these parameters by the method of maximum likelihood is not tractable due to the need of numerical evaluation of high-dimensional integrals. The GEE approach avoids this difficulty when the variance of the random effects satisfies some conditions.

Thus our results may be regarded as an extension of Feddag, Grama and Mesbah GEE2 methods (Feddag et al., 2002, 2003) for classical mixed Rasch model to the longitudinal mixed Rasch model. To reach this goal, we have generalized Sutradhar and Rao (2001) approximations to the multivariate random effects case.

The specific outline of the paper is as follow. In Section 2, we present the model considered. In Section 3, we give theoretical results of approximations of the marginal likelihood. Then we derive the approximate joint moments of the variables. The estimating equations for the fixed effects parameters and the covariance matrix of random effects are constructed in Section 4. Next, we establish the asymptotic properties of the estimators. In Section 5, we present some simulation results in order to illustrate our methods. A real data set from a quality of life experiment is also used. A summary and discussions are presented in Section 6.

2. Model description

2.1. Classical mixed Rasch model

Consider a set of $K$ individuals having answered a questionnaire of $J$ dichotomous items. Let $X_{ij}$ be the answers of individual $i$ to item $j$ and $\theta_i$ be the univariate random effects. Three assumptions, common to all IRT logistic models, are as follows:

- Given $\theta_i$, the variables $X_{i1}, \ldots, X_{iJ}$ are mutually independent.
- The probability $p_{ij}$ of the response of the $i$th individual to the $j$th item is given by
  
  \[ p_{ij} = P(X_{ij} = x_{ij} \mid \theta_i, \beta_j) = \frac{\exp((\theta_i - \beta_j)x_{ij})}{1 + \exp(\theta_i - \beta_j)}, \]  

  where $\beta_j$ is the fixed effects parameter associated to item $j$, $x_{ij} = 1$ if the $i$th individual has a positive response (correct, agree) for item $j$ and $x_{ij} = 0$ (false, disagree) otherwise.
- The variables $\theta_1, \ldots, \theta_K$ are mutually independent with a common underlying distribution $G$. 

Generally, we assume that $G$ is the cumulative distribution function of the normal distribution with mean 0 and variance $\sigma^2$.

2.2. Longitudinal mixed Rasch model

We extend now the Rasch model defined before to the repeated measurements data. In such case the random effects become multivariate and the proposed extension should be such that at each time the probability of individual answers conditionally to the random effects follow a Rasch model.

From now on, we will consider a sample of $K$ independent random multivariate binary observations $X = (X^T_{ij})$, where $X^T_{ij}$ is a binary response of individual $i$ to item $j$ at time $t$, $i=1, \ldots, K, j=1, \ldots, J, t=1, \ldots, T$. Let $X^T_{ij}=(X^T_{i1}, \ldots, X^T_{ij})'$ the vector of response at time $t$ of individual $i$, $X^T=(X^T_1, \ldots, X^T_K)'$ the $K \times J$ matrix of variables, $X^T=(X^T_1, \ldots, X^T_T)'$ the vector response of individual $i$ and $\theta_i=(\theta_{i1}, \ldots, \theta_{iT})'$ the multivariate random effects associated to individual $i, i=1, \ldots, K$. The longitudinal Rasch model satisfies the following assumptions:

- Given the random effects $\theta_i, i=1, \ldots, K$, we have

$$P(X^T_{ij} = x^T_{ij} | \theta_i, \beta) = \prod_{t=1}^{T} \prod_{j=1}^{J} P(X^T_{ij} = x^T_{ij} | \theta_{it}, \beta_j), \quad (2)$$

where $\beta = (\beta_1, \ldots, \beta_J)'$ is the fixed effects parameter.

- For all $i, j, t; i=1, \ldots, K, j=1, \ldots, J, t=1, \ldots, T$, we have

$$P(X^T_{ij} = x^T_{ij} | \theta_{it}, \beta_j) = \frac{\exp\left\{ (\theta_{it} - \beta_j)x^T_{ij} \right\}}{1 + \exp(\theta_{it} - \beta_j)}, \quad (3)$$

- The latent variables $\theta_1, \ldots, \theta_K$, are independent and identically distributed with a multivariate normal with mean vector 0 and covariance matrix $\Sigma = (\sigma_{ij})_{j,l=1,\ldots,T}$.

We are interested in estimating $(\beta, \alpha, \gamma)$, where $\beta=(\beta_1, \ldots, \beta_J)'$, $\alpha=(\alpha_{11}, \sigma_{22}, \ldots, \sigma_{TT})'$ and $\gamma = (\sigma_{12}, \ldots, \sigma_{1T}, \sigma_{23}, \ldots, \sigma_{T-1,T})'$. Usually, one uses the maximum likelihood approach. The marginal likelihood of $X$ is given by

$$L(\beta, \alpha, \gamma | x) = \prod_{i=1}^{K} \int_{\mathbb{R}^T} \prod_{t=1}^{T} \prod_{j=1}^{J} \frac{\exp\left\{ (\theta_{it} - \beta_j)x^T_{ij} \right\}}{1 + \exp(\theta_{it} - \beta_j)} \phi(\theta_i, \alpha, \gamma) d\theta_i, \quad (4)$$

where

$$\phi(\theta_j, \alpha, \gamma) = \frac{1}{(2\pi)^{T/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} \theta^T \Sigma^{-1} \theta \right\}$$

is the distribution of the multivariate normal with mean vector 0 and covariance matrix $\Sigma$. 
The maximization of the above function is computationally difficult and requires iterative
techniques. The EM algorithm is not a practical solution in this case since the computation
of E step and the M step are not tractable due to the need of numerical evaluation of high
dimensional integrals. To avoid these difficulties, we propose the GEE approach, which
turns out to be computationally less intensive.

3. The approximation of marginal likelihood and joint moments

The aim of this section is to give the approximations of the joint moments of the observed
variable $X_i$, which we shall use later on in the GEE approach. Their computation requires
an approximation of the marginal likelihood of $X$, given in Eq. (4). Theses approximations
are obtained by using a Taylor series expansion about the random effects $\theta_i$.

Up to now, we assume that the joint moments from order six of the random effects $\theta_i$
satisfies the following condition:

$$q_r = E(\|\theta_i\|^r) = o(f_r(\Sigma)) \text{ for } r \geq 6,$$

where $f_r$ is polynomial function of $r$ and $\Sigma$.

The following

$$E(\theta_{it}^6), \ E(\theta_{it}^5 \theta_{ih}), \ E(\theta_{it}^4 \theta_{ih}^2), \ E(\theta_{it}^3 \theta_{ih}^3), \ E(\theta_{it}^2 \theta_{ih}^2 \theta_{iu}^2), \ E(\theta_{it} \theta_{ih} \theta_{iu} \theta_{iv} \theta_{iw} \theta_{iz}),$$

where $t, h, u, v, w, z = 1, \ldots, T$, are the 6th order joint moments. Under assumption (5),
these joint moments must be negligible. For our case, $\theta_i$ is normally distributed with mean
0 and covariance matrix $\Sigma$, these quantities are easily obtained as function of elements of
$\Sigma$. This condition generalizes assumption given in Sutradhar and Rao (2001) and Feddag et
al. (2002) to the unidimensional case where the variance of the random effects is supposed
to be small. Here, when all elements of $\Sigma$ are small, assumption made in (5) is satisfied.
Nevertheless, as the simulations given in Section 5.1 show that even some elements of $\Sigma$
are not small, we obtain the convergence of the algorithm.

