Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2002/08

Alternative specifications of bivariate multilevel probit ordinal response models

Leonardo Grilli, Carla Rampichini

Università degli Studi
di Firenze

Applied Statistics

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Alternative specifications of bivariate multilevel probit ordinal response models

Leonardo Grilli
Associate research assistant
Dept. of Statistics "G. Parenti", University of Florence
Viale Morgagni, 59, 50134 Florence, Italy
Tel: +39 0554237271
Fax: +39 0554223560
e-mail: grilli@ds.unifi.it

Carla Rampichini
Professor of Statistics
Dept. of Statistics "G. Parenti", University of Florence
Viale Morgagni, 59, 50134 Florence, Italy
Tel: +39 0554237246
Fax: +39 0554223560
e-mail: carla@ds.unifi.it

Abstract

In the last years several methods for the analysis of ordinal multivariate multilevel data have been proposed (Muthén, 1994; Rabe-Hesketh *et al.*, 2001; Mazzolli, 2001; Lillard and Panis 2000). The present paper highlights the interpretation of the variance-covariance parameters of the assumed multivariate distribution of the latent variables. Moreover, under the hypothesis of a multivariate Gaussian distribution, the paper illustrates some alternative specifications of the model, which have been proposed in order to use certain estimation algorithms yet implemented in the existing statistical software.

1 Basic specification of the model

Let $Y_{ij}^{(h)}$ be the h-th observed ordinal variable ($h = 1, 2, \cdots, H$) for the i-th subject ($i = 1, 2, \cdots n_j$) of the j-th cluster ($j = 1, 2, \cdots J$). In our application the clusters are the courses, the subjects are the questionnaires and the ordinal variables are the ratings on two items of the questionnaire (i.e. $H = 2$). Now assume that each of the observed ratings $Y_{ij}^{(h)}$, which takes values in $\{1, 2, \ldots, C\}$ (letting C be the same for all h for simplicity), is generated by a latent variable $\tilde{Y}_{ij}^{(h)}$ through the following relationship:

$$\left\{Y_{ij}^{(h)} = c^{(h)}\right\} \quad \Leftrightarrow \quad \left\{\gamma_{c^{(h)}-1}^{(h)} < \tilde{Y}_{ij}^{(h)} \leq \gamma_{c^{(h)}}^{(h)}\right\},$$

where the thresholds satisfy $-\infty = \gamma_0^{(h)} \leq \gamma_1^{(h)} \leq \ldots \leq \gamma_{C-1}^{(h)} \leq \gamma_C^{(h)} = +\infty$.

Now let us consider the following two-level null model for the latent variables:

$$\begin{aligned} \tilde{Y}_{ij}^{(1)} &= \alpha^{(1)} + u_j^{(1)} + \varepsilon_{ij}^{(1)} \\ \tilde{Y}_{ij}^{(2)} &= \alpha^{(2)} + u_j^{(2)} + \varepsilon_{ij}^{(2)}, \end{aligned} \tag{1}$$

where, for each h, $\alpha^{(h)}$ is the mean, $u_j^{(h)}$ is the cluster's random effect (level two error) and $\varepsilon_{ij}^{(h)}$ is the subject's disturbance (level one error). The errors are assumed to be distributed as

$$\begin{aligned} \left(\varepsilon_{ij}^{(1)}, \varepsilon_{ij}^{(2)}\right)' &\overset{iid}{\sim} N(\mathbf{0}, \mathbf{\Sigma}_\varepsilon) \\ \left(u_j^{(1)}, u_j^{(2)}\right)' &\overset{iid}{\sim} N(\mathbf{0}, \mathbf{\Omega}) \end{aligned} \tag{2}$$

with

$$\boldsymbol{\Sigma}_\varepsilon = \begin{pmatrix} \sigma^2_{\varepsilon_1} & \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \tau_1^2 & \\ \tau_{12} & \tau_2^2 \end{pmatrix}; \tag{3}$$

moreover the first and second level errors are assumed to be independent, so $\mathrm{Cov}(\varepsilon_{ij}^{(h)}, u_j^{(k)}) = 0, \forall i, j, h, k$.

