A quasi-score approach to the analysis of ordered catogorical data via a mixed heteroskedastic threshold model

Florence Jaffrezic, Christèle Robert-Granié, J.L. Foulley

Original article

A quasi-score approach to the analysis of ordered categorical data via a mixed heteroskedastic threshold model

Florence Jaffrézic, Christèle Robert-Granié, Jean-Louis Foulley*

Station de génétique quantitative et appliquée, Institut national
de la recherche agronomique, Centre de recherches de Jouy-en-Josas,
78352 Jouy-en-Josas cedex, France

Abstract – This article presents an extension of the methodology developed by Gilmour et al. [19], for ordered categorical data, taking into account the heterogeneity of residual variances of latent variables. Heterogeneity of residual variances is described via a structural linear model on log-variances. This method involves two main steps: i) a 'marginalization' with respect to the random effects leading to quasi-score estimators; ii) an approximation of the variance–covariance matrix of the observations which leads to an analogue of the Henderson mixed model equations for continuous Gaussian data. This methodology is illustrated by a numerical example of footshape in sheep. © Inra/Elsevier, Paris

generalized linear mixed models / quasi-score / heterogeneity of variances / threshold response model

Résumé – Une approche de quasi-score pour l'analyse de variables qualitatives ordonnées par un modèle mixte à seuils hétéroscédastique. Cet article présente une extension de la méthodologie développée par Gilmour et al. [19] dans le cas de variables qualitatives ordonnées, prenant en compte l'hétérogénéité des variances résiduelles des variables latentes. L'hétérogénéité des variances résiduelles est décrite par un modèle linéaire structurel sur les logarithmes des variances. Cette méthode comprend deux étapes principales : i) une « marginalisation » par rapport aux effets aléatoires qui conduit, grâce aux équations de quasi-score, à l'estimation des paramètres ; ii) une approximation de la matrice de variance–covariance des observations qui aboutit à un système analogue aux équations du modèle mixte d'Henderson dans le cas de variables continues gaussiennnes. Cette méthodologie est illustrée par un exemple sur la forme des pieds chez le mouton. © Inra/Elsevier, Paris

modèles linéaires généralisés mixtes / quasi-score / variances hétérogènes / modèle à seuils

* Correspondence and reprints
E-mail: foulley@jouy.inra.fr

1. INTRODUCTION

The threshold model is one of the most popular models for analysing ordered categorical data especially in population [36, 37] and quantitative [7] genetics as well as in animal breeding [16].

Recently Foulley and Gianola [8] extended the standard threshold model to a model allowing for heterogeneous variances of the Gaussian latent variables using a log-linear model for the residual variances. In the case of mixed models, they proposed to base inference about threshold cutoff points, location and dispersion parameters of the latent distribution on the mode of the a posteriori (MAP) distribution. This approach is basically a conditional one (given random effects) and is similar to penalized quasi-likelihood [1, 31], iterated re-weighted restricted maximum likelihood [5] and hierarchical likelihood of generalized linear mixed models [28] for one parameter exponential families. As discussed by Foulley and Manfredi [10] and Engel and Keen [6], these procedures are likely to have some drawbacks regarding the estimation of fixed effects due to the approximation in integrating out random effects.

One simple way to overcome the difficulty of an exact integration of random effects is the quasi-score approach of Mc Cullagh and Nelder [30] which only requires the mean and variance of the data distribution. In particular, an appealing version of the quasi-score approach for computing estimations of fixed effects was proposed by Gilmour et al. [18, 19] using an approximation of the variance–covariance matrix. One of the main advantages of this method is that it mimics the mixed model equations of Henderson [23] making the estimation of fixed effects computationally easier and providing analogues of BLUP (best linear unbiased predictor) of random effects as by-products. Moreover, this quasi-score method, via linearization, was proven to be quite general [1, 21, 39]. Initially derived by Gilmour et al. [18] for binary data modelled with logit or probit links, it was applied to ordered categorical data by the same authors [19], to Poisson data with a log link by Foulley and Im [9] and to a log link exponential model by Trottier [35]. The purpose of this paper is to show how this procedure can also cope with heterogeneous residual variances in the case of ordered polytomics modelled via Gaussian latent variables. Section 2 entitled 'Theory' outlines the model, the quasi-score equations and their GAR [19] counterpart and by-products. Section 3 illustrates the theory using the numerical example of footshape in sheep presented by GAR [19].

2. THEORY

2.1. Model

The model assumptions and notations are basically the same as in Foulley and Gianola [8]. First, it is assumed that the population can be stratified according to an index i $(i = 1, 2, \ldots, I)$ such that the between subgroup variation corresponds to systematic influences of identified factors and the within group variation to random noise.

There are J response categories indexed by j such that $\boldsymbol{y}_{i+} = \{y_{ij+}\}$ represents the vector of the counts of responses for subpopulation i in the

different categories j. The vector $\boldsymbol{y}_{i+}$ can be expressed as the sum $\boldsymbol{y}_{i+} = \sum_{r=1}^{n_i} \boldsymbol{y}_{ir}$ of indicator vectors $\boldsymbol{y}_{ir} = (y_{i1r}, y_{i2r}, \ldots, y_{ijr}, \ldots, y_{iJr})'$ such that $y_{ijr} = 1$ if response of observation r in subpopulation i is in category j and $y_{ijr} = 0$ otherwise.

In the threshold approach, the probability of a response in category j for an observation of population i, say π_{ij}, is described by the distribution of continuous latent variables ℓ_{ir}. The expression of these variables is discretized via threshold values $(\xi_1, \xi_2, \ldots, \xi_j, \ldots, \xi_{J-1})$, $(\xi_0 = -\infty$ and $\xi_J = +\infty)$ such that:

$$y_{ijr} = 1 \iff \xi_{j-1} < \ell_{ir} \leqslant \xi_j \tag{1}$$

A mixed model structure is hypothesized on the latent variable:

$$\ell_{ir} = \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma_{u_i}\boldsymbol{z}_i'\boldsymbol{u}^* + e_{ir} \tag{2}$$

where $\eta_i = E(\ell_{ir})$ is decomposed as a linear function $\boldsymbol{x}_i'\boldsymbol{\beta}$ of explanatory variables (row vector $\boldsymbol{x}_i'$) with unknown coefficients $\boldsymbol{\beta} \in \mathrm{IR}^p$; $\sigma_{u_i}\boldsymbol{z}_i'\boldsymbol{u}^*$ represents the contribution of random effects to the model with $\boldsymbol{u}^*$ being a $(q \times 1)$ vector of scaled deviations, $\boldsymbol{z}_i'$ the corresponding row incidence vector and σ_{u_i} the square root of the $\boldsymbol{u}$-component of variance, which may vary between subpopulations.

Classical assumptions are made regarding the distribution of $\boldsymbol{u}^*$ and $\boldsymbol{e}_i = \{e_{ir}\}$, i.e. $\boldsymbol{u}^* \sim \mathcal{N}(0, \boldsymbol{I}_q)$ or, in genetics, $\boldsymbol{u}^* \sim \mathcal{N}(0, \boldsymbol{A})$ where $\boldsymbol{A}$ represents the known relationship matrix, $\boldsymbol{e}_i \sim \mathcal{N}(0, \sigma_{e_i}^2 \boldsymbol{I}_{n_i})$ and $\mathrm{Cov}(\boldsymbol{u}, \boldsymbol{e}_i') = 0$.

Homogeneity of the covariate structure is assumed within the subpopulation i, i.e. $\boldsymbol{x}_{ir} = \boldsymbol{x}_i$ and $\boldsymbol{z}_{ir} = \boldsymbol{z}_i$. If not (e.g. when $\boldsymbol{x}_i$ is a continuous covariate), smaller units will be considered, even at the limit elementary units $(n_i = 1)$.