Under the assumption made in (5), we obtain the approximation of the marginal likelihood
given by Eq. (4) and the joint density of $X_j, i = 1, \ldots, K$. These approximations are given
in Theorems A.1 and A.2 presented, respectively, in the Appendices A.1 and A.2.

Using Theorem A.2, we derive the approximate joint moments of the variables $X_j, i =
1, \ldots, K$, given in the following proposition.

**Proposition 1.** For all $i, j, t, h, m, s$ satisfying $1 \leq i \leq K, \ 1 \leq j \leq J, \ 1 \leq t \neq h \leq T, \ m, s =
2, \ldots, J$, the following holds:

(1) The marginal mean $\mu_{ij}^t$ of $X_{ij}$ is given by

$$E(X_{ij}) = \frac{1}{1 + e^{\beta_j}} + \frac{\sigma_{tt}}{2} \frac{e^{\beta_j}(e^{\beta_j} - 1)}{(1 + e^{\beta_j})^3} + \frac{\sigma_{tt}^2}{8} \frac{e^{\beta_j}(e^{3\beta_j} - 11e^{2\beta_j} + 11e^{\beta_j} - 1)}{(1 + e^{\beta_j})^5}. \quad (6)$$
The \( m \)th order marginal joint moment within times is

\[
E(X_{t_1}^{i_1} \ldots X_{t_m}^{i_m}) = \prod_{l=1}^{m} \frac{1}{1 + e^{\beta_{i_l}}} \\
\times \left[ 1 + \frac{\sigma_{tt}}{2} ((\hat{A}_{i_l,j_l}^{t_1 \ldots j_m})^2 - B_{j_1 \ldots j_m}) + \frac{\sigma_{th}^2}{8} \hat{Q}_{i_l,j_l}^{t_1 \ldots j_m} \right].
\]

(7)

(3) The \((m, s)\)th order marginal joint moment between times is

\[
E(X_{t_1}^{i_1} \ldots X_{t_m}^{i_m} X_{h_1}^{i_1} \ldots X_{h_s}^{i_s}) = \prod_{l=1}^{m} \frac{1}{1 + e^{\beta_{i_l}}} \prod_{r=1}^{s} \frac{1}{1 + e^{\beta_{h_r}}} \\
\times \left[ 1 + \frac{1}{2} (\sigma_{tt} \hat{R}_{i_r,j_r}^{t_1 \ldots j_m} + \sigma_{hh} \hat{R}_{i_r,j_r}^{h_1 \ldots j_s}) \right. \\
\left. + \sigma_{th} \hat{A}_{i_r,j_r}^{t_1 \ldots j_m} \hat{A}_{h_r,j_r}^{h_1 \ldots j_s} + \frac{1}{2} (\hat{A}_{i_r,j_r}^{t_1 \ldots j_m} \hat{P}_{i_r,j_r}^{h_1 \ldots j_s} \sigma_{th} \sigma_{hh} \right) \\
+ \hat{P}_{i_r,j_r}^{t_1 \ldots j_m} \hat{A}_{i_r,j_r}^{h_1 \ldots j_s} \sigma_{tt} \sigma_{th} \\
\left. + \frac{1}{4} \hat{R}_{i_r,j_r}^{t_1 \ldots j_m} \hat{R}_{i_r,j_r}^{h_1 \ldots j_s} \sigma_{tt} \sigma_{hh} + 2 \sigma_{th}^2 \right) \\
\left. + \frac{1}{8} (\sigma_{tt}^2 \hat{Q}_{i_r,j_r}^{t_1 \ldots j_m} + \sigma_{hh}^2 \hat{Q}_{i_r,j_r}^{h_1 \ldots j_s}) \right],
\]

(8)

where \( \hat{A}_{i,j,k}^{r_1 \ldots j_m} \), \( \hat{P}_{i,j,k}^{r_1 \ldots j_m} \), \( \hat{Q}_{i,j,k}^{r_1 \ldots j_m} \) \((r = t, h, k = m, s)\) are, respectively, the values of \( A_{i,j,k}^{r_1 \ldots j_m} \), \( P_{i,j,k}^{r_1 \ldots j_m} \), \( Q_{i,j,k}^{r_1 \ldots j_m} \) and \( R_{i,j,k}^{r_1 \ldots j_m} \), quantities given in Theorem 2, at \((x_{i_1}^{r_1}, \ldots, x_{i_k}^{r_k}) = (1, \ldots, 1)\).

The proof of this proposition is given in Appendix A.3.

The covariances \( \text{Cov}(X_{i_j}^{t_1}, X_{i_k}^{t_1}) \) and \( \text{Cov}(X_{i_j}^{h_1}, X_{i_k}^{h_1}) \), for \( 1 \leq t \neq h \leq T, \ 1 \leq j \neq k \leq J \), are easily derived with this proposition and are given by the expressions

\[
\text{Cov}(X_{i_j}^{t_1}, X_{i_k}^{t_1}) = \frac{\sigma_{tt}}{(1 + e^{\beta_j})^2 (1 + e^{\beta_k})^2} \times \left[ e^{\beta_j} e^{\beta_k} + \frac{\sigma_{tt}}{2} H_1(\beta) \right],
\]

(9)

\[
\text{Cov}(X_{i_j}^{h_1}, X_{i_k}^{h_1}) = \frac{\sigma_{th}}{(1 + e^{\beta_j})^2 (1 + e^{\beta_k})^2} \times \left[ e^{\beta_j} e^{\beta_k} + \frac{1}{2} H_2(\beta, \gamma, \gamma) \right],
\]

(10)

where

\[
H_1(\beta) = \frac{(e^{3\beta_j} - 4e^{2\beta_j} - e^{\beta_j})e^{\beta_k}}{(1 + e^{\beta_j})^2} + \frac{e^{\beta_j}(e^{\beta_j} - 1)e^{\beta_k}(e^{\beta_k} - 1)}{(1 + e^{\beta_j})(1 + e^{\beta_k})} \\
+ \frac{e^{\beta_j}(3e^{3\beta_k} - 4e^{2\beta_k} + e^{\beta_k})}{(1 + e^{\beta_k})^2},
\]
\[ H_2(\beta, \alpha, \gamma) = \sigma_{tt} \frac{(e^{3\beta_j} - 4e^{2\beta_j} + e^{\beta_j})e^{\beta_k}}{(1 + e^{\beta_j})^2} + \sigma_{th} \frac{e^{\beta_j}(e^{\beta_j} - 1)e^{\beta_k}(e^{\beta_k} - 1)}{(1 + e^{\beta_j})(1 + e^{\beta_k})} + \sigma_{hh} \frac{e^{\beta_j}(e^{3\beta_k} - 4e^{2\beta_k} + e^{\beta_k})}{(1 + e^{\beta_k})^2}. \]

The qualities of the first two joint moments of these approximations are illustrated graphically in Feddag et al. (2002).