The previous model specification implies the following conditional covariance structure for the two latent variables $\tilde{Y}_{ij}^{(h)}$:

$$\begin{aligned} \mathrm{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)} \mid u_j^{(h)}, u_j^{(k)}) &= \mathrm{E}(\varepsilon_{ij}^{(h)} \varepsilon_{i'j'}^{(k)}) \\ &= \begin{cases} \sigma^2_{\varepsilon_h} & \text{if} \quad k = h, j = j', i = i' \\ \sigma_{\varepsilon_1\varepsilon_2} & \text{if} \quad k \neq h, j = j', i = i' \\ 0 & \qquad\qquad \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

Then the ratio $\sigma_{\varepsilon_1\varepsilon_2}/\sigma_{\varepsilon_1}\sigma_{\varepsilon_2}$ can be interpreted as the conditional polychoric correlation (see Drasgow, 1981).

The unconditional covariance structure is:

$$\mathrm{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)}) = \mathrm{E}(\varepsilon_{ij}^{(h)} \varepsilon_{i'j'}^{(k)}) + \mathrm{E}(u_j^{(h)} u_{j'}^{(k)}), \tag{5}$$

with $\mathrm{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)}) = 0$ if $j \neq j'$, while the expression of $\mathrm{Cov}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j'}^{(k)})$ for $j = j'$ is reported in Table 1.

Table 1: $\mathrm{Cov}(\tilde{y}_{ij}^{(h)}, \tilde{y}_{i'j'}^{(k)})$ for $j = j'$ (same cluster).

	$i = i'$	$i \neq i'$
$h = k$	$\sigma^2_{\varepsilon_h} + \tau_h^2$	τ_h^2
$h \neq k$	$\sigma_{\varepsilon_1\varepsilon_2} + \tau_{12}$	τ_{12}

¿From the expressions reported in Table 1, three types of correlation can be defined:

- the correlation between the same variable for two distinct subjects of the same cluster, that is the intraclass correlation coefficient, ICC, representing also the proportion of variance explained by the clusters:

$$\mathrm{Corr}(\tilde{Y}_{ij}^{(h)}, \tilde{Y}_{i'j}^{(h)}) = \tau_h^2/(\sigma^2_{\varepsilon_h} + \tau_h^2) \qquad h = 1, 2;$$

- the correlation between the two variables for the same subject (marginal polychoric correlation):

$$\mathrm{Corr}(\tilde{Y}_{ij}^{(1)}, \tilde{Y}_{ij}^{(2)}) = (\sigma_{\varepsilon_1\varepsilon_2} + \tau_{12}) / \sqrt{(\sigma_{\varepsilon_1}^2 + \tau_1^2)(\sigma_{\varepsilon_2}^2 + \tau_2^2)};$$

- the correlation between the two variables for two distinct subjects of the same cluster:

$$\mathrm{Corr}(\tilde{Y}_{ij}^{(1)}, \tilde{Y}_{i'j}^{(2)}) = \tau_{12} / \sqrt{(\sigma_{\varepsilon_1}^2 + \tau_1^2)(\sigma_{\varepsilon_2}^2 + \tau_2^2)}.$$

The cluster random effects $u_j^{(1)}$ and $u_j^{(2)}$ may be viewed as factors, so the model described so far may be interpreted as a two-factor model. The one-factor version is obtained by specifying

$$u_j^{(h)} = \lambda_h w_j, \qquad h = 1, 2,$$

where $w_j \overset{iid}{\sim} N(0,1)$ and the λ_h's are parameters. In this case $u_j^{(1)}$ and $u_j^{(2)}$ have a distinct variances, but they are perfectly correlated. The unconditional covariances (5) are easily derived posing $\tau_h = \lambda_h$ and $\tau_{12} = \lambda_1\lambda_2$.

