Supplementary Materials for: A Mixed Hidden Markov Model for Multivariate Monotone Disease Processes in the Presence of Measurement Errors

Lizbeth Naranjo [1], Emmanuel Lesaffre [2], and Carlos J. Pérez [3]

[1] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, Mexico
[2] L-BioStat, School of Public Health, KU Leuven, Leuven, Belgium
[3] Departamento de Matemáticas, Universidad de Extremadura, Cáceres, Spain

Address for correspondence: Lizbeth Naranjo, Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, C.P. 04510, Ciudad de México, Mexico.
E-mail: lizbethna@ciencias.unam.mx.
Phone: (+52) 55 5622 3899 Ext.:45735.
Fax: 01 55 5622 4859.

Abstract: Motivated by a longitudinal oral health study, the Signal-Tandmobiel® study, an inhomogeneous mixed hidden Markov model with continuous state-space is proposed to explain the caries disease process in children between 6 and 12 years of age. The binary caries experience outcomes are subject to misclassification. We modeled this misclassification process via a longitudinal latent continuous response subject to a measurement error process and showing a monotone behaviour. The baseline distributions of the unobservable continuous processes are defined as a function of the covariates through the specification of conditional distributions making use of the Markov property. In addition, random effects are considered to model the relationships among the multivariate responses. Our approach is in contrast with a previous approach working on the binary outcome scale. This method requires conditional independence of the possibly corrupted binary outcomes on the true binary outcomes. We assumed conditional independence on the latent scale, which is a weaker assumption than conditional independence on the binary scale. The aim of this paper is therefore to show the properties of a model for a progressive longitudinal response with misclassification on the manifest scale but modeled on the latent scale. The model parameters are estimated in a Bayesian way using an efficient Markov chain Monte Carlo method. The model performance is shown through a simulation-based example, and the analysis of the motivating dataset is presented.

Key words: Bayesian analysis; Conditional independence; Measurement error; Misclassification; Mixed hidden Markov model; Monotone continuous process

1 Conditional independence

Based on García-Zattera et al. (2007), we have adapted some of the results to show that conditional independence for a latent continuous vector does not imply conditional independence for a binary response vector for a simplified version of our model.

Proposition: Assume that the vector $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_K^*)$ follows a monotone homogeneous hidden Markov model with continuous state-space, considering measurement error and parameterized by the following:

$$\begin{aligned}
Y_k^* &= \begin{cases} 1 & \text{if } W_k^* > 0 \\ 0 & \text{if } W_k^* \leq 0 \end{cases}, \\
W_k^* &\sim \mathrm{N}(W_k, \sigma^2), \\
W_1 &\sim \mathrm{N}(\eta_1, 1), \\
W_k | W_{k-1} = w_{k-1} &\sim \mathrm{N}(\eta_k, 1)\mathrm{I}[W_k \geq w_{k-1}].
\end{aligned}$$

Assume that the measurement error process is characterized by the equivalent univariate versions of Assumptions (A.1)-(A.6), see Appendix A in the paper. Then, the responses are not conditionally independent.

Proof: The conditional independence assumptions (A.1)-(A.6) show that:

$$\begin{aligned}
\mathrm{p}(W_1^*, W_2^* | W_1, W_2) &= \mathrm{p}(W_1^* | W_1, W_2)\mathrm{p}(W_2^* | W_1, W_2), \\
\mathrm{p}(W_2^*, W_1, W_3 | W_2) &= \mathrm{p}(W_2^* | W_2)\mathrm{p}(W_1, W_3 | W_2),
\end{aligned}$$

and this imply that:

$$\begin{aligned}
\mathrm{p}(W_1^*, W_3^* | W_2) &= \mathrm{p}(W_1^* | W_2)\mathrm{p}(W_3^* | W_2), \\
\mathrm{p}(W_1, W_3 | W_2) &= \mathrm{p}(W_1 | W_2)\mathrm{p}(W_3 | W_2).
\end{aligned}$$

However, this does not happen with the response binary variable $\boldsymbol{Y}^*$, i.e.:

$$\mathrm{p}(Y_1^*, Y_3^* | Y_2^*) \neq \mathrm{p}(Y_1^* | Y_2^*)\mathrm{p}(Y_3^* | Y_2^*)$$