4. Estimation of the parameters

4.1. Estimating equations

Consider \( X_i, \ i = 1, \ldots, K \), to be the outcomes of the longitudinal Rasch model defined in Section 2.2. Our approach for estimating parameters \( \beta, \alpha \) and \( \gamma \) is as follows. Along with the basic estimating equations for the mean of \( X_i, \ i = 1, \ldots, K \), which gives estimators for the regression parameter \( \beta \), we shall use secondary equations to estimate the parameters \( \alpha \) and \( \gamma \). These equations are based on the empirical covariances which are unbiased estimators of the true covariance of the vector \( X_i \). Let \( S_i^t = (S_{ij}^t)_{t=1,\ldots,T} \) be a vector where \( S_{ij}^t \) is \( J(J-1)/2 \times 1 \) vector of empirical pairwise covariances associated with \( X_i^t \) defined by

\[ S_{ij}^t = (X_{ij}^t - \mu_{ij}^t)(X_{il}^t - \mu_{il}^t) \tag{11} \]

and \( W_i^t = (W_{ij}^{th})_{1 \leq t < h \leq T} \), a vector with elements \( W_{ij}^{th} \) be a \( J(J-1)/2 \times 1 \) vector of empirical pairwise covariance associated with \( (X_i^t, X_j^h) \), defined by

\[ W_{ij}^{th} = (X_{ij}^t - \mu_{ij}^t)(X_{il}^h - \mu_{il}^h). \tag{12} \]

To apply the GEE approach, we consider the vectors \( (X_i, S_i, W_i), \ i = 1, \ldots, K \), as observations with means

\[
\mu = E(X_i) = (\mu_{ij}^t)_{t=1,\ldots,T}, \quad \eta = E(S_i) = (\eta_{ij}^t)_{t=1,\ldots,T}, \quad \upsilon = E(W_i) = (\upsilon_{ij}^{th})_{1 \leq t < h \leq T},
\]

where

\[
\mu_{ij}^t = E(X_{ij}^t) = (\mu_{ij}^t)_{j=1,\ldots,J}, \quad \eta_{ij}^t = (\eta_{ijk}^t)_{1 \leq j < k \leq J}, \quad \upsilon_{ij}^{th} = (\upsilon_{ijk}^{th})_{1 \leq j < k \leq J}
\]

and with covariance matrix

\[
V = \begin{pmatrix}
V_{11} & V_{12} & V_{13} \\
V_{21} & V_{22} & V_{23} \\
V_{31} & V_{32} & V_{33}
\end{pmatrix}, \tag{14}
\]

where

\[
V_{11} = \text{Var}(X_i), \quad V_{22} = \text{Var}(S_i), \quad V_{33} = \text{Var}(W_i), \\
V_{12} = \text{Cov}(X_i, S_i), \quad V_{13} = \text{Cov}(X_i, W_i), \quad V_{23} = \text{Cov}(S_i, W_i).
\]
The quantities \( t^i_j \), \( t^j_k \), \( \vdots \) are, respectively, given by Eqs. (6), (9) and (10). The elements of the matrix \( V \) are computed using the joint moments of \( X_i \) up to order four given in Proposition 1. The elements of \( V_{11} \) are obtained by formulas (6) and (9). The computation of \( V_{12}, V_{13}, V_{21}, \) and \( V_{31} \) requires the joint moments up to order three and the elements of the others matrices requires the joint moments up to order four.

The parameters \((\hat{\beta}, \hat{\alpha}, \hat{\gamma})\) will be estimated by the solution of the following generalized estimating equations (GEE):

\[
U(\beta, \alpha, \gamma) = D'V^{-1} \sum_{i=1}^{K} \zeta_i = 0, \tag{15}
\]

where

\[
\zeta_i = \left( \begin{array}{c} X_i - \mu \\ S_i - \eta \\ W_i - \nu \end{array} \right), \quad D = \left( \begin{array}{ccc} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{array} \right) 
\]

with

\[
D_{11} = \frac{\partial \mu}{\partial \beta}, \quad D_{12} = \frac{\partial \mu}{\partial \alpha}, \quad D_{13} = \frac{\partial \mu}{\partial \gamma}, \quad D_{21} = \frac{\partial \eta}{\partial \beta}, \\
D_{22} = \frac{\partial \eta}{\partial \alpha}, \quad D_{23} = \frac{\partial \eta}{\partial \gamma}, \quad D_{31} = \frac{\partial \nu}{\partial \beta}, \quad D_{32} = \frac{\partial \nu}{\partial \alpha}, \quad D_{33} = \frac{\partial \nu}{\partial \gamma}. 
\]

All these derivative matrices are given in the Appendix A.4.

Let \((\hat{\beta}, \hat{\alpha}, \hat{\gamma})\) be the root of Eq. (15).

4.2. Asymptotic properties

As a consequence of the theorem of Liang and Zeger (1986), Prentice (1988), and Feddag et al. (2003), we derive the following asymptotic properties of the estimator \((\hat{\beta}, \hat{\alpha}, \hat{\gamma})\).

- \((\hat{\beta}, \hat{\alpha}, \hat{\gamma})\) is consistent for \((\beta, \alpha, \gamma),\)
- \(K^{1/2} \{ (\hat{\beta} - \beta)', (\hat{\alpha} - \alpha)', (\hat{\gamma} - \gamma)' \} \to N(0, W), \tag{16}\)

where

\[
W = \lim_{K \to \infty} \frac{1}{K} (A_1^{-1} A_2 A_1^{-1}) \tag{17}
\]

with

\[
A_1 = D' V^{-1} D, \quad A_2 = D' V^{-1} \left( \sum_{i=1}^{K} \text{Cov}(\zeta_i) \right) V^{-1} D. 
\]
The covariance matrix $W$ is consistently estimated by

$$
\hat{W} = \lim_{K \to \infty} \frac{1}{K} (\hat{A}_1^{-1} \hat{A}_2 \hat{A}_1^{-1}),
$$

where

$$
\hat{A}_1 = \hat{D}' \hat{V}^{-1} \hat{D}, \quad \hat{A}_2 = \hat{D}' \hat{V}^{-1} \left( \sum_{i=1}^{K} \hat{\xi}_i \hat{\xi}_i' \right) \hat{V}^{-1} \hat{D},
$$

and $\hat{D}, \hat{V}$ are, respectively, the values of $D$ and $V$ at $(\hat{\beta}, \hat{\xi}, \hat{\gamma})$.

To compute $(\hat{\beta}, \hat{\xi}, \hat{\gamma})$, we use the Fisher-scoring algorithm. The iterative procedure at step $(s + 1)$ is given by

$$
\begin{pmatrix}
\hat{\beta}^{(s+1)} \\
\hat{\xi}^{(s+1)} \\
\hat{\gamma}^{(s+1)}
\end{pmatrix} = \begin{pmatrix}
\hat{\beta}^{(s)} \\
\hat{\xi}^{(s)} \\
\hat{\gamma}^{(s)}
\end{pmatrix} + \frac{1}{K} (\hat{D}' \hat{V}^{-1} \hat{D})^{-1} \left( \hat{D}' \hat{V}^{-1} \sum_{i=1}^{K} \hat{\xi}_i \right),
$$

where $\hat{D}, \hat{V}, \hat{\xi}_i$ are, respectively, the values of $D, V$ and $\xi_i$ at $(\hat{\beta}^{(s)}, \hat{\xi}^{(s)}, \hat{\gamma}^{(s)})$.

### 5. Illustrations

We will illustrate the methods defined before with simulations and with a real data analysis from a quality of life study.