To make the ordinal model identifiable, it is necessary to impose some constraints: in the following we assume $\gamma_1^{(1)} = \gamma_1^{(2)} = 0$ and $\sigma_{\varepsilon_1} = \sigma_{\varepsilon_2} = 1$. Note that the model has four estimable variance-covariance parameters, three at cluster level (τ_1^2, τ_2^2, τ_{12}) and one at subject level ($\sigma_{\varepsilon_1\varepsilon_2}$, the conditional polychoric correlation). In the following we denote with θ the set of all estimable parameters.

The full model likelihood can be derived in the following steps. First, the conditional likelihood for subject i of cluster j is

$$L_{ij}(\boldsymbol{\theta} \,|\, \mathbf{u}) = \prod_{\mathbf{c} \in \mathcal{C}} \left[P\left(\bigcap_{h=1}^{2} \left\{ Y_{ij}^{(h)} = c^{(h)} \right\} \;\mid\; u^{(1)}, u^{(2)} \right) \right]^{d_{ij\mathbf{c}}}, \qquad (6)$$

where $\mathcal{C}$ is the set of all admissible values of the vector $\mathbf{c} = (c^{(1)}, c^{(2)})$ and $d_{ij\mathbf{c}}$ is the indicator function of the event $\bigcap_{h=1}^{2}\{Y_{ij}^{(h)} = c^{(h)}\}$. Note that the relationship between the observed and latent variables and the hypotheses on the latent model imply that

$$P\left(\bigcap_{h=1}^{2} \left\{ Y_{ij}^{(h)} = c^{(h)} \right\} \;\mid\; u^{(1)}, u^{(2)} \right) \qquad (7)$$

$$= P\left(\bigcap_{h=1}^{2}\left\{\gamma^{(h)}_{c^{(h)}-1} < \tilde{Y}^{(h)}_{ij} \le \gamma^{(h)}_{c^{(h)}}\right\} \mid u^{(1)}, u^{(2)}\right)$$
$$= E_{\boldsymbol{\varepsilon}}\left[\prod_{h=1}^{2} I\left\{\gamma^{(h)}_{c^{(h)}-1} - \alpha^{(h)} - u^{(h)} < \varepsilon^{(h)} \le \gamma^{(h)}_{c^{(h)}} - \alpha^{(h)} - u^{(h)}\right\} \mid u^{(1)}, u^{(2)}\right],$$

where $\boldsymbol{\varepsilon} = (\varepsilon^{(1)}, \varepsilon^{(2)})$; therefore, computation of the probability involves an integral with respect to a bivariate Gaussian density.

Second, the marginal likelihood for cluster j is

$$L_j(\boldsymbol{\theta}) = E_{\mathbf{u}}\left[\prod_{i=1}^{n_j} L_{ij}(\boldsymbol{\theta} \mid \mathbf{u})\right], \tag{8}$$

involving another integral with respect to a bivariate Gaussian density. Finally, the overall marginal likelihood is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{J} L_j(\boldsymbol{\theta}). \tag{9}$$

Maximization of the marginal likelihood (9) requires the solution of the double integrals at subject and cluster levels. The NLMIXED procedure of the SAS system (SAS Institute, 1999), which allows to specify an arbitrary conditional likelihood programmable with SAS statements, can do the job: in fact the probabilities (7) can be written using the bivariate Gaussian distribution function and are calculated through finite differences, while the integration with respect to the random effects $u^{(1)}, u^{(2)}$ is performed through Gaussian quadrature. The approximated marginal likelihood is maximized using a dual quasi-Newton algorithm.

2 Alternative specifications of the model

In general, to fit a (univariate) two-level model it is necessary to integrate out the second level errors (random effects). The fitting of the bivariate two-level model defined in the previous Section involves some additional computational difficulties, due to the integration required for the bivariate normal distribution of the first level errors. Moreover, if the outcomes are more than two the dimension of the multivariate conditional distribution increases, ans so the order of the integrals.