Moreover, as in Foulley and Gianola [8], the ratio $\rho_i = \sigma_{u_i}/\sigma_{e_i}$ is assumed to be constant (ρ) across populations which is equivalent to supposing homogeneous intra-class correlations (e.g. constant heritability or repeatability) across environments. Thus,

$$E(\ell_{ir}) = \eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{3}$$

$$\mathrm{Var}(\ell_{ir}) = \sigma_{e_i}^2(1 + \rho^2\boldsymbol{z}_i'\boldsymbol{A}\boldsymbol{z}_i) = \sigma_{e_i}^2(1 + \rho^2\alpha_i) \tag{4}$$

with $\alpha_i = \boldsymbol{z}_i'\boldsymbol{A}\boldsymbol{z}_i$. In many applications, α_i is a constant or even $\alpha_i = 1$, but this simplification is not mandatory throughout this paper. In fact, the theory is presented here with a single random factor but it can be easily extended to any number K of independent random vectors $\boldsymbol{u}_k^*$.

Similarly for the expectations, a structure is postulated for residual variances so as to account for the effects of factors causing heteroskedasticity. As in Foulley et al. [13, 14], heterogeneity of residual variances is described by a structural linear model and a log link function, as follows:

$$\ln \sigma_{e_i} = \boldsymbol{p}_i'\boldsymbol{\delta} \tag{5}$$

where $\boldsymbol{p}_i'$ is the $(1 \times r)$ row vector of covariates and $\boldsymbol{\delta}$ is the $(r \times 1)$ vector of real-valued dispersion parameters.

2.2. Estimation

The estimation procedure described here includes two steps. The first step consists in setting up the quasi-score equations based on the first two marginal moments according to the quasi-likelihood theory [30] and its extension to correlated observations [29]. The second step lies in replacing the variance–covariance matrix of observations by an approximation which is analogous to solving for fixed effects using the mixed model equations of Henderson [23].

2.2.1. Quasi-score equations

Let $\boldsymbol{\theta} = (\boldsymbol{\xi}', \boldsymbol{\beta}', \boldsymbol{\delta}')'$ be the $(J - 1 + p + r)$ vector of parameters of interest, where $\boldsymbol{\xi}_{((J-1)\times 1)}$ are the thresholds, $\boldsymbol{\beta}_{(p\times 1)}$ the location parameters, and $\boldsymbol{\delta}_{(r\times 1)}$ the dispersion parameters. The quasi-score equations are:

$$U(\boldsymbol{\theta}, \widetilde{\boldsymbol{y}}) = \boldsymbol{D}'\boldsymbol{\Sigma}^{-1}(\widetilde{\boldsymbol{y}} - \boldsymbol{\mu}) = 0 \tag{6}$$

where $\widetilde{\boldsymbol{y}}_{(I(J-1)\times 1)} = (\widetilde{\boldsymbol{y}}_1', \widetilde{\boldsymbol{y}}_2', \ldots, \widetilde{\boldsymbol{y}}_i', \ldots, \widetilde{\boldsymbol{y}}_I')'$ is the vector of the observed cumulative proportions with $\widetilde{\boldsymbol{y}}_{i_{((J-1)\times 1)}} = \boldsymbol{L}y_{i+}/n_i$, $\boldsymbol{L}$ is a $((J - 1) \times J)$ matrix built from a lower triangular matrix of 1s, the last row of which is removed. In addition, $\boldsymbol{\mu} = E(\widetilde{\boldsymbol{y}})$, $\boldsymbol{\Sigma} = \mathrm{Var}(\widetilde{\boldsymbol{y}})$ and $\boldsymbol{D}' = \partial\boldsymbol{\mu}'/\partial\boldsymbol{\theta}$ with dimension $((J + p + r - 1) \times I(J - 1))$.

Equations in (6) need to specify $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ which can be performed as follows.

Let $\mu_{ij} = \sum_{k=1}^{j} \pi_{ik}$. The conditional expectation of μ_{ij} given realized values of the random effects $\boldsymbol{u}^*$ is defined as $\mu_{ij}(\boldsymbol{u}^*) = \mathrm{Pr}(\ell_{ir} \leqslant \xi_j | \boldsymbol{u}^*)$ which, due to the distribution assumptions made, can be expressed as a normal cumulative density function (CDF):

$$\mu_{ij}(\boldsymbol{u}^*) = \Phi\Big(\frac{\xi_j - \boldsymbol{x}_i'\boldsymbol{\beta} - \sigma_{u_i}\boldsymbol{z}_i'\boldsymbol{u}^*}{\sigma_{e_i}}\Big) \tag{7}$$

In the marginal model, μ_{ij} is the expectation of $\mu_{ij}(\boldsymbol{u}^*)$ with respect to the distribution of $\boldsymbol{u}^*$. Remember that if $X \sim \mathcal{N}(\mu, \sigma^2)$, the $E(\Phi(X)) = \Phi(\mu(1 + \sigma^2)^{-1/2})$ [2]. Here, the expectation of (7) reduces to:

$$\mu_{ij} = \Phi(\gamma_{ij}) \tag{8}$$

with

$$\gamma_{ij} = \frac{\xi_j - \boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma_{e_i}\sqrt{1 + \alpha_i\rho^2}}$$

As shown in detail in the Appendix, the variance–covariance matrix $\boldsymbol{\Sigma}$ of the observations can be decomposed as the sum of two components:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_B \tag{9}$$

The first component $\boldsymbol{\Sigma}_A$ is a $(I(J-1) \times I(J-1))$ block diagonal matrix such that:

$$\boldsymbol{\Sigma}_A = \overset{I}{\underset{i=1}{\oplus}} \boldsymbol{\Sigma}_{A,ii} \tag{10}$$

$$\boldsymbol{\Sigma}_{A,ii} = \frac{1}{n_i}(\boldsymbol{\Sigma}_{0,ii} - \boldsymbol{\Sigma}_{B,ii}) \tag{11}$$

In equation (11), $\boldsymbol{\Sigma}_{0,ii}$ is a $((J-1) \times (J-1))$ matrix whose general term is $(\sigma_{0,ii})_{jk} = \mu_{ij}(1 - \mu_{ik})$ for $j, k = 1, \ldots, (J-1)$, so that $\boldsymbol{\Sigma} = \overset{I}{\underset{i=1}{\oplus}}(\boldsymbol{\Sigma}_{0,ii})/n_i$ is the variance–covariance matrix of observations for multinomial data (i.e. a purely fixed model).

The second component $\boldsymbol{\Sigma}_B$ corresponds to the covariance terms for off-diagonal blocks, i.e.:

$$\forall i \neq i', \ \boldsymbol{\Sigma}_{i,i'_{((J-1)\times(J-1))}} = \boldsymbol{\Sigma}_{B,ii'} \tag{12}$$

For any pair of blocks (diagonal $i = i'$ or off-diagonal $i \neq i'$) its general term (jk) can be expressed as:

$$(\sigma_{B,ii'})_{jk} = \Phi_2(\gamma_{ij}, \gamma_{i'k}; t_{ii'}) - \Phi(\gamma_{ij})\Phi(\gamma_{i'k}) \tag{13}$$

where $t_{ii'}$ is the correlation coefficient between ℓ_{ir} and $\ell_{i'r'}$ and $\Phi_2(a, b; r)$ is the CDF of the standardized binormal distribution with arguments a, b, and correlation r.