#### 5.1. Simulation results

In this section, we study the sample performance in estimating the fixed effects parameters $\beta$ and the parameters $\alpha$ and $\gamma$ of the random effects by the GEE approach. We performed 500 replications of the longitudinal Rasch model with parameters fixed as $J = 6$, $T = 3$, $\beta = (-2, -1, -0.5, 0.5, 1, 2)$, two different covariance matrices of the random effects defined as follows:

$$
\Sigma_1 = \begin{pmatrix}
0.4 & 0.1 & 0.2 \\
0.1 & 0.5 & 0.3 \\
0.2 & 0.3 & 0.6
\end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix}
0.8 & 0.5 & 0.6 \\
0.5 & 0.9 & 0.7 \\
0.6 & 0.7 & 1
\end{pmatrix}
$$

and two sample size $K = 100, 300$.

We denote by $\alpha_1 = (0.4, 0.5, 0.6)$, $\gamma_1 = (0.1, 0.2, 0.3)$ the parameters of $\Sigma_1$ and by $\alpha_2 = (0.8, 0.9, 1)$, $\gamma_2 = (0.5, 0.6, 0.7)$ the parameters of $\Sigma_2$.

The software package Splus is used for the simulations.

The simulation results for the classical mixed Rasch model (see Feddag et al., 2002, 2003), shows that the specification of the third and fourth joint moments of the variables do not improve the estimates of different parameters. Then we have considered in this case $V$ matrix where all elements containing 3rd and 4th order joint moments are supposed equal.
Table 1
Simulation results for the longitudinal Rasch model with parameters $\beta$, $\alpha_1$ and $\gamma_1$

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\hat{\beta}$ Mean</th>
<th>$\hat{\alpha}_1$ Mean</th>
<th>$\hat{\gamma}_1$ Mean</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$-2.027$</td>
<td>$-1.001$</td>
<td>$-0.492$</td>
<td>0.204</td>
<td>0.147</td>
<td>0.139</td>
<td>0.145</td>
<td>0.138</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$-2.007$</td>
<td>$-0.996$</td>
<td>$-0.497$</td>
<td>0.104</td>
<td>0.089</td>
<td>0.079</td>
<td>0.073</td>
<td>0.084</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Simulation results for the longitudinal Rasch model with parameters $\beta$, $\alpha_2$ and $\gamma_2$

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\hat{\beta}$ Mean</th>
<th>$\hat{\alpha}_2$ Mean</th>
<th>$\hat{\gamma}_2$ Mean</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
<th>s.e</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$-2.029$</td>
<td>$-0.982$</td>
<td>$-0.478$</td>
<td>0.241</td>
<td>0.172</td>
<td>0.160</td>
<td>0.155</td>
<td>0.174</td>
<td>0.242</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$-2.035$</td>
<td>$-0.981$</td>
<td>$-0.472$</td>
<td>0.133</td>
<td>0.102</td>
<td>0.085</td>
<td>0.085</td>
<td>0.098</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

zero. The different matrices $V_{jl}, j, l = 1, 2, 3$, are defined as follows:

- $V_{11}$ matrix with elements obtained by formulas (6) and (9).
- $V_{12}$ where the elements $\text{Cov}(X_{ij}^t, S_{i,kl}^i) = \text{Cov}(X_{ij}^t, S_{i,kl}^h) = 0$ for $j \neq k, l$ and $t \neq h$.
- $V_{13}$ where $\text{Cov}(X_{ij}^t, W_{i,kl}^sh) = 0$ ($t \neq s \neq h$), $\text{Cov}(X_{ij}^t, W_{i,kl}^{th}) = \text{Cov}(X_{ij}^t, W_{i,kl}^{ht}) = 0$ ($j \neq k, l$).
- $V_{22}$ and $V_{33}$ are considered diagonal matrices.
- $V_{23} = 0$.

We give in Tables 1 and 2 the mean and the standard error (s.e) of the 500 estimates values obtained for each parameter.

From Tables 1 and 2, its seen that all estimates are unbiased. Table 1 shows that the bias for GEE estimation of both the fixed effects and component of variance relative to their true values is less than 2 per cent. From Table 2, we note that the largest bias for the fixed effects estimates is equal to 3 per cent and less than 3 per cent for the variance components estimates. We note non-surprisingly that the standard errors (s.e), of all estimates increases slightly as the elements of the covariance matrix $\Sigma$ increase or $K$ decreases. These results confirm the theoretical asymptotic properties of these estimators when the elements of the
Table 3
Description of the emotional behavior subscale of SIP (\(K = 131\), \(J = 6\), \(T = 3\))

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.

Table 4
Parameter estimates for \((\beta, x, \gamma)\) and their standard errors (s.e)

<table>
<thead>
<tr>
<th>(\hat{\beta})</th>
<th>(0.768)</th>
<th>(0.937)</th>
<th>(0.836)</th>
<th>(0.292)</th>
<th>(1.561)</th>
<th>(0.357)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e</td>
<td>(0.103)</td>
<td>(0.115)</td>
<td>(0.111)</td>
<td>(0.103)</td>
<td>(0.161)</td>
<td>(0.070)</td>
</tr>
</tbody>
</table>

\((\hat{x}, \hat{\gamma})\)

<table>
<thead>
<tr>
<th>(\hat{x})</th>
<th>(0.852)</th>
<th>(1.091)</th>
<th>(0.734)</th>
<th>(0.183)</th>
<th>(-0.254)</th>
<th>(-0.240)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s.e</td>
<td>(0.249)</td>
<td>(0.275)</td>
<td>(0.249)</td>
<td>(0.210)</td>
<td>(0.178)</td>
<td>(0.168)</td>
</tr>
</tbody>
</table>

covariance matrix are small. We note also that this approach is less intensive with large value of \(K\).

5.2. Example

In this section, we illustrate the application of this methodology to the analysis from a quality of life data. The sample is composed of 470 depressive patients who answer at 3 occasions to a French version of the sickness impact profile (SIP) questionnaire (see Bergner et al. (1973) for the international version). This questionnaire is composed of 12 subscales of items, each related to a particular aspect from a quality of life. We present here the result for the Emotional Behavior dimension (the items contents are given in Table 3). Each person have to choose the items that describe surely his actual situation and that are related to his health. If an item is checked then we code the response 1. So this subscale measure the decrease of the quality of life. Only 131 individuals have responded at all three occasions.

In Table 4, we report the estimations of different parameters and their standard errors (s.e) based on expression (18).

From Table 4, we note that the most difficult item is number 5 (with largest estimation) and the easiest is item 4 (with smaller estimation). All measures of quality of life are correlated. The first and the second random effects are positively correlated and the correlation between the third and the two first random effects are negative.

6. Summary and discussions

In this paper, we used the GEE2 approach for the longitudinal mixed Rasch models. This method has been used by several authors for the generalized linear models with fixed effects (GLMs). Zeger et al. (1988) have used this approach to estimate only the regression parameter in GLMMs through the use of approximations to the mean and the covariance
matrix. Breslow and Clayton (1993), and Breslow and Lin (1995) have used penalized quasi-likelihood approach and the marginal quasi-likelihood method to estimate the regression parameters and the univariate component of the random effects in GLMMs. Recently Sutradhar and Rao (2001) have proposed approximations of the joint moments up to order four to estimate separately the regression parameters and the small univariate component of the random effects. Feddag et al. (2002) have proposed estimating equations to estimate simultaneously the fixed effects parameters and the variance of the random effects to the classical mixed Rasch model. This paper illustrate how the estimating equations proposed by Feddag et al. (2003), can be extended to the longitudinal mixed Rasch model. The approximations of the joint moments are based on the generalizations of those used by Sutradhar and Rao (2001) to GLMMs with univariate random effects. The obtained estimators are consistent and asymptotically normal. The simulation results confirms the theoretical asymptotic properties of the estimators and the computation are less intensive. In summary, various models for longitudinal binary data have been considered. It would be interesting to further compare our current results with those of Sutradhar and Sinha (2002) models.