One way to handle the computational complexity of the model with several responses, is to set up a convenient reparametrization based on the following decomposition of the first level error:

$$\varepsilon_{ij}^{(h)} = v_{ij}^{(h)} + \xi_{ij}^{(h)}, \tag{10}$$

where, still considering the case of two items ($H = 2$), the v and ξ errors are independent with

$$\begin{array}{lcl} \left(v_{ij}^{(1)}, v_{ij}^{(2)}\right)' & \stackrel{iid}{\sim} & \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{v}}) \\ \left(\xi_{ij}^{(1)}, \xi_{ij}^{(2)}\right)' & \stackrel{iid}{\sim} & \mathrm{N}(\mathbf{0}, \boldsymbol{I}) \end{array} \tag{11}$$

with

$$\mathbf{\Sigma}_v = \begin{pmatrix} \sigma_{v_1}^2 & \\ \sigma_{v_1 v_2} & \sigma_{v_2}^2 \end{pmatrix}.$$

The constraints on the parameters of the model described in the previous Section must be reported also in the reparametrized version: this amounts to impose two constraints on the three parameters of $\mathbf{\Sigma}_v$, for example $\sigma_{v_1}^2 = \sigma_{v_2}^2 = 1$.

The error decomposition (10), which is exploited by the GLLAMM software (Rabe-Haskett *et al.*, 2001), is the standard trick used to fit multivariate multilevel models by the estimation routines for univariate multilevel models (Snijders and Bosker, 1999). In our context, the bivariate two-level latent model (1) can be viewed as a univariate three-level model where the two responses form the new bottom level. Note that the two responses are independent conditionally on the random effects at subject level $(v_{ij}^{(1)}, v_{ij}^{(2)})$ and cluster level $(u_j^{(1)}, u_j^{(2)})$. Under the proposed decomposition, equations (1) become

$$\begin{array}{lcl} \tilde{Y}_{ij}^{(1)} & = & \alpha^{(1)} + u_j^{(1)} + v_{ij}^{(1)} + \xi_{ij}^{(1)} \\ \tilde{Y}_{ij}^{(2)} & = & \alpha^{(2)} + u_j^{(2)} + v_{ij}^{(2)} + \xi_{ij}^{(2)} \end{array} \tag{12}$$

The conditional covariance structure for the two latent variables is as (4), with $\sigma_{\varepsilon_h}^2 = 1 + \sigma_{v_h}^2$ and $\sigma_{\varepsilon_1 \varepsilon_2} = \sigma_{v_1 v_2}$. Note that the conditional polychoric correlation is now $\sigma_{v_1 v_2} / \sqrt{(\sigma_{v_1}^2 + 1)(\sigma_{v_2}^2 + 1)}$, which is different from the correlation between the subject's random effects $v_{ij}^{(1)}$ and $v_{ij}^{(2)}$.

It is important to note that the original model, based on equations (1), and the reparametrized model, based of equations (12), differ in the scaling factor used to insure identifiability. In fact, in the reparametrized model it is natural to fix to one the standard deviations of $\xi^{(1)}$ and $\xi^{(2)}$, while in the original model the corresponding constraint is imposed on the standard deviations of $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$. In terms of the reparametrized model, the constraint used in the original model amounts to scale the parameters of the h-th latent equation by a factor $\sqrt{1+\sigma^2_{v_h}}$. Therefore the parameters of the original model are smaller in magnitude, specifically they are $1/\sqrt{1+\sigma^2_{v_h}}$ times the corresponding parameters of the reparametrized model.

Though the reparametrized model is equivalent to the original one, the way in which the likelihood is written down is not the same. In particular, the probability (7) is now written as

$$\begin{aligned} & P\left(\bigcap_{h=1}^{2}\left\{Y_{ij}^{(h)}=c^{(h)}\right\} \mid u^{(1)},u^{(2)}\right) \qquad (13)\\ = \; & P\left(\bigcap_{h=1}^{2}\left\{\gamma^{(h)}_{c^{(h)}-1}<\tilde{Y}_{ij}^{(h)}\leq\gamma^{(h)}_{c^{(h)}}\right\} \mid u^{(1)},u^{(2)}\right)\\ = \; & E_{\mathbf{v}}\left[P\left(\bigcap_{h=1}^{2}\left\{\gamma^{(h)}_{c^{(h)}-1}<\tilde{Y}_{ij}^{(h)}\leq\gamma^{(h)}_{c^{(h)}}\right\} \mid u^{(1)},u^{(2)},v^{(1)},v^{(2)}\right)\right]\\ = \; & E_{\mathbf{v}}\left[\prod_{h=1}^{2}P\left(\left\{\gamma^{(h)}_{c^{(h)}-1}<\tilde{Y}_{ij}^{(h)}\leq\gamma^{(h)}_{c^{(h)}}\right\} \mid u^{(h)},v^{(h)}\right)\right]\\ = \; & E_{\mathbf{v}}\left[\prod_{h=1}^{2}\left(F_{\xi^{(h)}}(\gamma^{(h)}_{c^{(h)}}-\eta^{(h)})-F_{\xi^{(h)}}(\gamma^{(h)}_{c^{(h)}-1}-\eta^{(h)})\right)\right], \end{aligned}$$