The system in equation (6) can be solved by Fisher's iterative algorithm as follows:

$$(\boldsymbol{D}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{(t)}\Delta\boldsymbol{\theta}^{(t+1)} = (\boldsymbol{D}'\boldsymbol{\Sigma}^{-1})^{(t)}(\tilde{\boldsymbol{y}} - \boldsymbol{\mu}^{(t)}) \tag{14}$$

where $\Delta\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}$.
$\boldsymbol{D}' = \partial\boldsymbol{\mu}'/\partial\boldsymbol{\theta}$ can be decomposed as $(\partial\boldsymbol{\gamma}'/\partial\boldsymbol{\theta})(\partial\boldsymbol{\mu}'/\partial\boldsymbol{\gamma})$.
Now $\mu_{ij} = \Phi(\gamma_{ij})$ so that:

$$\frac{\partial\boldsymbol{\mu}'}{\partial\boldsymbol{\gamma}} = \boldsymbol{\phi} \tag{15}$$

with $\boldsymbol{\phi} = \overset{I}{\underset{i=1}{\oplus}} \boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_i = \text{diag}\{\phi(\gamma_{ij})\}$ for $j = 1, 2, \ldots, (J-1)$, where $\phi(.)$ is the standardized normal density function. The second element can be written as the product:

$$\frac{\partial\boldsymbol{\gamma}'}{\partial\boldsymbol{\theta}} = \boldsymbol{T}'\boldsymbol{H}' \tag{16}$$

where

$$\boldsymbol{T}' = \begin{pmatrix} \boldsymbol{I}_{(J-1)} & 0 & 0 \\ 0 & \boldsymbol{X}' & 0 \\ 0 & 0 & \boldsymbol{P}' \end{pmatrix}$$

with

$$\boldsymbol{X}'_{(p \times I)} = \quad (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i, \ldots, \boldsymbol{x}_I)$$

$$\boldsymbol{P}'_{(r \times I)} = \quad (\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_i, \ldots, \boldsymbol{p}_I)$$

and

$$\boldsymbol{H}_{(I(J-1) \times (2I+J-1))} = (\boldsymbol{H}_{\boldsymbol{\xi}}, \boldsymbol{H}_{\boldsymbol{\eta}}, \boldsymbol{H}_{\boldsymbol{\nu}})$$

with

$$\boldsymbol{H}'_{\boldsymbol{\xi}_{((J-1) \times I(J-1))}} = \frac{\partial \gamma'}{\partial \boldsymbol{\xi}} = \left\{ \frac{\partial \gamma_{ij}}{\partial \xi_j} = \frac{1}{\Psi_i \sigma_{e_i}} \right\} \tag{17}$$

$$\boldsymbol{H}'_{\boldsymbol{\eta}_{(I \times I(J-1))}} = \frac{\partial \gamma'}{\partial \boldsymbol{\eta}} = \left\{ \frac{\partial \gamma_{ij}}{\partial \eta_i} = -\frac{1}{\Psi_i \sigma_{e_i}} \right\} \tag{18}$$

$$\boldsymbol{H}'_{\boldsymbol{\nu}_{(I \times I(J-1))}} = \frac{\partial \gamma'}{\partial \boldsymbol{\nu}} = \left\{ \frac{\partial \gamma_{ij}}{\partial \nu_i} = -\gamma_{ij} \right\} \tag{19}$$

and $\Psi_i = \sqrt{1 + \rho^2 \alpha_i}$.

Replacing $\boldsymbol{D}'$ in equation (14) by its combined expression $\boldsymbol{D}' = \boldsymbol{T}'\boldsymbol{H}'\phi$ from equations (15) and (16) leads to an iterative generalized least square system:

$$\boldsymbol{T}'(\boldsymbol{H}'\boldsymbol{W}\boldsymbol{H})^{(t)}\boldsymbol{T}\theta^{(t+1)} = \boldsymbol{T}'(\boldsymbol{H}'\boldsymbol{W})^{(t)}\boldsymbol{v}^{(t)} \tag{20}$$

where $\boldsymbol{W}_{(I(J-1) \times I(J-1))} = \phi \boldsymbol{\Sigma}^{-1} \phi$ is a matrix of weights, and $\boldsymbol{v} = \boldsymbol{H}\boldsymbol{T}\theta + \phi^{-1}(\widetilde{\boldsymbol{y}} - \boldsymbol{\mu})$ is a working variable. Both are updated from round (t) to round $(t+1)$ of iteration using the current value $\boldsymbol{\theta}^{(t)}$ of $\boldsymbol{\theta}$.

2.2.2. The GAR procedure

The size $(I(J-1))$ of the $\boldsymbol{\Sigma}$ matrix to invert in $\boldsymbol{W}$ may be very large in some types of applications (e.g. genetic evaluation of field data). This precludes the use of the equation system (20) for computing $\boldsymbol{\theta}$ estimates. This was the basic reason why Gilmour et al. [18] proposed an alternative procedure based on a convenient approximation of $\boldsymbol{\Sigma}$, whose principle was explained in detail in Foulley et al. [12].

Let $\Omega(a, b; r) = \Phi_2(a, b; r) - \Phi(a)\Phi(b)$. Using Tallis's [34] result viz $\partial\Omega/\partial r = \phi_2(a, b; r)$ ($\phi_2(.)$: standardized bivariate density with arguments a, b and correlation r), the first order Taylor expansion of $\Omega(a, b; r)$ about $r = 0$ is $\Omega(a, b; r) = r\phi(a)\phi(b) + o(r^2)$. Applying this to $a = \gamma_{ij}$, $b = \gamma_{i'k}$ and $r = t_{ii'}$, which occur in the general term of $\boldsymbol{\Sigma}_{B,ii'}$ (cf. equation (13)), leads to:

$$(\sigma_{B,ii'})_{jk} \approx t_{ii'}\phi(\gamma_{ij})\phi(\gamma_{i'k}) \tag{21}$$

This can also be written as:

$$\boldsymbol{\Sigma}_{B,ii'} = \phi_i \boldsymbol{M}_i \boldsymbol{z}'_i \boldsymbol{G} \boldsymbol{z}_{i'} \boldsymbol{M}'_{i'} \phi_{i'} \tag{22}$$

where ϕ_i and z_i' are as previously defined, $G = A\rho^2$, $M_i = -\mathbf{1}_{(J-1)}\Psi_i^{-1}$ ($\mathbf{1}_{(k)}$ is a vector of k ones and the minus sign is used for the convenience of calculation).

Letting $Z_{(I \times q)} = (z_1, z_2, \ldots, z_i, \ldots, z_I)'$, $M_{(I(J-1) \times I)} = \bigoplus\limits_{i=1}^{I} M_i$ and $Z^*_{(I(J-1) \times q)} = MZ$, Σ and its components can be expressed in condensed form as:

$$\Sigma_B \approx \phi Z^* G Z^{*'} \phi \tag{23}$$

$$\Sigma \approx \Sigma_A^* + \phi Z^* G Z^{*'} \phi \tag{24}$$

where Σ_A^* is the same as defined in equations (10) and (11) with block diagonal terms of Σ_B replaced by their approximations given in equation (23). Substituting Σ in $W^{-1} = \phi^{-1}\Sigma\phi^{-1}$ by its expression in equation (24), one has:

$$W^{-1} = \phi^{-1}\Sigma_A^* \phi^{-1} + Z^* G Z^{*'} \tag{25}$$

which displays the classical form $R + Z^* G Z^{*'}$ of a variance–covariance matrix of data under a linear mixed model. This structure enables us to solve for θ in (20) using the Henderson mixed model equations [23], i.e. here with:

$$\begin{pmatrix} T'(H'R^{-1}H)^{(t)}T & T'(H'R^{-1}M)^{(t)}Z \\ Z'(M'R^{-1}H)^{(t)}T & Z'(M'R^{-1}M)^{(t)}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\theta}^{(t+1)} \\ \widehat{u}^{(t+1)} \end{pmatrix}$$
$$= \begin{pmatrix} T'(H'R^{-1}v)^{(t)} \\ Z'(M'R^{-1}v)^{(t)} \end{pmatrix} \tag{26}$$

R^{-1} can be directly calculated due to the peculiar structure of Σ_0 which has a tridiagonal inverse (see Appendix). Detailed expressions for the elements of the coefficient matrix and the right hand side of (26) can be found in the Appendix.

Moreover, arguing as Gilmour et al. [19] from the mixed model structure of equation (26), one can extract two by-products of this system:

i) a BLUP-type prediction of the random effects represented by the $\widehat{u}$ solution to equation (26).

ii) a EM-REML-type estimation of the variance component, say here ρ^2 via:

$$\widehat{\rho}^{2(n+1)} = \frac{\widehat{u}'^{(n)} A^{-1} \widehat{u}^{(n)} + \text{trace}(A^{-1} C_{uu})}{q} \tag{27}$$

where C_{uu} is the portion of the inverse of the coefficient matrix in equation (26) corresponding to u.

In some instances, one may consider a backtracking procedure [3] to reach convergence, i.e. at the beginning of the iterative process, compute $\alpha^{(k+1)} = (\theta'^{(k+1)}, u'^{(k+1)})'$ as $\alpha^{(k+1)} = \alpha^{(k)} + \varpi^{(k+1)}\Delta\alpha^{(k+1)}$ with $0 < \varpi^{(k+1)} \leqslant 1$.