This is because:

$$\begin{aligned}
&\mathrm{p}(Y_1^* = 1, Y_3^* = 1 | Y_2^* = 1) \qquad (1.1) \\
&= \mathrm{p}(W_1^* > 0, W_3^* > 0 | W_2^* > 0) \\
&= \frac{\int_0^\infty \int_0^\infty \int_0^\infty f(w_1^*, w_3^*, w_2^*) \mathrm{d}w_1^* \mathrm{d}w_3^* \mathrm{d}w_2^*}{\int_0^\infty f(w_2^*) \mathrm{d}w_2^*} \\
&= \frac{\int_0^\infty \int_0^\infty \int_0^\infty \int\int\int f(w_3^*|w_3) f(w_2^*|w_2) f(w_1^*|w_1) f(w_3|w_2) f(w_2|w_1) f(w_1) \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3 \mathrm{d}w_1^* \mathrm{d}w_3^* \mathrm{d}w_2^*}{\int_0^\infty \int\int f(w_2^*|w_2) f(w_2|w_1) f(w_1) \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_2^*} \\
&= \frac{\int\int\int \Phi(w_3; 0, \sigma^2)\Phi(w_2; 0, \sigma^2)\Phi(w_1; 0, \sigma^2) \frac{\phi(w_3; \eta_3, 1)}{1 - \Phi(w_2; \eta_3, 1)} \frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)} \phi(w_1; \eta_1, 1) \mathrm{d}w_1 \mathrm{d}w_2 \mathrm{d}w_3}{\int\int \Phi(w_2; 0, \sigma^2) \frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)} \phi(w_1; \eta_1, 1) \mathrm{d}w_1 \mathrm{d}w_2},
\end{aligned}$$

where $\Phi(w^*; w, \sigma^2)$ denotes the cumulative distribution function of the random variable (r.v.) W^* having a distribution $W^* \sim \mathrm{N}(w, \sigma^2)$; and $\phi(w^*; w, \sigma^2)$ denotes the probability density function of the r.v. $W^* \sim \mathrm{N}(w, \sigma^2)$ evaluated in w^*. But,

$$\begin{aligned}
&\mathrm{p}(Y_1^* = 1|Y_2^* = 1)\mathrm{p}(Y_3^* = 1|Y_2^* = 1) \qquad (1.2)\\
&= \mathrm{p}(W_1^* > 0|W_2^* > 0)\mathrm{p}(W_3^* > 0|W_2^* > 0)\\
&= \frac{\int_0^\infty \int_0^\infty \int\int f(w_1^*, w_2^*, w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_1^*\mathrm{d}w_2^*}{\int_0^\infty \int\int f(w_2^*, w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_2^*}\\
&\quad\times\frac{\int_0^\infty \int_0^\infty \int\int\int f(w_3^*, w_2^*, w_1, w_2, w_3)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3\mathrm{d}w_3^*\mathrm{d}w_2^*}{\int_0^\infty \int\int f(w_2^*, w_1, w_2)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_2^*}\\
&= \frac{\int_0^\infty \int_0^\infty \int\int f(w_2^*|w_2)f(w_1^*|w_1)f(w_2|w_1)f(w_1)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_1^*\mathrm{d}w_2^*}{\int_0^\infty \int\int f(w_2^*|w_2)f(w_2|w_1)f(w_1)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_2^*}\\
&\quad\times\frac{\int_0^\infty \int_0^\infty \int\int\int f(w_3^*|w_3)f(w_2^*|w_2)f(w_3|w_2)f(w_2|w_1)f(w_1)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3\mathrm{d}w_3^*\mathrm{d}w_2^*}{\int_0^\infty \int\int f(w_2^*|w_2)f(w_2|w_1)f(w_1)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_2^*}\\
&= \frac{\int\int \Phi(w_2; 0, \sigma^2)\Phi(w_1; 0, \sigma^2)\frac{\phi(w_2;\eta_2,1)}{1-\Phi(w_1;\eta_2,1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2}{\int\int \Phi(w_2; 0, \sigma^2)\frac{\phi(w_2;\eta_2,1)}{1-\Phi(w_1;\eta_2,1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2}\\
&\quad\times\frac{\int\int\int \Phi(w_3; 0, \sigma^2)\Phi(w_2; 0, \sigma^2)\frac{\phi(w_3;\eta_3,1)}{1-\Phi(w_2;\eta_3,1)}\frac{\phi(w_2;\eta_2,1)}{1-\Phi(w_1;\eta_2,1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2\mathrm{d}w_3}{\int\int \Phi(w_2; 0, \sigma^2)\frac{\phi(w_2;\eta_2,1)}{1-\Phi(w_1;\eta_2,1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2}.
\end{aligned}$$