where $\eta^{(h)}=\alpha^{(h)}+u^{(h)}+v^{(h)}$, $\mathbf{v}=(v^{(1)},v^{(2)})$ and $F_{\xi^{(h)}}$ is the distribution function of the item-specific error $\xi^{(h)}$. Therefore the reparametrized model involves an additional step of integration and so it is likely to be computationally more heavy.

However, the computational burden can be greatly reduced by eliminating one of the two errors at subject level. In fact, the identifiability constraints imply that only one variance-covariance parameter is estimated at subject level, so instead of using two random effects with three parameters and two constraints, one can insert a single random effect. In this way the errors' decomposition (10) becomes:

$$\varepsilon_{ij}^{(h)}=v_{ij}+\xi_{ij}^{(h)}, \qquad (14)$$

with $v_{ij} \overset{iid}{\sim} \mathrm{N}(0, \sigma_v^2)$, and the bivariate two-level latent model (12) is replaced by the equations

$$\begin{aligned} \tilde{Y}_{ij}^{(1)} &= \alpha^{(1)} + u_j^{(1)} + v_{ij} + \xi_{ij}^{(1)} \\ \tilde{Y}_{ij}^{(2)} &= \alpha^{(2)} + u_j^{(2)} + v_{ij} + \xi_{ij}^{(2)}. \end{aligned} \tag{15}$$

The conditional polychoric correlation is now $\sigma_v^2/(1+\sigma_v^2)$, which is restricted to be positive. A negative correlation is obtained multiplying v_{ij} by -1 in one of the two equations. In the model selection process one should try both versions to discover the sign of the correlation.

An alternative parametrization, suggested by Rabe-Haskett (2002), which allows the conditional polychoric correlation to assume both positive and negative values, is the following:

$$\begin{aligned} \tilde{Y}_{ij}^{(1)} &= \alpha^{(1)} + u_j^{(1)} + v_{ij} + \xi_{ij}^{(1)} \\ \tilde{Y}_{ij}^{(2)} &= \alpha^{(2)} + u_j^{(2)} + \lambda v_{ij} + \xi_{ij}^{(2)}. \end{aligned} \tag{16}$$

with $v_{ij} \overset{iid}{\sim} \mathrm{N}(0, 1)$ and λ is a parameter that determines the conditional polychoric correlation: $\lambda/\sqrt{2(\lambda^2+1)}$. Since the total variance of the subject level random term is not identified, it does not matter that the variances are different.

The two-level bivariate probit model based on equations (15) or (16) can be easily fitted by means of softwares such as GLLAMM (Rabe-Hesketh *et al.*, 2001) and aML (Lillard and Panis, 2000), which use Gaussian quadrature to solve the integrals that appear in the likelihood expression.

Finally, it should be noted that the reparametrized model is equivalent to the original one only because of the hypothesis of Gaussian disturbances at item and subject levels. Without such an assumption the model based on the errors' decomposition (10) in general does not correspond to a model with a well-known multivariate distribution at the subject level.

3 Application

The models presented in the previous Sections have been used to analyze some of the data gathered in the survey of course quality carried out by the University of Florence, in all schools of the university, for classes in the 2000-2001 academic year.