Acknowledgements

The authors thank the referees for valuable comments and suggestions leading to the improvement of the paper.

Appendix A.

A.1. Approximation of the marginal likelihood

**Theorem A.1.** Under the assumption made in (5), the marginal likelihood of \( X \) given by Eq. (4), can be approximated by

\[
L^*(\beta, \alpha, \gamma \mid x) = \prod_{i=1}^{K} L^*_i(\beta, \alpha, \gamma \mid x_i)
= \prod_{i=1}^{K} (1 + P_T(x_i, \beta, \alpha, \gamma)) \prod_{i=1}^{T} \prod_{j=1}^{J} g_{ij}^*(x_{ij}; \beta_j),
\]

where

\[
P_T(x_j, \beta, \alpha, \gamma) = \int_{\mathbb{R}^T} \sum_{t=1}^{T} \sum_{1 \leq t_1 < \ldots < t_T \leq T} h_{i_1}^T(x_j, \theta_{i_1}, \beta) \ldots h_{i_T}^T
\]
\[
\times (x_j, \theta_i, \beta) \phi(\theta_i, \alpha, \gamma) \, d\theta_i
\]
with

\[ h_{ij}^i(x_i, \theta_i, \beta) = A^i_1 \theta_{itj} + \frac{1}{2} R_{ij}^i \theta_{itj}^2 + \frac{1}{6} P_{ij}^i \theta_{itj}^3 + \frac{1}{24} Q_{ij}^i \theta_{itj}^4, \]

\[ g_{ij}^i(x_{ij}^i; \beta_j) = \exp(-x_{ij}^i \beta_j - \ln(1 + e^{-\beta_j})), \]

\[ A_{i,j}^i = x_{ij}^i - a_{j}^{(1)}, \quad B_j = a_{j}^{(2)}, \quad C_j = a_{j}^{(3)}, \quad F_j = a_{j}^{(4)}, \]

\[ R_{i,j}^i = (A_{i,j}^i)^2 - B_j, \quad P_{i,j}^i = (A_{i,j}^i)^3 - 3A_{i,j}^i B_j - C_j, \]

\[ Q_{i,j}^i = (A_{i,j}^i)^4 - 6(A_{i,j}^i)^2 B_j - 4A_{i,j}^i C_j + 3B_j^2 - F_j, \]

\[ R_i^i = (A_i^i)^2 - B, \quad P_i^i = (A_i^i)^3 - 3A_i^i B - C, \]

\[ Q_i^i = (A_i^i)^4 - 6(A_i^i)^2 B - 4A_i^i C + 3B^2 - F, \]

\[ A_i^i = \sum_{j=1}^{J} A_{i,j}^i, \quad B = \sum_{j=1}^{J} B_j, \quad C = \sum_{j=1}^{J} C_j, \quad F = \sum_{j=1}^{J} F_j, \]

\[ a_{j}^{(1)} = \frac{1}{(1 + e^{\beta_j})}, \quad a_{j}^{(2)} = \frac{e^{\beta_j}}{(1 + e^{\beta_j})^{2}}, \quad a_{j}^{(3)} = \frac{e^{\beta_j}(e^{\beta_j^{-1}})}{(1 + e^{\beta_j})^{3}}, \]

\[ a_{j}^{(4)} = \frac{e^{\beta_j}(e^{2\beta_j} - 4e^{\beta_j} + 1)}{(1 + e^{\beta_j})^{4}}. \]

The computation of the expressions \( P_T, i = 1, \ldots, K, \) requires the joint moments up to order four of the variables \( \theta_i, \ i = 1, \ldots, K, \) given for all \( t, h, u, v = 1, \ldots, T, \) by:

\[ E(\theta_{it}) = 0, \quad E(\theta_{it}^2) = \sigma_{it}, \quad E(\theta_{it}^3) = 0, \quad E(\theta_{it}^4) = 3\sigma_{it}^2, \quad E(\theta_{it} \theta_{ih}) = \sigma_{th}, \]

\[ E(\theta_{it}^2 \theta_{ih}^2) = \sigma_{it} \sigma_{hh} + 2\sigma_{th}^2, \quad E(\theta_{it} \theta_{ih}^3) = 3\sigma_{th} \sigma_{hh}, \]

\[ E(\theta_{it} \theta_{ih} \theta_{iu} \theta_{iv}) = \sigma_{ih} \sigma_{uv} + \sigma_{iu} \sigma_{hv} + \sigma_{iv} \sigma_{hu}. \]

We give the expression \( P_T(\cdot), i = 1, \ldots, K \) for three values of \( T : \ T = 1, 2, 3. \)

- \( T = 1, \Sigma = \sigma_{11} : \ P_T(x_i, \beta, \alpha) = \frac{1}{2} R_i \sigma_{11} + \frac{1}{8} Q_i \sigma_{11}^2. \)
- \( T = 2, \)

\[ P_T(x_i, \beta, \alpha, \gamma) = \frac{1}{2} (R_i \sigma_{11} + R_i^2 \sigma_{22}) + \frac{1}{8} (Q_i \sigma_{11}^2 + Q_i^2 \sigma_{22}^2) + A_i \sigma_{12}^2 \]

\[ + \frac{1}{2} (A_i P_i \sigma_{12} \sigma_{22} + P_i A_i^2 \sigma_{11} \sigma_{12}) + \frac{1}{4} R_i \sigma_{11} \sigma_{22} + 2\sigma_{12}^2. \]
• $T = 3$,

$$P_T(x_{it}; \beta, \alpha, \gamma) = \frac{1}{2}(R_i^1\sigma_{i1} + R_i^2\sigma_{i2} + R_i^3\sigma_{i3}) + \frac{1}{8}(Q_i^1\sigma_{i1}^2 + Q_i^2\sigma_{i2}^2 + Q_i^3\sigma_{i3}^2) + A_i^1A_i^2\sigma_{i12} + A_i^1A_i^3\sigma_{i13} + A_i^2A_i^3\sigma_{i23} + \frac{1}{2}[A_i^1P_i^2\sigma_{i12}\sigma_{i22} + P_i^1A_i^2\sigma_{i11}\sigma_{i12} + A_i^1P_i^3\sigma_{i11}\sigma_{i33} + P_i^1A_i^3\sigma_{i11}\sigma_{i13} + A_i^2P_i^3\sigma_{i23}\sigma_{i33} + P_i^2A_i^3\sigma_{i22}\sigma_{i23}]$$

$$+ \frac{1}{4}[R_i^1R_i^2(\sigma_{i11}\sigma_{i22} + 2\sigma_{i22}^2) + R_i^1R_i^3(\sigma_{i11}\sigma_{i33} + 2\sigma_{i13}^2) + (\sigma_{i22}\sigma_{i33} + 2\sigma_{i23}^2)R_i^2R_i^3] + \frac{1}{2}[A_i^1A_i^2R_i^3(\sigma_{i12}\sigma_{i33} + 2\sigma_{i13}\sigma_{i23}) + A_i^1R_i^2A_i^3(\sigma_{i13}\sigma_{i22} + 2\sigma_{i12}\sigma_{i23}) + R_i^1A_i^2A_i^3(\sigma_{i11}\sigma_{i22} + 2\sigma_{i12}\sigma_{i13})].$$

**Proof of Theorem A.1.** For all $i = 1, \ldots, K$, $j = 1, \ldots, J$, $t = 1, \ldots, T$, let

$$g_{ij}(x_{ij}; \theta_{it}, \beta_j) = \exp\{x_{ij}(\theta_{it} - \beta_j) - a_{ij}\},$$

where $a_{ij} = \ln(1 + \exp(\theta_{it} - \beta_j))$.