3. NUMERICAL EXAMPLE

The preceding theory is now illustrated with a small example. For pedagogical reasons, the data set used is the same as the one analysed by Gilmour et al. [19]. The data consisted of footshape scores recorded in three categories on 2 513 lambs observed over a 2 year period, out of five mating groups [17] later on referred to as 'breeds' for simplicity, and sired by 34 rams which are assumed to be unrelated.

The data set is listed in *table I*. As the year (Y_i; $i = 1, 2$) and breed (B_j; $j = 1, 2, 3, 4, 5$) factors are disconnected, parametrization is not standard. Following Searle's [32] 'cell means models', the parametrization adopted here is defined from the elementary estimable parameters, i.e. here the cell location (η_{ij}) and dispersion (ν_{ij}) parameters.

Table I. Distribution of footshape records (from Gilmour et al. [19]).

Sire	Fixed effects		Categories			Sire	Fixed effects		Categories		
	Year	Breed	[1]	[2]	[3]		Year	Breed	[1]	[2]	[3]
1	1	1	52	25	0	18	2	4	47	24	12
2	1	1	49	17	1	19	2	4	46	25	9
3	1	1	50	13	1	20	2	4	79	32	11
4	1	1	42	9	0	21	2	4	50	23	5
5	1	1	74	15	0	22	2	4	63	18	8
6	1	1	54	8	0	23	2	5	30	20	9
7	1	1	96	12	0	24	2	5	31	33	3
8	1	2	57	52	9	25	2	5	28	18	4
9	1	2	55	27	5	26	2	5	42	27	4
10	1	2	70	36	4	27	2	5	35	22	2
11	1	2	70	37	3	28	2	5	33	18	3
12	1	2	82	21	1	29	2	5	35	17	4
13	1	2	75	19	0	30	2	5	26	13	2
14	1	3	17	12	10	31	2	5	37	15	2
15	1	3	13	23	3	32	2	5	36	14	1
16	1	3	21	17	3	33	2	5	63	20	3
17	2	4	37	41	23	34	2	5	41	8	1

The chosen functions are as follows:

$$\beta_0 = \mu_{11}$$

$$\beta_1 = \sum_{k=4}^{5} \frac{n_{2k}}{n_{2+}} \mu_{2k} - \sum_{j=1}^{3} \frac{n_{1j}}{n_{1+}} \mu_{1j}$$

$$\beta_2 = \mu_{12} - \mu_{11}; \quad \beta_3 = \mu_{13} - \mu_{11}; \quad \beta_4 = \mu_{25} - \mu_{24}$$

with $n_{1+} = \sum_{j=1}^{3} n_{1j}$ and $n_{2+} = \sum_{k=4}^{5} n_{2k}$.

β_0 represents the effect of a reference population (breed 1 in year 1); β_1 is a possible measure of a 'year' effect; β_2, β_3 and β_4 stand for within year contrasts between breeds.

Letting those estimable functions expressed as $\boldsymbol{\beta} = \boldsymbol{B\mu}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)'$, $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{24}, \mu_{25})'$ and $\boldsymbol{B}$ is the (5×5) matrix of coefficients given previously, the incidence matrix $\boldsymbol{X}$ used in equations (3) and (16) is obtained simply as $\boldsymbol{X} = \boldsymbol{B}^{-1}$ (since $\boldsymbol{\mu} = \boldsymbol{X\beta} = \boldsymbol{XB\mu}$). Note that this parametrization not only makes sense as far as its practical interpretation is concerned, but also generates an intercept β_0 (since $\boldsymbol{b}'_i\boldsymbol{1} = 0$, $\forall i$) which can be substracted from the original threshold values ξ_j making computations easier (see Foulley et al. [12], formula 17.85 p. 392, and Gilmour et al. [19] formula 2).

The same $\boldsymbol{B}$ transformation applies to the δ_{ij} as linear functions of the $\nu_{ij} = \ln\sigma_{ij}$. The interpretation of parameters is similar to previously, but with the geometric means replacing arithmetic means and ratios replacing differences as shown below:

$$\delta_0 = \ln\sigma_{11}$$

$$\delta_1 = \ln\sqrt{\frac{\sigma_{2*}^2}{\sigma_{1*}^2}}$$

$$\delta_2 = \ln\sqrt{\frac{\sigma_{12}^2}{\sigma_{11}^2}}; \ \delta_3 = \ln\sqrt{\frac{\sigma_{13}^2}{\sigma_{11}^2}}; \ \delta_4 = \ln\sqrt{\frac{\sigma_{25}^2}{\sigma_{24}^2}}$$

with $\sigma_{1*}^2 = [\prod_{j=1}^{3}(\sigma_{1j}^2)^{n_{1j}}]^{1/n_{1+}}$, idem for σ_{2*}^2.

The general procedure presented here was applied to both standard (S-TM) and heteroskedastic (H-TM) threshold models with the fixed parametrization effects described above for the location and dispersion parameters, and random sire effects within year × breed subclasses.

Data were not analysed in detail since the main purpose of this numerical illustration is to serve as a test example. Parameter estimates under both models are shown in *table II*. The intra-class correlation (sire variance) was estimated as 0.0622 and 0.0630 under the S-TM and H-TM, respectively. Differences between sire predictions under the two models are distinct but small, suggesting, as expected, a wider spread of predictions under the H-TM (+ 0.8 %).

The estimations of fixed effects for location parameters under the S-TM model are not directly comparable with those obtained by Gilmour et al. [19] owing to different parametrizations. The estimates and Wald's tests (*table III*) provide strong evidence for heterogeneity in residual variances. Marked differences can be observed between year 2 and year 1 (ratio: $\sigma_{Y_2}^2/\sigma_{Y_1}^2 = \exp(2 * 0.3145) = 1.88$) and between breeds especially in year 2 (ratios: $\sigma_{B_3}^2/\sigma_{B_1}^2 = \exp(2 * 0.3389) = 1.97$ and $\sigma_{B_5}^2/\sigma_{B_4}^2 = \exp(2 * (-0.3016)) = 0.55$).

It is worth noting that, in the H-TM model, year and breed contrasts within year 2 are not significant factors of variation of the mean but greatly influence the residual variance contrarily to what happened with the breed contrast

Table II. GAR estimates of the parameters and their precision.

Parameters		Homogeneous		Heterogeneous	
		Estimate	SE	Estimate	SE
Thresholds	$\xi_2-\xi_1$	1.2705	0.0425	1.8601	0.2411
Location					
β_0	$\mu_{11}-\xi_1$	−0.9061	0.1142	−0.8716	0.1150
β_1	y_2-y_1	0.3316	0.1024	0.1659	0.1581
β_2	b_2-b_1 in y_1	0.4500	0.1612	0.3216	0.2106
β_3	b_3-b_1 in y_1	1.0759	0.2117	1.1359	0.2801
β_4	b_5-b_4 in y_2	−0.2018	0.1426	−0.0469	0.2617
Scale					
δ_1	y_2-y_1			0.3145	0.0786
δ_2	b_2-b_1 in y_1			0.2502	0.1466
δ_3	b_3-b_1 in y_1			0.3389	0.1749
δ_4	b_5-b_4 in y_2			−0.3016	0.0919
Random effects	(Sires 1–5)				
s_1		0.2944	0.1585	0.3040	0.1590
s_2		0.1771	0.1626	0.1903	0.1632
s_3		0.0629	0.1640	0.0717	0.1646
s_4		−0.0523	0.1712	−0.0567	0.1719
s_5		−0.0827	0.1543	−0.0886	0.1549
Variance component					
ρ^2		0.0622		0.0630	

y_1 and y_2 stand for year 1 and year 2, b_1 to b_4 stand for breeds 1 to 4.