The survey form was based on the proposal of a unique standard questionnaire for the evaluation of courses by the students formulated by the National Committee for the Evaluation of the University System (Chiandotto and Gola, 1999). Specifically, we considered the ratings relative to the courses held in the School of Pharmacy, excluding the courses with less than five respondents. We also excluded the questionnaires (3.05% of the total) with a missing response in either of the two items which enter the analysis, i.e. course workload (Q3) and clarity of the teacher (Q13). Altogether, 2888 questionnaires have been considered, corresponding to 87 courses (see Table 2). The number of respondents per course goes from 6 to 136 (median=32, mean=34).

Table 2: Courses evaluated and respondents by year. The University of Florence, School of Pharmacy, academic year 2000-2001.

Year	Courses evaluated	Respondents			
		Tot	Average	min	max
1	18	814	46.72	7	136
2	22	677	31.63	12	67
3	21	678	32.90	6	72
4	10	267	27.40	17	42
5	16	452	29.81	7	71
Tot	87	2888	34.24	6	136

The questionnaire begins with a preliminary section, containing information about the course (its code, the name of the professor, the number of attendant students) and goes on with six sections concerning the evaluation of various aspects of the course. All the questions in these sections require the same type of ordinal response: 1. decidedly no; 2. more no than yes; 3. more yes than no; 4. decidedly yes.

We jointly analyzed two of the questions posed in the questionnaire, that is: course workload (Q3) and clarity of the teacher (Q13). Table 3 reports the sample bivariate distribution of Q3 and Q13. A standard measure of association among the two items is the polychoric correlation (Drasgow, 1981), which is the correlation coefficient of the underlying bivariate normal distribution. Ignoring the hierarchical structure of the data, the polychoric correlation can be estimated by means of a null one-level bivariate probit model, which is fitted by the 'plcorr' option of the SAS FREQ procedure applied to the 4×4 table of frequencies. The resulting estimate is 0.4483 (s.e. 0.0185, see model (a) of Table 5).

Table 3: Respondents by course workload (Q3) and clarity of the teacher (Q13). The University of Florence, School of Pharmacy, academic year 2000-2001.

	Teacher's clarity				*Total*	
Workload	1	2	3	4	N	%
1	44	24	30	6	104	3.60
2	64	138	186	76	464	16.07
3	116	268	718	486	1588	54.99
4	26	56	198	452	732	25.35
Total	250	486	1132	1020	2888	
%	8.66	16.83	39.20	35.32		100.00

In general, the rating of a student to a given item for a certain course may depend on the characteristics of the following hierarchical levels: a) the student (background, expectations etc.); b) the course (subject-matter, organization, professor, readings); c) the curriculum or the school or department (halls, laboratories, sections, orientation etc.); d) the university.

Therefore a full analysis would require a complex multivariate multilevel model (Goldstein, 1995; Snijders and Bosker, 1999). However in our application we consider a single school, so there is no need for the school and university levels; morover in this first stage of the analysis we omit the covariates.

3.1 The univariate models

In order to make an initial assessment of the proportion of variance in the ratings which is linked to the course, we used the NLMIXED procedure of SAS to fit a two-level probit model without covariates (null model) for each of the two considered items (Table 4).

The variance component is highly significant in both cases. The null model provides an estimate of the intraclass correlation coefficient (ICC). This estimate is very different for the two considered items: the proportion of variability of the evaluations which is attributable to the courses is 0.24 for the course workload and 0.41 for the teacher's clarity.

Table 4: Variance decomposition by univariate null models. The University of Florence, School of Pharmacy, academic year 2000-2001.

Model	n.of param.	-2logL	ICC
Course workload (Q3)			
Null, without var.comp.	3	6297.1	
Null, with var.comp.	4	5920.0	0.2369
Teacher's clarity of the (Q13)			
Null, without var.comp.	3	7199.2	
Null, with var.comp.	4	6171.4	0.4143
Number of quadrature points=10, number of observations=2888			

3.2 The bivariate model

Now let us consider the bivariate two-level model defined in Section 1, which can be used to jointly analyze the items Q3 and Q13 (Muthén, 1994; Rabe-Hesketh *et al.*, 2001). The bivariate model includes the correlation structure between the two items, which is interesting in itself and might also influence the other parameters.