Expanding the function $g_{ij}$ in a Taylor series about $\theta_{it} = 0$ up to order four, we obtain the expression

$$g_{ij}(x_{ij}; \beta_j, \theta_{it}) = g_{ij}^*(x_{ij}; \beta_j) \left(1 + A_{ij}^1\theta_{it} + \frac{1}{2}R_{ij}^1\theta_{it}^2 + \frac{1}{6}P_{ij}^1\theta_{it}^3 + \frac{1}{24}Q_{ij}^1\theta_{it}^4 + O(\theta_{it}^5)\right),$$

where

$$g_{ij}^*(x_{ij}; \beta_j) = \exp(-x_{ij}\beta_j - \ln(1 + e^{-\beta_j})), \quad a_j = \ln(1 + e^{-\beta_j}),$$

$$A_{ij}^1 = x_{ij} - a_j^{(1)}, \quad B_j = a_j^{(2)}, \quad C_j = a_j^{(3)}, \quad F_j = a_j^{(4)},$$

$$R_{ij}^1 = (A_{ij}^1)^2 - B_j, \quad P_{ij}^1 = (A_{ij}^1)^3 - 3A_{ij}^1B_j - C_j, \quad Q_{ij}^1 = (A_{ij}^1)^4 - 6(A_{ij}^1)^2B_j - 4A_{ij}^1C_j + 3B_j^2 - F_j.$$

Then we obtain the following approximation:

$$\prod_{j=1}^{J} g_{ij}(x_{ij}; \beta_j, \theta_{it}) \simeq \prod_{j=1}^{J} g_{ij}^*(x_{ij}; \beta_j) \prod_{j=1}^{J} \left(1 + A_{ij}^1\theta_{it} + \frac{1}{2}R_{ij}^1\theta_{it}^2 + \frac{1}{6}P_{ij}^1\theta_{it}^3 + \frac{1}{24}Q_{ij}^1\theta_{it}^4\right),$$

$$= T_{i}(x_{ij}; \beta, \theta_{it}) \prod_{j=1}^{J} g_{ij}^*(x_{ij}; \beta_j),$$

(21)
where

\[ T_t(x_t; \beta, \theta_{it}) = \left( 1 + A_i^t \theta_{it} + \frac{1}{2} R_i^t \theta_{it}^2 + \frac{1}{6} P_i^t \theta_{it}^3 + \frac{1}{24} Q_i^t \theta_{it}^4 \right), \]  

(22)

\[ R_i^t = (A_i^t)^2 - B, \quad P_i^t = (A_i^t)^3 - 3A_i^t B - C, \]

\[ Q_i^t = (A_i^t)^4 - 6(A_i^t)^2 B - 4A_i^t C + 3B^2 - F, \]

\[ A_i^t = \sum_{j=1}^J A_{i,j}^t, \quad B = \sum_{j=1}^J B_j, \quad C = \sum_{j=1}^J C_j, \quad F = \sum_{j=1}^J F_j. \]

Then

\[ \prod_{t=1}^T \prod_{j=1}^J g_{ij}(x_{ij}; \beta_j) \simeq T(x_i; \beta, \theta_i) \prod_{t=1}^T \prod_{j=1}^J g_{ij}^*(x_{ij}; \beta_j), \]  

(23)

where

\[ T(x_i; \beta, \theta_i) \]

\[ = 1 + \sum_{t=1}^T \sum_{1 \leq t_1 < \ldots < t_r \leq T} \left( A_i^{t_1} \theta_{i t_1} + \frac{1}{2} R_i^{t_1} \theta_{i t_1}^2 + \frac{1}{6} P_i^{t_1} \theta_{i t_1}^3 + \frac{1}{24} Q_i^{t_1} \theta_{i t_1}^4 \right) \]

\[ \ldots \left( A_i^{t_r} \theta_{i t_r} + \frac{1}{2} R_i^{t_r} \theta_{i t_r}^2 + \frac{1}{6} P_i^{t_r} \theta_{i t_r}^3 + \frac{1}{24} Q_i^{t_r} \theta_{i t_r}^4 \right). \]  

(24)

Taking the expectation of expression (23) over \( \theta_i \), we obtain expression (19) of Theorem A.1.

A.2. Approximation of joint density

Theorem A.2. Under the assumption made in (5), we have the following

1. The joint density of \((X_{ij_1}^t, \ldots, X_{ij_m}^t)\), \(i = 1, \ldots, K, \ m = 1, \ldots, J, \ t = 1, \ldots, T\), is given by

\[ L_{i,(m)}^*(x_{ij_1}^t, \ldots, x_{ij_m}^t; \beta, \gamma) = \left( 1 + \frac{1}{2} \sigma_{hh} R_{i,j_1 \ldots j_m}^t + \frac{1}{8} \sigma_{hh}^2 Q_{i,j_1 \ldots j_m}^t \right) \]

\[ \times \prod_{l=1}^m g_{ij_l}(x_{ij_l}; \beta_{jl}). \]  

(25)
(2) The joint density of \( (X_{ij}^t, \ldots, X_{ij,m}^t, X_{ij1}^h, \ldots, X_{ij,s}^h) \), \( i = 1, \ldots, K, \ m, s = 1, \ldots, J, \ 1 < t \neq h < T \), is given by

\[
L^*_i(m,s)(x_{ij1}^t, \ldots, x_{ij,m}^t, x_{ij1}^h, \ldots, x_{ij,s}^h; \beta, \alpha, \gamma) = \prod_{l=1}^m g^{n}_{ijl}(x_{ij1}^t; \beta_{ji}) \prod_{s} g^{n}_{ij}(x_{ij1}^h; \beta_{ji}) \left[ 1 + \frac{1}{2} (\sigma_{tt} R_{i,j1}^{t} \ldots m + \sigma_{hh} R_{i,j1}^{h} \ldots j) \\
+ \sigma_{th} A_{i,j1}^{t} \ldots m A_{i,j1}^{h} \ldots j + \frac{1}{2} (A_{i,j1}^{t} \ldots m P_{i,j1}^{h} \ldots j) \sigma_{th} \sigma_{hh} \\
+ P_{i,j1}^{t} \ldots m A_{i,j1}^{h} \ldots j (\sigma_{tt} \sigma_{hh} + 2 \sigma_{th}) \\
+ \frac{1}{8} (\sigma_{tt} Q_{i,j1}^{t} \ldots m + \sigma_{hh} Q_{i,j1}^{h} \ldots j) \right].
\]

(26)

The quantities involved in this theorem are defined in Theorem A.1 and below:

\[
A_{i,j1}^{t} \ldots m = \sum_{l=1}^{m} A_{i,j1}^{l} \ldots m, \quad B_{j1} \ldots m = \sum_{l=1}^{m} B_{jl} \ldots m, \quad C_{j1} \ldots m = \sum_{l=1}^{m} C_{jl} \ldots m, \quad F_{j1} \ldots m = \sum_{l=1}^{m} F_{jl} \ldots m.
\]

\[
R_{i,j1}^{t} \ldots m = (A_{i,j1}^{t} \ldots m B_{j1} \ldots m)^2 - B_{j1} \ldots m, \quad P_{j1}^{t} \ldots m = (A_{j1}^{t} \ldots m B_{j1} \ldots m)^2 - A_{j1}^{t} \ldots m C_{j1} \ldots m.
\]

\[
Q_{i,j1}^{t} \ldots m = (A_{i,j1}^{t} \ldots m B_{j1} \ldots m)^4 - 6 (A_{i,j1}^{t} \ldots m B_{j1} \ldots m)^2 B_{j1} \ldots m
- 4 A_{i,j1}^{t} \ldots m C_{j1} \ldots m + 3 B_{j1} \ldots m - F_{j1} \ldots m.
\]

Proof of Theorem A.2. The proof of this theorem requires the following lemma.