Expressions (1.1) and (1.2) are different. Expression (1.1) is larger than expression (1.2), because

$$E(g_1(X)g_2(X)) \geq E(g_1(X))E(g_2(X)),$$

holds for nondecreasing functions g_1 and g_2.

Although there are no simple solutions of these integrals, it is clear that

$$\mathrm{p}(Y_1^*, Y_3^*|Y_2^*) \quad \geq \quad \mathrm{p}(Y_1^*|Y_2^*)\mathrm{p}(Y_3^*|Y_2^*).$$

This shows that conditional independence on the latent scale does not imply conditional independence on the observed scale and vice versa. ■

2 Simulation-based experiment for conditional independence

By simulation it can be shown how there is conditional independence on latent scale, but not on the manifest scale.

Data have been simulated by considering multivariate normal distribution, $\boldsymbol{W} \sim \mathrm{N}(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is the variance-covariance matrix, where the variances vary in the interval $(0.1, 10)$, and with zero non-diagonal elements, with the property that always:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Samples of size $n = 10000$ are simulated for $\boldsymbol{W}$.

Then, the binary responses were generated by dichotomizing $\boldsymbol{W}$, i.e. $Y = 1$ if $W > 0$ and $Y = 0$ if $W \leq 0$. The association between Y_j and Y_l conditional on $Y_h = y$ $(Y_j, Y_l | Y_h = y)$ is reviewed, where y is 0 or 1, first the tables of 2×2 are obtained, and second the Pearson's contingency coefficient are computed.

The `R` code is the following.

```
library(MASS)
library(corpcor)
library(DescTools)
vv = seq(0.1,10,0.5)   ### different values for the precisions
m = length(vv)
RHO = matrix(NA,m^3,3*4)   ### save results here
N = 10000   ### sample size
idx = 0
for(i1 in 1:m){  for(i2 in 1:m){  for(i3 in 1:m){
    v1 = vv[i1]   ### diagonal 1
    v2 = vv[i2]   ### diagonal 2
    v3 = vv[i3]   ### diagonal 3
    R = matrix(c(1,0,0, 0,1,0, 0,0,1),3,3) ### partial correlation matrix
    A = matrix(0,3,3)
    diag(A) = c(v1,v2,v3)   ### precisions = 1/variances
    V = solve(sqrt(A) %*% R %*% sqrt(A))   ### variance covariance matrix
    W = mvrnorm(n=N, mu=rep(0,3), Sigma=V)   ### simulate latent scale
    Y = ifelse(W>0,1,0)   ### manifiest scale
    idx = idx+1   ### index to save results
    RHO[idx,1:3] = c(v1,v2,v3)   ### precisions
    RHO[idx,4:6] = c(V[1,1],V[2,2],V[3,3])   ### variances
    ### Pearson's contingency coefficient
    RHO[idx,7] = ContCoef(Y[Y[,3]==0,1],Y[Y[,3]==0,2])
    RHO[idx,8] = ContCoef(Y[Y[,3]==1,1],Y[Y[,3]==1,2])
```

```
    RHO[idx,9] = ContCoef(Y[Y[,2]==0,1],Y[Y[,2]==0,3])
    RHO[idx,10] = ContCoef(Y[Y[,2]==1,1],Y[Y[,2]==1,3])
    RHO[idx,11] = ContCoef(Y[Y[,1]==0,2],Y[Y[,1]==0,3])
    RHO[idx,12] = ContCoef(Y[Y[,1]==1,2],Y[Y[,1]==1,3])
} } }
hist(c(RHO[,7],RHO[,8],RHO[,9],RHO[,10],RHO[,11],RHO[,12]) , nclass=50,
     main="Pearson's contingency coefficient",xlab="")
```

In order to compute the variance-covariance matrix $\mathbf{V}$ from the partial correlation matrix $\mathbf{R}$ we use the property that $\mathbf{R} = \text{diag}(\mathbf{V}^{-1})^{-1/2}\mathbf{V}^{-1}\text{diag}(\mathbf{V}^{-1})^{-1/2}$, see e.g. Whittaker (1990).