In order to identify the model, for each item we fix to zero the first threshold $\gamma_1^{(h)}$ and to one the first level standard deviation σ_{ε_h}, i.e. at the first level only the correlation is estimated.

Interpreting the random effects as factors, two alternative hypotheses can be set up for the second level model:

1. *one-factor model*: there is a single random effect at course level, entering the two linear predictors with different factor loadings;

2. *two-factor model*: there are two random effects at course level (one for each item), whose variances and covariance can be estimated.

The one-factor model is a special case of the two-factor model in which the factors are perfectly correlated. Having a single factor is useful in that the courses can be easily ranked on the basis of the predicted values of that factor; however such a model should be used only if supported by the data.

Table 5 reports the results for three types of bivariate model: (a) single-level (no factor) model, (b) two-level one-factor model and (c) two-level two-factor model. The estimates were obtained with the NLMIXED procedure

Table 5: Bivariate null models. The University of Florence, School of Pharmacy, academic year 2000-2001.

Parameter	(a) no u's		(b) same u's†		(c) different u's†	
	Estim	*s.e.*	*Estim*	*s.e.*	*Estim*	*s.e.*
fixed						
$\alpha^{(Q3)}$	1.797	0.0440	1.809	0.0463	1.988	0.0632
$\gamma_2^{(Q3)}$	0.946	0.0413	0.974	0.0424	1.081	0.0471
$\gamma_3^{(Q3)}$	2.458	0.0479	2.526	0.0503	2.785	0.0563
$\alpha^{(Q13)}$	1.355	0.0329	1.677	0.0578	1.571	0.0530
$\gamma_2^{(Q13)}$	0.693	0.0288	0.941	0.0383	0.940	0.0382
$\gamma_3^{(Q13)}$	1.729	0.0357	2.327	0.0494	2.332	0.0495
covariance						
$\rho_{\tilde{y}^{(Q3)}\tilde{y}^{(Q13)}\mid u^{(Q3)},u^{(Q33)}}$	.	.	0.401	0.0227	0.421	0.0205
$\rho_{\tilde{y}^{(Q3)}\tilde{y}^{(Q13)}}$	0.448	0.0185	0.45	0.0188	0.477	0.0245
$\mathrm{ICC}^{(Q3)}$	0	.	0.056	0.0111	0.245	0.0289
$\mathrm{ICC}^{(Q13)}$	0	.	0.426	0.0229	0.501	0.0248
$\rho_{u^{(Q3)}u^{(Q13)}}$	.	.	1	.	0.624	0.0503
-2logL	13058		12029		11738	
n. of param.	7		9		10	
obs=2888, †non-adaptive Gaussian quadrature with 21 points.						

of SAS, using for models (b) and (c) non-adaptive Gaussian quadrature with 21 points.

In terms of deviances ($-2logL$) the two-factor model is clearly preferable over the one-factor model. Even if the two models lead to a similar estimate of the marginal polychoric correlation (0.477 versus 0.450), in the two-factor model the correlation between the factors (0.62) is farther from unity, which is the value assumed by the one-factor model. The consequences of this incorrect restriction seem to concern mainly the item Q3: in the one-factor model the ICC goes down to 0.056, causing an attenuation in the intercept and thresholds.

Therefore the ranking of the courses on the basis of the considered items cannot rely upon a single measure. Instead, the predicted values of both factors (second level residuals) should be computed and plotted: the best courses are then the ones lying in the I quadrant, while the worst courses are the ones lying in the III quadrant. Extreme cases should be selected for further investigation.

4 Computational remarks

The choice of the number of quadrature points at the higher level is crucial in the case of two or more factors, like model (c) of Table 5, because especially the variance-covariance parameter estimates are quite sensitive to the number of quadrature points. As an example consider the results reported in Table 6.

First consider the original model defined in Section 1 and fitted with SAS NLMIXED. In this case there is only one level which need numerical integration and we found that 21 points were adequate. Note that with 5 points the variance-covariance estimates are totally misleading and change in a substantive manner if the order of the equations is inverted! (this phenomenon was noted also in univariate models with two random effects, in which the estimates obtained with few quadrature points are sensitive to the ordere in which the random effects enter the equation).