Lemma A.1. Let \( Y \) be a random variable with density function defined by

\[
f_{Y}(y) = \exp(y\theta - a(\theta) + c(y)),
\]

where \( a(\cdot) \) and \( c(\cdot) \) are functions, respectively, of \( \theta \) and \( y \). We have the following results:

\[
E(Y) = a^{(1)}(\theta), \quad V(Y) = a^{(2)}(\theta), \quad E((Y - a^{(1)}(\theta))^3) = a^{(3)}(\theta),
\]

\[
E((Y - a^{(1)}(\theta))^4) = a^{(4)}(\theta) + 3 a^{(2)}(\theta)^2,
\]

\[
E((Y - a^{(1)}(\theta))^5) = a^{(5)}(\theta) + 10 a^{(3)}(\theta) a^{(2)}(\theta),
\]

\[
E(Y(Y - a^{(1)}(\theta)^r) = E((Y - a^{(1)}(\theta)^r + 1) E((Y - a^{(1)}(\theta)^r), \quad (r = 2, 3, 4).
\]

For each \( t = 1, \ldots, T \); \( P_t(x_{ij}^t, \beta, \alpha) = \frac{1}{2} R_{i,j1}^{t} \sigma_{tt} + \frac{1}{8} Q_{i,j1}^{t} \sigma_{tt}^2 \). Then the density of \( X_{ij}^t \) is given by

\[
L^*_i(x_{ij1}^t; \beta, \alpha, \gamma) = \prod_{j=1}^J g^{n}_{ij}(x_{ij1}^t; \beta_{ji}) \left( 1 + \frac{1}{2} \sigma_{tt} R_{i,j1}^{t} + \frac{1}{8} \sigma_{tt}^2 Q_{i,j1}^{t} \right)
\]

(27)
and the density of \((X_{t,ij}, \ldots, X_{t,ijm}^T), \ t = 1, \ldots, T, \ i = 1, \ldots, K\) is obtained by

\[
L^*_{i,(m)}(x_{t,ij}^1, \ldots, x_{t,ijm}^T; \beta, \alpha, \gamma) = \sum_{\sum_{l=0}^{t,ijm} I_l \notin \{j_1, \ldots, j_m\}} L^*_i(x_{ij}^l; \beta, \alpha, \gamma) \tag{28}
\]

by using the previous lemma for the density \(g^*_{ij}\) for all \(j \notin \{j_1, \ldots, j_m\}\), we deduce that

\[
E(R_{i,j}^t) = E(Q_{i,j}^t) = 0.
\]

These results yield expression (25) of Theorem A.2.

The joint density of \((X_{t}^i, X_{t}^h)\) is given by

\[
L^*_i(x_{t}^i, x_{t}^h; \beta, \alpha, \gamma) = \prod_{l=1}^J g^*_{ij}(x_{ij}^l; \beta_{jl}) g^*_{ij}(x_{ij}^h; \beta_{jh})
\times \left[1 + \frac{1}{2} (\sigma_{ii} R_{i}^t + \sigma_{hh} R_{h}^t) + \sigma_{ii} A_{i}^l A_{i}^h
+ \frac{1}{2} (A_{i}^l P_{i}^h \sigma_{hh} + P_{i}^l A_{i}^h \sigma_{ii})
+ \frac{1}{4} R_{i}^t R_{h}^t (\sigma_{ii} \sigma_{hh} + 2 \sigma_{ih}) + \frac{1}{8} (\sigma_{ii}^2 Q_{i}^t + \sigma_{hh}^2 Q_{h}^t)\right]. \tag{29}
\]

The density of \((X_{t,ij}^1, \ldots, X_{t,ijm}^T, X_{t,j1}^h, \ldots, X_{t,jk}^h)\), is obtained by

\[
L^*_{i,(m,s)}(x_{t,ij}^1, \ldots, x_{t,ijm}^T, x_{t,j1}^h, \ldots, x_{t,jk}^h; \beta, \alpha, \gamma)
= \sum_{\sum_{l=0}^{t,ijm} I_l \notin \{j_1, \ldots, j_m\}} \sum_{\sum_{r=0}^{t,jk} I_r \notin \{j_1, \ldots, j_k\}} L^*_i(x_{ij}^l; x_{ij}^h; \beta, \alpha, \gamma). \tag{30}
\]

By using the previous lemma for the density \(g^*_{ij}\) for all \(j \notin \{j_1, \ldots, j_m, j_1, \ldots, j_k\}\), we deduce that

\[
E(R_{i,j}^t) = E(R_{i,j}^h) = E(Q_{i,j}^t) = E(Q_{i,j}^h) = 0.
\]

These results yield expression (26) of Theorem A.2.

A.3. Proof of Proposition 1

We obtain the assertions of this Proposition by substituting the derivatives \(a_{ij}^{(r)}\), \(r = 1, \ldots, 4\), by their explicit expressions given by Theorem A.1 in the
following equations:

\[ E(X_{ij}^t) = \sum_{x_{ij}^t=0}^1 x_{ij}^t L_{i,(1)}^*(x_{ij}^t; \beta, \alpha, \gamma) = L_{i,(1)}^*(1; \beta, \alpha, \gamma), \]

\[ E(X_{ij_1}^t \ldots X_{ij_m}^t) = \sum_{x_{ij_1}^t=0}^1 \ldots \sum_{x_{ij_m}^t=0}^1 x_{ij_1}^t \ldots x_{ij_m}^t L_{i,(m)}^*(x_{ij_1}^t, \ldots, x_{ij_m}^t; \beta, \alpha, \gamma) = L_{i,(m)}^*(1, \ldots, 1; \beta, \alpha, \gamma), \]

\[ E(X_{ij_1}^t \ldots X_{ij_m}^t X_{ij_1}^h \ldots X_{ij_s}^h) = L_{i,(m,s)}^*(1, \ldots, 1, 1, \ldots, 1; \beta, \alpha, \gamma). \]