Figure 1 shows their contingency coefficients for $Y_j, Y_l|Y_h = y$. Notice that with conditional independence on latent scale, one can assume conditional dependence on manifest scale, and Figure 1 shows the size of this dependency. This figure shows that the conditional dependence on the manifest scale is rather limited, but the conditional dependence could have a more important effect with more responses as in our study.

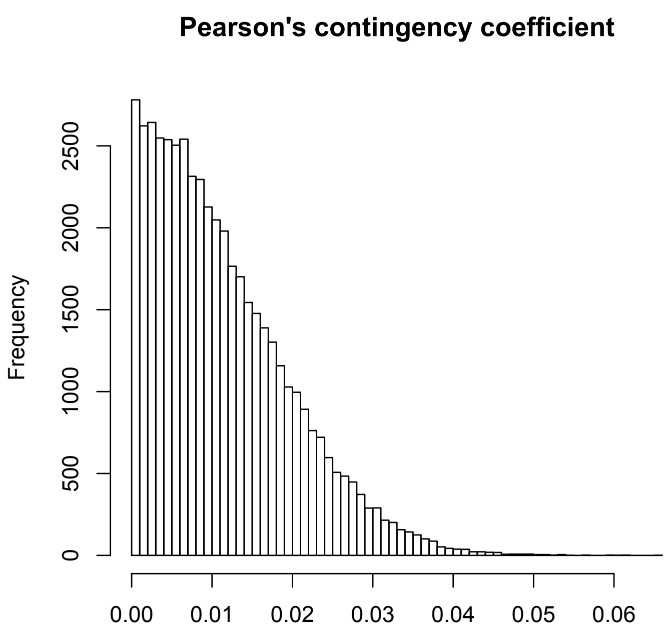

Figure 1: Pearson's contingency coefficients for $Y_j, Y_l|Y_h = y$.

3 Identifiability

Based on the propositions 2 and 3 presented in García-Zattera et al. (2012), we have derived a proposition that shows the identifiability of the parameters for a simplified version of our model.

Proposition: Assume that the vector $\boldsymbol{Y}^* = (Y_1^*, \ldots, Y_K^*)$ follows a monotone homogeneous hidden Markov model with continuous state-space, considering measurement error, parameterized by the following:

$$\begin{aligned} Y_k^* &= \begin{cases} 1 & \text{if } W_k^* > 0 \\ 0 & \text{if } W_k^* \leq 0 \end{cases}, \\ W_k^* &\sim \mathrm{N}(W_k, \sigma^2), \\ W_1 &\sim \mathrm{N}(\eta_1, 1), \\ W_k | W_{k-1} = w_{k-1} &\sim \mathrm{N}(\eta_k, 1)\mathrm{I}[W_k \geq w_{k-1}]. \end{aligned}$$

Assume that the measurement error process is characterized by the equivalent univariate versions of Assumptions (A.1)-(A.6), see Appendix A in the paper. Then, the parameters are identified.