The situation is more complex for the reparametrized model defined in Section 2, equations (15), which requires numerical integration at two levels of the hierarchy. However, with both GLLAMM and AML, it is clear that the course level, which has two random effects, is more demanding, in terms of quadrature points, than the subject level, which has a single random effect. For example, the estimates obtained with 21 points of quadrature at both the levels are the same as the estimates obtained with 21 points at the second level and 10 points at the first level. Also using the same number of quadrature points at the second level, in this particular case, the three software give estimates that are not exactly the same. This can be due to the different implemented algorithm, but more investigation is needed to understand what is going on.

The presented model can be easily extended to the case of more then two items, by means of the reparametrization proposed in (12) or (15). The NLMIXED SAS procedure does not allow to estimate such an extended model, instead one can use the GLLAMM procedure of STATA or the aML software. GLLAMM is a flexible procedure, but the the time requested for the estimation increases rapidly with the complexity of the model. Provided with good initial values, aML seems to be very quickly, but the software does not allow, at the present moment, to estimate the second level residuals.

Table 6: Bivariate null model (c) of Table 5 estimated with different software and number of quadrature points.

Parameters	SAS NLMIXED			aML	GLLAMM
	5 qp	*10 qp*	*21 qp*	*10;21 qp*	*10;21 qp*
Fixed					
$\alpha^{(Q3)}$	1.934	1.854	1.988	2.055	2.055
$\gamma_2^{(Q3)}$	1.072	1.082	1.081	1.081	1.081
$\gamma_3^{(Q3)}$	2.764	2.782	2.785	2.785	2.786
$\alpha^{(Q13)}$	1.547	1.342	1.571	1.877	1.879
$\gamma_2^{(Q13)}$	0.924	0.928	0.940	0.936	0.935
$\gamma_3^{(Q13)}$	2.295	2.314	2.332	2.325	2.325
Covariance					
$\rho_{\tilde{y}^{(Q3)}\tilde{y}^{(Q13)}\|u^{(Q3)},u^{(Q13)}}$	0.425	0.422	0.421	0.424	0.423
$ICC^{(Q3)}$	0.194	0.219	0.245	0.228	0.228
$ICC^{(Q13)}$	0.302	0.522	0.501	0.401	0.400
$\rho_{u^{(Q3)}u^{(Q13)}}$	0.289	0.677	0.624	0.552	0.550
-2logL	11781	11767	11738	11732	11732

References

Chiandotto B., Gola M. M., 1999 Questionario di base da utilizzare per l'attuazione di un programma per la valutazione della didattica da parte degli studenti, *Osservatorio Nazionale per la Valutazione del Sistema Universitario*, Roma.

Drasgow, 1981 Polychoric and polyserial correlations, Encyclopedia of Statistics, Johnson and Kotz Ed.

Lillard L.A., Panis C.W.A., 2000 *aML Multilevel Multiprocess Statistical Software, Release 1.0.* EconWare, Los Angeles, California.

Mazzolli, 2001 A multilevel structural equation model with polytomous and dichotomous data, *mimeo* presented at the Third International Conference on Multilevel Analysis, Amsterdam April 9-10, 2001

Muthén B.O., 1994 Multilevel Covariance Structure Analysis, *Sociological Methods & Research*, **22**, 376-398.

Rabe-Hesketh S., Skrondal A. and A. Pickles, 2001. Generalized multilevel structural equation modelling, paper submitted for publication.

Rabe-Hesketh S., A. Pickles and Skrondal A., 2001. GLLAMM Manual, *Technical report2001/01*, Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London.

Snijders T.A.B., Bosker R.J., 1999 *Multilevel Analysis. An introduction to basic and advanced multilevel modelling.* Sage, London.

Acknowledgements

We are grateful to Sophia Rabe-Hesketh for her constructive and detailed comments on a preliminary version of the paper and GLAMM use and to Stan Panis for his useful suggestions on aML use and model specification.