Table III. Wald's tests.

	df*	Homogeneous		Heterogeneous	
		Wald	P-value	Wald	P-value
Model for $\boldsymbol{\beta}$	4	35.827	0.000	18.780	0.001
Y	1	10.490	0.001	1.101	0.294
$B:Y_1$	2	26.670	0.000	17.046	0.000
$B:Y_2$	1	2.002	0.157	0.032	0.858
Model for $\boldsymbol{\delta}$	4			26.631	0.000
Y	1			15.990	0.000
$B:Y_1$	2			4.089	0.129
$B:Y_2$	1			10.761	0.001

* Degree of freedom
The breed effects within year 1 $(B : Y_1)$ and within year 2 $(B : Y_2)$ were tested separately.

within year 1. Thus one may apply in practice a more parsimonious H-TM model which has in that case as many parameters as the S-TM model (i.e. four fixed effects + one variance component) but fits the data set better (χ^2 Pearson statistics = 27.0 and 11.8 for 4 degrees of freedom for purely fixed models).

4. DISCUSSION

New perspectives are opened in the analysis of ordinal data by the use of heteroskedastic threshold models [38]. The justifications for considering this extension were discussed at length by Foulley and Gianola [8] and include alternatives such as the variable threshold concept and its relationship with H-TM. Here, heterogeneous variances were considered within the framework of the usual mixed linear model with heteroskedasticity described by structural models [14, 15]. This problem can also be tackled under different model structures such as for instance the multilevel models of Golstein [20].

In the H-TM context, the GAR quasi-likelihood procedure turns out to be a natural alternative to the MAP approach proposed by Foulley and Gianola [8]. The main advantage of the MAP approach lies in both its conceptual and computational simplicity. Part of this simplicity results, however, from an inference based on the mode of the joint posterior distribution of β and u rather than from posterior expectations or marginal modes. In other words, due to its equivalence to Schall's approach [31], it can be viewed as a procedure based on a linearization of a conditional model [35].

On the contrary, the GAR quasi-score method is an attempt to integrate out u in order to estimate the β and δ fixed effects. There is, however, a trade-off for an easy integration, i.e. i) to replace the data distribution by its quasi-likelihood counterpart, and ii) to approximate the variance–covariance matrix of data by a Taylor expansion about small intra-class correlations. Moreover, as pointed out by Knuiman and Laird [27], $\hat{u}$ solutions to equation (26) have no clear justification.

An additional level of approximation resorts to estimating G from formula (27) which mimics classical EM-type formulae for linear models. Model approximations, when estimating G as originally proposed by Harville and Mee [22] and Foulley et al. [11], could also be especially critical. For example, with binary data, such approximations may yield seriously biased estimators [4]. Fully EM marginal maximum likelihood or Bayesian posterior analysis based on MCMC (Monte Carlo Markov chains) methods [33] would be useful to improve the estimation of variance components and to get further in the implementation of heteroskedastic models [15]. An alternative procedure based on simulated moments was proposed by Jiang [26]; it provides consistent estimators for both fixed effects and random components. Although computationally attractive, this method can be quite inefficient when applied to small samples.

A comparison between MAP and GAR has been carried out via Monte Carlo simulation for an S-TM and for binary data by Hoeschele and Gianola [24]. Contrarily to expectations, MAP estimators of fixed effects were superior to GAR estimators in terms of bias and MSE, and $\hat{u}$ predictions were very close to each other. Preliminary simulation work carried out on a sire–maternal grand sire design assuming G known and relatively large differences in variances exactly indicate the same tendency as far as the comparison of these two methods is concerned: smaller MSE for β and δ fixed effects with MAP and very little difference in u prediction (-0.5 % in MSE and $+1.0$ % in R^2 for MAP versus GAR). That simulation, however, clearly shows the interest of selecting H-TM versus S–TM based EBVs (7 % in MSE and $+5$ to 14 % in R^2).

ACKNOWLEDGEMENTS

REFERENCES

[1] Breslow N.E., Clayton D.G., Approximate inference in generalized linear mixed models, J. Am. Stat. Assoc. 88 (1993) 9–25.

[2] Curnow R., Progeny testing for all-or-none traits when a multifactorial model applies, Biometrics 40 (1984) 375–382.

[3] Denis J.E. Jr, Schnabel R.B., Numerical Methods for Unconstrained Optimization and Non-linear Equations, Prentice-Hall Inc., Englewood Cliffs, 1983.

[4] Engel B., A simple illustration of the failure of PQL, IRREML and APHL as approximate ML methods for mixed models for binary data, Biom. J. 40 (1998) 141–154.

[5] Engel B., Keen A., A simple approach for the analysis of generalized linear mixed models, Statistica Neerlandica 48 (1994) 1–22.

[6] Engel B., Keen A., Contribution to the discussion of Lee and Nelder: Hierarchical generalized linear models, J. R. Statist. Soc. 58 (1996) 656–657.

[7] Falconer D.S., The inheritance of liability to certain diseases estimated from the incidence among relatives, Ann. Hum. Genet. 29 (1965) 51–76.

[8] Foulley J.L., Gianola D., Statistical analysis of ordered categorical data via a strutural heteroskedastic threshold model, Genet. Sel. Evol. 28 (1996) 249–273.

[9] Foulley J.L., Im S., A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models, Genet. Sel. Evol. 25 (1993) 101–107.

[10] Foulley J.L., Manfredi E., Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils, Genet. Sel. Evol. 23 (1991) 309–338.

[11] Foulley J.L., Im S., Gianola D., Hoeschele I., Empirical Bayes estimation of parameters for n polygenic binary traits, Genet. Sel. Evol. 19 (1987) 197–224.

[12] Foulley J.L., Gianola D., Im S., Genetic evaluation for discrete polygenic traits in animal breeding, in: Gianola D., Hammond K. (Eds.), Advances in Statistical Methods for Genetic Improvements of Livestock, Springer-Verlag, Heidelberg, 1990, pp. 361–396.

[13] Foulley J.L., Gianola D., San Cristobal M., Im S., A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models, J. Dairy Sci. 73 (1990) 1612–1624.

[14] Foulley J.L., San Cristobal M., Gianola D., Im S., Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models, Comput. Stat. Data Anal. 13 (1992) 291–305.

[15] Foulley J.L., Quaas R.L., Thaon d'Arnoldi C., A link function approach to heterogeneous variance components, Genet. Sel. Evol. 30 (1998) 27–43.

[16] Gianola D., Theory and analysis of threshold characters, J. Anim. Sci. 54 (1982) 1079–1096.

[17] Gilmour A.R., The estimation of genetic parameters for categorical traits, Ph.D. thesis, Massey University, NZ, 1983.

[18] Gilmour A.R., Anderson R.D., Rae A.L., The analysis of binomial data by a generalized linear mixed model, Biometrika 72 (1985) 593–599.

[19] Gilmour A.R., Anderson R.D., Rae A.L., Variance components on an underlying scale for ordered multiple threshold categorical data using a generalized linear mixed model, J. Anim. Breed. Genet. 104 (1987) 149–155.

[20] Goldstein H., Multilevel mixed linear models analysis using iterative generalized least squares, Biometrika 73 (1986) 43–56.

[21] Goldstein H., Non linear models with an application to discrete response data, Biometrika 78 (1991) 45–51.

[22] Harville D.A., Mee R.W., A mixed model procedure for analysing ordered categorical data, Biometrics 40 (1984) 393–408.

[23] Henderson C.R., Applications of Linear Models in Animal Breeding, University of Guelph, Guelph, 1984.

[24] Hoeschele I., Gianola D., Bayesian versus maximum quasi-likelihood methods for sire evaluation with categorical data, J. Dairy Sci. 72 (1989) 1569–1577.

[25] Householder A.S., The Theory of Matrices in Numerical Analysis, Dover Publications Inc., New York, 1975.

[26] Jiang J., Consistent estimators in generalized linear mixed models, J. Am. Stat. Assoc. 93 (1998) 720–729.

[27] Knuiman M., Laird N., Parameter estimation in variance component models for binary response data, in: Gianola D., Hammond K. (Eds.), Advances in Statistical Methods for Genetic Improvement of Livestock, Springer-Verlag, Heidelberg, 1990, pp. 194–207.

[28] Lee Y., Nelder J.A., Hierarchical generalized linear models, J. R. Statist. Soc. 58 (1996) 619–678.

[29] Liang K.Y., Zeger S.L., Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[30] Mc Cullagh P., Nelder J., Generalized Linear Models, 2nd ed., Chapman and Hall, London, 1989.