Proof: The proof is based on the expression of the parameters of interest as functions of other identified quantities. Assume that two time points are considered, $K = 2$. Let $Q_1 = \mathrm{p}(Y_1^* = 0, Y_2^* = 0)$, $Q_2 = \mathrm{p}(Y_1^* = 1, Y_2^* = 0)$, $Q_3 = \mathrm{p}(Y_1^* = 0, Y_2^* = 1)$, and $Q_4 = \mathrm{p}(Y_1^* = 1, Y_2^* = 1)$, be the corresponding probabilities, which are identified quantities. These probabilities are the following functions of the parameters of interest:

$$\begin{aligned} Q_1 &= \mathrm{p}(Y_1^* = 0, Y_2^* = 0) = \mathrm{p}(W_1^* \leq 0, W_2^* \leq 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{p}(W_2^* \leq 0 | w_2) \mathrm{p}(W_1^* \leq 0 | w_1) \mathrm{p}(w_2 | w_1) \mathrm{p}(w_1) \mathrm{d}w_1 \mathrm{d}w_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(0; w_2, \sigma^2) \Phi(0; w_1, \sigma^2) \frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)} \phi(w_1; \eta_1, 1) \mathrm{d}w_1 \mathrm{d}w_2, \end{aligned}$$

where $\Phi(w^*; w, \sigma^2)$ denotes the cumulative distribution function of the random variable (r.v.) W^* having a distribution $W^* \sim \mathrm{N}(w, \sigma^2)$; and $\phi(w^*; w, \sigma^2)$ denotes the probability density function of the r.v. $W^* \sim \mathrm{N}(w, \sigma^2)$ evaluated in w^*. Also,

$$\begin{aligned} Q_2 &= \mathrm{p}(Y_1^* = 1, Y_2^* = 0) = \mathrm{p}(W_1^* > 0, W_2^* \leq 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{p}(W_2^* \leq 0 | w_2) \mathrm{p}(W_1^* > 0 | w_1) \mathrm{p}(w_2 | w_1) \mathrm{p}(w_1) \mathrm{d}w_1 \mathrm{d}w_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi(0; w_2, \sigma^2) [1 - \Phi(0; w_1, \sigma^2)] \frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)} \phi(w_1; \eta_1, 1) \mathrm{d}w_1 \mathrm{d}w_2, \end{aligned}$$

$$
\begin{aligned}
Q_3 &= \mathrm{p}(Y_1^* = 0, Y_2^* = 1) = \mathrm{p}(W_1^* \le 0, W_2^* > 0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{p}(W_2^* > 0|w_2)\mathrm{p}(W_1^* \le 0|w_1)\mathrm{p}(w_2|w_1)\mathrm{p}(w_1)\mathrm{d}w_1\mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [1 - \Phi(0; w_2, \sigma^2)]\Phi(0; w_1, \sigma^2)\frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2,
\end{aligned}
$$

$$
\begin{aligned}
Q_4 &= \mathrm{p}(Y_1^* = 1, Y_2^* = 1) = \mathrm{p}(W_1^* > 0, W_2^* > 0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{p}(W_2^* > 0|w_2)\mathrm{p}(W_1^* > 0|w_1)\mathrm{p}(w_2|w_1)\mathrm{p}(w_1)\mathrm{d}w_1\mathrm{d}w_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [1 - \Phi(0; w_2, \sigma^2)][1 - \Phi(0; w_1, \sigma^2)]\frac{\phi(w_2; \eta_2, 1)}{1 - \Phi(w_1; \eta_2, 1)}\phi(w_1; \eta_1, 1)\mathrm{d}w_1\mathrm{d}w_2.
\end{aligned}
$$

Therefore, parameters η_1, η_2 and σ^2 can be identified by functions of identified quantities Q_1, Q_2, Q_3 and Q_4. ■

The proof of the multivariate case is difficult to derive, because there is no exact deterministic version to show the identifiability of the parameters.

4 Identifiability on covariance matrix of the random effects

Usually, the prior distributions for covariance matrix Ω^P and Ω^I are Wishart distributions. However, in this paper, Ω^P and Ω^I are covariance matrices of the random intercepts $\boldsymbol{u}_i^P = (u_{i1}^P, \ldots, u_{iJ}^P)'$ and $\boldsymbol{u}_i^I = (u_{i1}^I, \ldots, u_{iJ}^I)'$, because $\boldsymbol{u}_i^P \sim \mathrm{N}_J(\mathbf{0}, \Omega^P)$ and $\boldsymbol{u}_i^I \sim \mathrm{N}_J(\mathbf{0}, \Omega^I)$, and they are part of the linear predictor,