A.4. Derivative of the D matrix

- The matrices \( D_{11}, D_{12} \) and \( D_{13} \) are given by

\[
D_{11} = (D_{11}^1, \ldots, D_{11}^T)', \quad D_{13} = 0,
\]

where \( D_{11}^t, \ (t = 1, \ldots, T) \) is diagonal matrix defined by

\[
D_{11}^t = \text{diag}(\frac{\partial \mu_j^t}{\partial \beta_j})_{j=1,\ldots,J}
\]

with

\[
\frac{\partial \mu_j^t}{\partial \beta_j} = -\frac{e^{\beta_j}}{(1 + e^{\beta_j})^2} + \frac{\sigma_{tt}}{2} \frac{-e^{3\beta_j} + 4e^{2\beta_j} - e^{\beta_j}}{(1 + e^{\beta_j})^4} + \frac{\sigma_{tt}^2}{8} \frac{-e^{5\beta_j} + 26e^{4\beta_j} - 66e^{3\beta_j} + 26e^{2\beta_j} - e^{\beta_j}}{(1 + e^{\beta_j})^6}
\]

and \( D_{12} = \text{diag}(D_{12}^t)_{t=1,\ldots,T} \) is block diagonal matrix with column vector \( D_{12}^t = (\frac{\partial \mu_j^t}{\partial \sigma_{tt}})_{j=1,\ldots,J} \) defined by

\[
\frac{\partial \mu_j^t}{\partial \sigma_{tt}} = \frac{e^{\beta_j}(e^{\beta_j} - 1)}{2(1 + e^{\beta_j})^3} + \frac{\sigma_{tt}}{4} \frac{e^{\beta_j}(e^{3\beta_j} - 11e^{2\beta_j} + 11e^{\beta_j} - 1)}{(1 + e^{\beta_j})^5}.
\]

- The matrices \( D_{21}, D_{22} \) and \( D_{23} \) are given by

\[
D_{21} = (D_{21}^1, \ldots, D_{21}^T)', \quad D_{22} = (D_{22}^1, \ldots, D_{22}^T)', \quad D_{23} = 0,
\]

where

\[
D_{21}^t = \left( \frac{\partial \eta_{jl}^t}{\partial \beta_k} \right)_{1 \leq j < l \leq J}^{1 \leq j \leq J}, \quad D_{22}^t = \left( \frac{\partial \eta_{jl}^t}{\partial \sigma_{tt}} \right)_{1 \leq j < l \leq J}^{1 \leq j \leq J}
\]
with
\[
\frac{\partial n'_{jl}}{\partial \beta_j} = \sigma_{tt} \frac{e^{\beta_j} e^{\beta_l} (1 - e^{\beta_j})}{(1 + e^{\beta_j})^3(1 + e^{\beta_l})^2} + \frac{\sigma_{tt}^2}{2} (Q_{11} + Q_{12} + Q_{13}),
\]
\[
\frac{\partial n'_{jl}}{\partial \beta_l} = \sigma_{tt} \frac{e^{\beta_j} e^{\beta_l} (1 - e^{\beta_l})}{(1 + e^{\beta_j})^2(1 + e^{\beta_l})^3} + \frac{\sigma_{tt}^2}{2} (Q_{21} + Q_{22} + Q_{23}),
\]
\[
\frac{\partial \sigma_{jl}}{\partial \beta_k} = 0 \text{ if } k \neq j, l,
\]
\[
\frac{\partial n'_{jl}}{\partial \sigma_{tt}} = \frac{e^{\beta_j} e^{\beta_l}}{(1 + e^{\beta_j})^2(1 + e^{\beta_l})^2} + \sigma_{tt} (Q_{31} + Q_{32} + Q_{33}).
\]

- The matrices $D_{31}$, $D_{32}$ and $D_{33}$ are given by
\[
D_{31} = (D_{31}^{th})'_{1 \leq t < h \leq T}, \quad D_{32} = \left( \frac{\partial v'^h}{\partial \sigma_{ss}} \right)_{s = 1, \ldots, T}^{1 \leq t < h \leq T}, \quad D_{33} = (D_{33}^{th})'_{1 \leq t < h \leq T},
\]
where
\[
D_{31}^{th} = \left( \frac{\partial v'^h_{jl}}{\partial \beta_k} \right)_{1 \leq j < l \leq J}^{1 \leq j = l \leq J}, \quad D_{33}^{th} = \left( \frac{\partial v'^h_{jl}}{\partial \sigma_{th}} \right)_{1 \leq j < l \leq J}^{1 \leq j = l \leq J}
\]
with
\[
\frac{\partial v'^h_{jl}}{\partial \beta_j} = \sigma_{th} \frac{e^{\beta_j} e^{\beta_l} (1 - e^{\beta_j})}{(1 + e^{\beta_j})^3(1 + e^{\beta_l})^2} + \frac{1}{2} (\sigma_{tt} \sigma_{th} Q_{11} + \sigma_{tt}^2 Q_{12} + \sigma_{hh} \sigma_{th} Q_{13}),
\]
\[
\frac{\partial v'^h_{jl}}{\partial \beta_l} = \sigma_{th} \frac{e^{\beta_j} e^{\beta_l} (1 - e^{\beta_l})}{(1 + e^{\beta_j})^2(1 + e^{\beta_l})^3} + \frac{1}{2} (\sigma_{tt} \sigma_{th} Q_{21} + \sigma_{tt}^2 Q_{22} + \sigma_{hh} \sigma_{th} Q_{23}),
\]
\[
\frac{\partial \sigma_{jl}}{\partial \beta_k} = 0 \text{ if } k \neq j, l,
\]
\[
\frac{\partial v'^h_{jl}}{\partial \sigma_{ss}} = \begin{cases} \sigma_{th} \frac{(e^{3\beta_j} - 4e^{2\beta_j} + e^{\beta_j})e^{\beta_l}}{(1 + e^{\beta_j})^4(1 + e^{\beta_l})^2} & \text{if } s = t, \\ \sigma_{th} \frac{(e^{3\beta_l} - 4e^{2\beta_l} + e^{\beta_l})e^{\beta_j}}{(1 + e^{\beta_l})^2(1 + e^{\beta_j})^4} & \text{if } s = h, \\ 0 & \text{if } s \neq t, h, \end{cases}
\]
\[
\frac{\partial v'^h_{jl}}{\partial \sigma_{th}} = \frac{e^{\beta_j} e^{\beta_l}}{(1 + e^{\beta_j})^2(1 + e^{\beta_l})^2} + \frac{1}{2} (\sigma_{tt} Q_{31} + 2\sigma_{th} Q_{32} + \sigma_{hh} Q_{33}).
\]
and the quantities $Q_{jk}$ ($j, k = 1, 2, 3$), are given by

$$Q_{11} = -e^{\beta_j} e^{\beta_i} (e^{3\beta_j} - 11e^{2\beta_j} + 11e^{\beta_j} - 1),$$

$$Q_{12} = e^{\beta_j} (1 - e^{\beta_j})(e^{3\beta_i} - 4e^{2\beta_i} + e^{\beta_i}),$$

$$Q_{13} = e^{\beta_j}(-e^{2\beta_j} + 4e^{\beta_j} - 1)e^{\beta_i}(e^{\beta_i} - 1),$$

$$Q_{21} = -e^{\beta_i} e^{\beta_j} (e^{3\beta_i} - 11e^{2\beta_i} + 11e^{\beta_i} - 1),$$

$$Q_{22} = e^{\beta_i} (1 - e^{\beta_i})(e^{3\beta_j} - 4e^{2\beta_j} + e^{\beta_j}),$$

$$Q_{23} = e^{\beta_i}(-e^{2\beta_i} + 4e^{\beta_i} - 1)e^{\beta_j}(e^{\beta_j} - 1),$$

$$Q_{31} = e^{\beta_j} e^{\beta_i} (e^{2\beta_j} - 4e^{\beta_j} + 1),$$

$$Q_{32} = e^{\beta_j} e^{\beta_i} (1 - e^{\beta_j})(1 - e^{\beta_i}),$$

$$Q_{33} = e^{\beta_j} e^{\beta_i} (e^{2\beta_i} - 4e^{\beta_i} + 1).$$

References


Multivariate Anal. 76, 1–34.