[31] Schall R., Estimation in generalized linear models with random effects, Biometrika 78 (1991) 719–727.

[32] Searle S.R., Linear Models for Unbalanced Data, John Wiley and Sons, New York, 1987.

[33] Sorensen D.A., Andersen S., Gianola D., Korsgaard I., Bayesian inference in threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229–249.

[34] Tallis G., The maximum likelihood estimation of correlation from contingency tables, Biometrics 18 (1962) 342–353.

[35] Trottier C., Estimation dans les modèles linéaires généralisés à effets aléatoires, Institut National Polytechnique de Grenoble: Mathématiques appliquées, 1998, 167 p.

[36] Wright S., An analysis of variability in number of digits in an inbred strain of guinea pigs, Genetics 19 (1934) 506–536.

[37] Wright S., The results of crosses between inbred strains of guinea pigs differing in number of digits, Genetics 19 (1934) 537–551.

[38] Xie M., Simpson D.G., Carroll R.J., Scaled link functions for heterogeneous ordinal response data, in: Gregoire T.G. et al. (Eds.), Modelling Longitudinal and Spatially Correlated Data, Lecture Notes in Statistics, 1997, 122, pp. 23–36.

[39] Zeger S.L., Liang K.Y., Albert P.S., Models for longitudinal data: a generalized estimating equation approach, Biometrics 44 (1988) 1049–1060.

APPENDIX

Decomposition of the data variance–covariance matrix

Let $\widetilde{\boldsymbol{y}}$ be the vector of the observed cumulative proportions of response in the $(J-1)$ first categories $\widetilde{\boldsymbol{y}} = (\widetilde{\boldsymbol{y}}_1', \ldots, \widetilde{\boldsymbol{y}}_i', \ldots, \widetilde{\boldsymbol{y}}_I')'$.

Let $\overline{\boldsymbol{y}}_i = (\overline{y}_{i1}, \overline{y}_{i2}, \ldots, \overline{y}_{ij}, \ldots, \overline{y}_{iJ})$ be the $(J \times 1)$ vector whose $\overline{y}_{ij}$ element is the frequency of responses in category j for subpopulation i.

Cumulative and elementary proportions are linked via:

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{L}\overline{\boldsymbol{y}}_i \tag{A1}$$

where $\boldsymbol{L}$ is a $((J-1) \times J)$ matrix built from a lower triangular $(J \times J)$ matrix of 1s, the last row of which is removed.

$\overline{\boldsymbol{y}}_i$ can, however, be expressed as a combination of indicator variables:

$$\overline{\boldsymbol{y}}_i = \frac{1}{n_i} \sum_{r=1}^{n_i} \boldsymbol{y}_{ir} \text{ with } \boldsymbol{y}_{ir} \text{ an indicator vector of dimension } (J \times 1), \text{ with 1 on}$$

the jth row if the observation r of the population i is in the category j. Then,

$$\text{Var}(\overline{\boldsymbol{y}}_i) = \frac{1}{n_i^2} \left[\sum_{r=1}^{n_i} \text{Var}(\boldsymbol{y}_{ir}) + \sum_{r \neq r'} \text{Cov}(\boldsymbol{y}_{ir}, \boldsymbol{y}_{ir'}) \right] \tag{A2}$$

$$= \frac{1}{n_i} \left[\text{Var}(\boldsymbol{y}_{ir}) - \text{Cov}(\boldsymbol{y}_{ir}, \boldsymbol{y}_{ir'}) \right] + \text{Cov}(\boldsymbol{y}_{ir}, \boldsymbol{y}_{ir'}) \tag{A3}$$

Now, $\text{Var}(\boldsymbol{y}_{ir}) = E(\boldsymbol{y}_{ir}\boldsymbol{y}_{ir}') - E(\boldsymbol{y}_{ir})E(\boldsymbol{y}_{ir})'$, with $E(\boldsymbol{y}_{ir}\boldsymbol{y}_{ir}') = \text{diag}\{E(\boldsymbol{y}_{ir})\}$, and $E(\boldsymbol{y}_{ir}) = \boldsymbol{\Pi}_i$, with $\boldsymbol{\Pi}_i$ corresponding to the $(J \times 1)$ vector of $(\pi_{ij})_{j=1,\ldots,J}$. Thus,

$$\text{Var}(\boldsymbol{y}_{ir}) = \text{diag}\{\boldsymbol{\Pi}_i\} - \boldsymbol{\Pi}_i\boldsymbol{\Pi}_i' \tag{A4}$$

Moreover, $\text{Cov}(\boldsymbol{y}_{ir}, \boldsymbol{y}_{ir'}') = E(\boldsymbol{y}_{ir}\boldsymbol{y}_{ir'}') - E(\boldsymbol{y}_{ir})E(\boldsymbol{y}_{ir'})'$. $E(\boldsymbol{y}_{ir}\boldsymbol{y}_{ir'}') = \boldsymbol{\Pi}_{2_{ii}}$, where $\boldsymbol{\Pi}_{2_{ii}}$ stands for a $(J \times J)$ matrix, for which the (jk) term represents the joint probability that an observation r of subpopulation i is in category j and that a different observation r' from the same subpopulation i is in category k, i.e. $(\boldsymbol{\Pi}_{2,ii})_{jk} = Pr(y_{ijr} = y_{ikr'} = 1)$. Thus,

$$\text{Cov}(\boldsymbol{y}_{ir}, \boldsymbol{y}_{ir'}') = \boldsymbol{\Pi}_{2_{ii}} - \boldsymbol{\Pi}_i\boldsymbol{\Pi}_i' \tag{A5}$$

Since, from equation (A1), $\text{Var}(\widetilde{\boldsymbol{y}}_i) = \boldsymbol{L}\text{Var}(\overline{\boldsymbol{y}}_i)\boldsymbol{L}'$, and using the decomposition in equation (A3) with formulae (A4) and (A5), one has:

$$\boldsymbol{L} \text{ diag}\{\boldsymbol{\Pi}_i\} \boldsymbol{L}' = \begin{pmatrix} \mu_{i1} & \mu_{i1} & \mu_{i1} & \cdots & \mu_{i1} \\ \mu_{i1} & \mu_{i2} & \mu_{i2} & \cdots & \mu_{i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{i1} & \mu_{i2} & \mu_{i3} & \cdots & \mu_{i(J-1)} \end{pmatrix}$$

$L\Pi_i\Pi_i'L' = \{\mu_{ij}\mu_{ik}\}_{(j,k=1,...,(J-1))}$, where $\mu_{ij} = \Phi(\gamma_{ij}) = \sum_{k=1}^{j}\pi_{ik}$, and because $L\operatorname{Var}(\boldsymbol{y}_{ir})L' = L\operatorname{diag}\{\Pi_i\}\ L' - L\ \Pi_i\Pi_i'\ L'$, this expression reduces to $L\operatorname{Var}(\boldsymbol{y}_{ir})L' = \Sigma_{0,ii}$ with

$$\Sigma_{0,ii} = \begin{pmatrix} \mu_{i1}(1-\mu_{i1}) & \mu_{i1}(1-\mu_{i2}) & \cdots & \cdots & \mu_{i1}(1-\mu_{i(J-1)}) \\ \mu_{i1}(1-\mu_{i2}) & \mu_{i2}(1-\mu_{i2}) & \cdots & \cdots & \mu_{i2}(1-\mu_{i(J-1)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_{i1}(1-\mu_{i(J-1)}) & \mu_{i2}(1-\mu_{i(J-1)}) & \cdots & \cdots & \mu_{i(J-1)}(1-\mu_{i(J-1)}) \end{pmatrix}$$

$$(A6)$$

As $\sum_{l=1}^{j}\sum_{m=1}^{k}\Pi_{2,ii}(lm) = \Phi_2(\gamma_{ij},\gamma_{ik};t_{ii})$, then

$$LE(\boldsymbol{y}_{ir}\boldsymbol{y}_{ir'}')L' = \{\Phi_2(\gamma_{ij},\gamma_{ik};t_{ii})\}_{(j,k=1,...(J-1))}$$

where t_{ii} stands for the correlation coefficient between ℓ_{ir} and $\ell_{ir'}$; Φ_2 is the CDF of the normal bivariate distribution, such that $\Phi_2(y_1,y_2;\rho) = P((Y_1 \leqslant y_1) \cap (Y_2 \leqslant y_2))$, Y_1 and Y_2 being identically distributed as $\mathcal{N}(0,1)$, with $\operatorname{Corr}(Y_1,Y_2) = \rho$. Inserting this in equation (32) leads to:

$$L\operatorname{Cov}(\boldsymbol{y}_{ir},\boldsymbol{y}_{ir'}')L' = \Sigma_{B,ii} = \{\Phi_2(\gamma_{ij},\gamma_{ik};t_{ii}) - \Phi(\gamma_{ij})\Phi(\gamma_{ik})\}_{(j,k=1,...,(J-1))}$$

$$(A7)$$

Hence, the (i,i) blocks of the variance–covariance matrix of $\widetilde{\boldsymbol{y}}$ can be written as:

$$\operatorname{Var}(\widetilde{\boldsymbol{y}}_i) = \frac{1}{n_i}(\Sigma_{0,ii} - \Sigma_{B,ii}) + \Sigma_{B,ii} \tag{A8}$$

The expression of the covariance calculated before can be easily generalized to the case of two distinct populations i and i'. Then, it follows:

$$L\operatorname{Cov}(\boldsymbol{y}_{ir},\boldsymbol{y}_{i'r'}')L' = \Sigma_{B,ii'} = \{\Phi_2(\gamma_{ij},\gamma_{i'k};t_{ii'}) - \Phi(\gamma_{ij})\Phi(\gamma_{i'k})\}_{(j,k=1,...,(J-1))}$$

$$(A9)$$

Finally, the variance–covariance matrix Σ of $\widetilde{\boldsymbol{y}}$ can be written as the sum of two components:

$$\Sigma = \Sigma_A + \Sigma_B \tag{A10}$$

where Σ_A is a block diagonal matrix defined as:

$$\Sigma_A = \bigoplus_{i=1}^{I}\Sigma_{A,ii} \tag{A11}$$

with $\Sigma_{A,ii} = \frac{1}{n_i}(\Sigma_{0,ii} - \Sigma_{B,ii})$, and $\Sigma_B = \{\Sigma_{B,ii'}\}_{(i,i'=1,...,I)}$, where $\Sigma_{0,ii}$ is given in equation (A6), and $\Sigma_{B,ii}$ and $\Sigma_{B,ii'}$ are given in equations (A7) and (A9).

Regarding correlation coefficients involved in equation (A7), one has

$$t_{ii} = \frac{\mathrm{Cov}(\ell_{ir}, \ell_{ir'})}{\sqrt{\mathrm{Var}(\ell_{ir})\mathrm{Var}(\ell_{ir'})}}$$

Since $\ell_{ir} = \mathbf{x}_i'\boldsymbol{\beta} + \sigma_{u_i}\mathbf{z}_i'\mathbf{u}^* + e_{ir}$, then:

$$\mathrm{Var}(\ell_{ir}) = \mathrm{Var}(\ell_{ir'}) = \mathbf{z}_i'\mathbf{A}\mathbf{z}_i\sigma_{u_i}^2 + \sigma_{e_i}^2$$

$$\mathrm{Cov}(\ell_{ir}, \ell_{ir'}) = \mathbf{z}_i'\mathbf{A}\mathbf{z}_i\sigma_{u_i}^2$$

Hence,

$$t_{ii} = \frac{\mathbf{z}_i'\mathbf{A}\mathbf{z}_i\sigma_{u_i}^2}{\mathbf{z}_i'\mathbf{A}\mathbf{z}_i\sigma_{u_i}^2 + \sigma_{e_i}^2} = \frac{\alpha_i\rho^2}{1 + \alpha_i\rho^2} \tag{A12}$$

Similarly for $t_{ii'}$, in equation (A9):

$$t_{ii'} = \frac{\mathbf{z}_i'\mathbf{A}\mathbf{z}_{i'}\sigma_{u_i}\sigma_{u_{i'}}}{\sqrt{\mathbf{z}_i'\mathbf{A}\mathbf{z}_i\sigma_{u_i}^2 + \sigma_{e_i}^2}\sqrt{\mathbf{z}_{i'}'\mathbf{A}\mathbf{z}_{i'}\sigma_{u_{i'}}^2 + \sigma_{e_{i'}}^2}} \tag{A13}$$

$$= \frac{\mathbf{z}_i'\mathbf{A}\mathbf{z}_{i'}\rho^2}{\sqrt{1 + \alpha_i\rho^2}\sqrt{1 + \alpha_{i'}\rho^2}} \tag{A14}$$

with $\alpha_i = \mathbf{z}_i'\mathbf{A}\mathbf{z}_i$ and $\rho^2 = \sigma_{u_i}^2/\sigma_{e_i}^2$

Calculation of $\mathbf{R}^{-1}$

$\mathbf{R}$ is a block diagonal matrix such that $\mathbf{R} = \phi^{-1}\boldsymbol{\Sigma}_A^*\phi^{-1} = \overset{I}{\underset{i=1}{\oplus}} \mathbf{R}_i$, and thus:

$$\mathbf{R}^{-1} = \overset{I}{\underset{i=1}{\oplus}} \mathbf{R}_i^{-1} \tag{A15}$$

Now,

$$\mathbf{R}_i = \phi_i^{-1}\boldsymbol{\Sigma}_{A,ii}^*\phi_i^{-1} = \frac{1}{n_i}(\mathbf{B}_i - \mathbf{1}\mathbf{1}'t_{ii}) \tag{A16}$$

where $\mathbf{B}_i = \phi_i^{-1}\boldsymbol{\Sigma}_{0,ii}\phi_i^{-1}$ and $\mathbf{1}\mathbf{1}'t_{ii} = \mathbf{z}_i^*\mathbf{G}\mathbf{z}_i^{*'}$. Using Woodbury's formula (see e.g. Householder [25] p. 124) for inverting $\mathbf{R}_i$, one obtains:

$$(\mathbf{B}_i - \mathbf{1}\mathbf{1}'t_{ii})^{-1} = \mathbf{B}_i^{-1} + \mathbf{B}_i^{-1}\mathbf{1}(t_{ii}^{-1} - \mathbf{1}'\mathbf{B}_i^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{B}_i^{-1} \tag{A17}$$

$$= \mathbf{B}_i^{-1}(\mathbf{B}_i + \frac{\mathbf{1}\mathbf{1}'}{c_i})\mathbf{B}_i^{-1} \tag{A18}$$

where $c_i = t_{ii}^{-1} - \mathbf{1}'\mathbf{B}_i^{-1}\mathbf{1}$.

Now, $\boldsymbol{B}_i^{-1}$ has a known tridiagonal structure (see e.g. Mc Cullagh and Nelder [30] p. 168) such that:

$$
\begin{aligned}
b_{ii}^{-1}(j,j) &= \phi(\gamma_{ij})^2(\pi_{ij}^{-1} + \pi_{i,j+1}^{-1}) \\
b_{ii}^{-1}(j,j+1) &= -\phi(\gamma_{ij})\phi(\gamma_{i(j+1)})\pi_{i,j+1}^{-1} = b_{ii}^{-1}(j+1,j)
\end{aligned}
\tag{A19}
$$

where ϕ stands for the density of $\mathcal{N}(0,1)$.

Expressions for the coefficients of equation (26)

The left hand size of the system is:

$$
\begin{pmatrix}
\boldsymbol{T}'(\boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{H})^{(t)}\boldsymbol{T} & \boldsymbol{T}'(\boldsymbol{H}'\boldsymbol{R}^{-1}\boldsymbol{M})^{(t)}\boldsymbol{Z} \\
\boldsymbol{Z}'(\boldsymbol{M}'\boldsymbol{R}^{-1}\boldsymbol{H})^{(t)}\boldsymbol{T} & \boldsymbol{Z}'(\boldsymbol{M}'\boldsymbol{R}^{-1}\boldsymbol{M})^{(t)}\boldsymbol{Z} + \boldsymbol{G}^{-1}
\end{pmatrix}
$$

which can be partitioned according to $\boldsymbol{\xi}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and $\boldsymbol{u}$, as:

$$
\begin{pmatrix}
\boldsymbol{IH}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_\xi\boldsymbol{I} & \boldsymbol{IH}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_\eta\boldsymbol{X} & \boldsymbol{IH}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_\nu\boldsymbol{P} & \boldsymbol{IH}_\xi'\boldsymbol{R}^{-1}\boldsymbol{M}\boldsymbol{Z} \\
& \boldsymbol{X}'\boldsymbol{H}_\eta'\boldsymbol{R}^{-1}\boldsymbol{H}_\eta\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{H}_\eta'\boldsymbol{R}^{-1}\boldsymbol{H}_\nu\boldsymbol{P} & \boldsymbol{X}'\boldsymbol{H}_\eta'\boldsymbol{R}^{-1}\boldsymbol{M}\boldsymbol{Z} \\
& & \boldsymbol{P}'\boldsymbol{H}_\nu'\boldsymbol{R}^{-1}\boldsymbol{H}_\nu\boldsymbol{P} & \boldsymbol{P}'\boldsymbol{H}_\nu'\boldsymbol{R}^{-1}\boldsymbol{M}\boldsymbol{Z} \\
\text{(symmetric)} & & & \boldsymbol{Z}'\boldsymbol{M}'\boldsymbol{R}^{-1}\boldsymbol{M}\boldsymbol{Z} + \boldsymbol{G}^{-1}
\end{pmatrix}
$$

$$
\boldsymbol{H}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_{\xi((J-1)\times(J-1))} = \sum_{i=1}^{I} \frac{\boldsymbol{R}_i^{-1}}{\sigma_{e_i}^2(1+\alpha_i\rho^2)}
$$

$$
\boldsymbol{H}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_{\eta((J-1)\times I)} = \left(\frac{-1}{\sigma_{e_i}^2(1+\alpha_i\rho^2)}\boldsymbol{R}_i^{-1}\boldsymbol{1}_{J-1} \;\middle|\; \frac{-1}{\sigma_{e_{i'}}^2(1+\alpha_{i'}\rho^2)}\boldsymbol{R}_{i'}^{-1}\boldsymbol{1}_{J-1} \right)
$$

where $(.|.)$ represents a description of the matrix column by column.

$$
\boldsymbol{H}_\xi'\boldsymbol{R}^{-1}\boldsymbol{H}_{\nu((J-1)\times I)} = \left(\frac{-1}{\sigma_{e_i}\sqrt{1+\alpha_i\rho^2}}\boldsymbol{R}_i^{-1}\gamma_i \;\middle|\; \frac{-1}{\sigma_{e_{i'}}\sqrt{1+\alpha_{i'}\rho^2}}\boldsymbol{R}_{i'}^{-1}\gamma_{i'} \right)
$$

where $\gamma_{i((J-1)\times 1)} = \{\gamma_{ij}\}_{(j=1,\ldots,J-1)}$

$$
\boldsymbol{H}_\xi'\boldsymbol{R}^{-1}\boldsymbol{M}_{((J-1)\times I)} = \left(\frac{-1}{\sigma_{e_i}(1+\alpha_i\rho^2)}\boldsymbol{R}_i^{-1}\boldsymbol{1}_{J-1} \;\middle|\; \frac{-1}{\sigma_{e_{i'}}(1+\alpha_{i'}\rho^2)}\boldsymbol{R}_{i'}^{-1}\boldsymbol{1}_{J-1} \right)
$$

$$
\boldsymbol{H}_\eta'\boldsymbol{R}^{-1}\boldsymbol{H}_{\eta(I\times I)} = \text{Diag}\{ \frac{1}{\sigma_{e_i}^2(1+\alpha_i\rho^2)}\boldsymbol{1}_{J-1}'\boldsymbol{R}_i^{-1}\boldsymbol{1}_{J-1} \}
$$

$$
\boldsymbol{H}_\eta'\boldsymbol{R}^{-1}\boldsymbol{H}_{\nu(I\times I)} = \text{Diag}\{ \frac{1}{\sigma_{e_i}\sqrt{1+\alpha_i\rho^2}}\boldsymbol{1}_{J-1}'\boldsymbol{R}_i^{-1}\gamma_i \}
$$

$$\boldsymbol{H_\eta}^{'}\boldsymbol{R}^{-1}\boldsymbol{M}_{(I\times I)} = \mathrm{Diag}\{\frac{1}{\sigma_{e_i}(1+\alpha_i\rho^2)}\boldsymbol{1}^{'}_{J-1}\boldsymbol{R}_i^{-1}\boldsymbol{1}_{J-1}\}$$

$$\boldsymbol{H_\nu}^{'}\boldsymbol{R}^{-1}\boldsymbol{H}_{\nu(I\times I)} = \mathrm{Diag}\{\gamma^{'}_i\boldsymbol{R}_i^{-1}\gamma_i\}$$

$$\boldsymbol{H_\nu}^{'}\boldsymbol{R}^{-1}\boldsymbol{M}_{(I\times I)} = \mathrm{Diag}\{\frac{1}{\sqrt{1+\alpha_i\rho^2}}\boldsymbol{1}^{'}_{J-1}\boldsymbol{R}_i^{-1}\gamma_i\}$$

$$\boldsymbol{M}^{'}\boldsymbol{R}^{-1}\boldsymbol{M}_{(I\times I)} = \mathrm{Diag}\{\frac{1}{(1+\alpha_i\rho^2)}\boldsymbol{1}^{'}_{J-1}\boldsymbol{R}_i^{-1}\boldsymbol{1}_{J-1}\}$$

The second part of the system is:

$$\begin{pmatrix} \boldsymbol{T}^{'}(\boldsymbol{H}^{'}\boldsymbol{R}^{-1}\boldsymbol{H})^{(t)}\boldsymbol{T}\theta^{(t)} + \boldsymbol{T}^{'}(\boldsymbol{H}^{'}\boldsymbol{R}^{-1}\phi^{-1})^{(t)}(\tilde{\boldsymbol{y}}-\mu^{(t)}) \\ \boldsymbol{Z}^{'}(\boldsymbol{M}^{'}\boldsymbol{R}^{-1}\boldsymbol{H})^{(t)}\boldsymbol{T}\theta^{(t)} + \boldsymbol{Z}^{'}(\boldsymbol{M}^{'}\boldsymbol{R}^{-1}\phi^{-1})^{(t)}(\tilde{\boldsymbol{y}}-\mu^{(t)}) \end{pmatrix}$$

with

$$\boldsymbol{H_\xi}^{'}\boldsymbol{R}^{-1}\phi^{-1}(\tilde{\boldsymbol{y}}-\mu)_{((J-1)\times 1)} = \sum_{i=1}^{I}\frac{1}{\sigma_{e_i}\sqrt{1+\alpha_i\rho^2}}\boldsymbol{R}_i^{-1}\varepsilon_i$$

$$\boldsymbol{H_\eta}^{'}\boldsymbol{R}^{-1}\phi^{-1}(\tilde{\boldsymbol{y}}-\mu)_{(I\times 1)} = \{\frac{-1}{\sigma_{e_i}\sqrt{1+\alpha_i\rho^2}}\boldsymbol{1}^{'}_{J-1}\boldsymbol{R}_i^{-1}\varepsilon_i\}$$

where $\varepsilon_i = \phi_i^{-1}(\tilde{\boldsymbol{y}}_i - \mu_i)$

$$\boldsymbol{H_\nu}^{'}\boldsymbol{R}^{-1}\phi^{-1}(\tilde{\boldsymbol{y}}-\mu)_{(I\times 1)} = \{-\gamma^{'}_i\boldsymbol{R}_i^{-1}\varepsilon_i\}$$

$$\boldsymbol{M}^{'}\boldsymbol{R}^{-1}\phi^{-1}(\tilde{\boldsymbol{y}}-\mu)_{(I\times 1)} = \{\frac{-1}{\sqrt{1+\alpha_i\rho^2}}\boldsymbol{1}^{'}_{J-1}\boldsymbol{R}_i^{-1}\varepsilon_i\}$$