$$
\begin{aligned}
\eta_{ij1} &= \alpha_1 + \boldsymbol{x}'_{ij}\boldsymbol{\beta}^P + \boldsymbol{z}'_{ij1}\boldsymbol{\gamma}^P + u_{ij}^P, \\
\eta_{ijk} &= \alpha_k + \boldsymbol{x}'_{ij}\boldsymbol{\beta}^I + \boldsymbol{z}'_{ijk}\boldsymbol{\gamma}^I + u_{ij}^I, \quad \text{for } k = 2, \ldots, K, \\
W_{ij1} &\sim \mathrm{N}(\eta_{ij1}, 1), \\
W_{ijk}|W_{ij,k-1} = w_{ij,k-1} &\sim \mathrm{N}(\eta_{ijk}, 1)\mathrm{I}[W_{ijk} \geq w_{ij,k-1}],
\end{aligned}
$$

where the W_{ijk}'s are latent variables.

A constraint must be given for Ω^P and Ω^I, otherwise convergence of MCMC chains will fail due to the lack of the identifiability of the parameters. In the paper, we have used the prior distributions proposed by Curtis (2010).

Based on the proposal of Curtis (2010), Ω^P is parametrized in terms of its Cholesky decomposition $\Omega^P = \Gamma^P\Gamma^{P\prime}$, where Γ^P is a lower triangular matrix, with entries equal to one on the diagonal, and unrestricted entries below the diagonal, i.e. $\Gamma_{ll}^P = 1$ for $l = 1, \ldots, J$, $\Gamma_{l_1 l_2}^P = 0$ for $l_1 = 1, \ldots, J-1$ and $l_2 = l_1+1, \ldots, J$, and $\Gamma_{l_1 l_2}^P \sim \mathrm{N}(0, 1)$ for $l_1 = 2, \ldots, J$ and $l_2 = 1, \ldots, l_1 - 1$. Setting the first element of Γ^P equal to one ensures that the first element of Ω^P is also equal to one, and therefore the first variance of Ω^P is constant to avoid lack of identifiability of parameters. Analogous prior distribution is defined on Ω^I.

Therefore, when a constraint with the parameters is used, the convergence is achieved, all the parameters are detectable, and there is no lack of identifiability. This effect is analogue to the one with the multivariate probit model. When latent variables are introduced, a constraint on the covariance matrix of the latent variables is needed. In fact, in the multivariate probit model, a correlation matrix is used, see e.g. Chib and Greenberg (1998).

In order to show the lack of identifiability and lack of convergence when no constraints are considered, we have present a simple example based on two simulated datasets. They have been simulated by the same way: $N = 1000$ subjects, $J = 3$ measures, $K = 5$ time points, $p = 2$ variables, $q = 2$ time-varying variables, where $x_{il} \sim \mathrm{U}(0, 1)$ and $z_{ikl} \sim \mathrm{U}(0, 1)$, $\boldsymbol{\alpha} = (-4, \ldots, -4)$, $\boldsymbol{\beta}^P = (1.5, 1.5)$, $\boldsymbol{\beta}^I = (1.5, 1.5)$, $\boldsymbol{\gamma}^P = (1.5, 1.5)$, $\boldsymbol{\gamma}^I = (1.5, 1.5)$, $Q = 4$ examiners where $(\sigma_1, \ldots, \sigma_Q) = (0.25, 0.5, 0.75, 1)$, and matrices

$$
\boldsymbol{\Omega}^P = \boldsymbol{\Omega}^I = \begin{pmatrix} 1 & 0.6247 & 0.5026 \\ 0.6247 & 1 & 0.7624 \\ 0.5026 & 0.7624 & 1 \end{pmatrix}.
$$

Figure 2 shows some traces and densities of the estimates by considering the constraint that one of the variances of $\boldsymbol{\Omega}^P$ and $\boldsymbol{\Omega}^I$ must be equal to 1.

On the contrary, Figure 3 shows some traces and densities of the estimates by considering that $\mathbf{\Omega}^P$ and $\mathbf{\Omega}^I$ are covariance matrices, i.e. without considering any constraints on the parameters. Note that the estimates on some parameters of $\mathbf{\Omega}^P$ and $\mathbf{\Omega}^I$ do not converge and the estimates of other parameters are more biased.

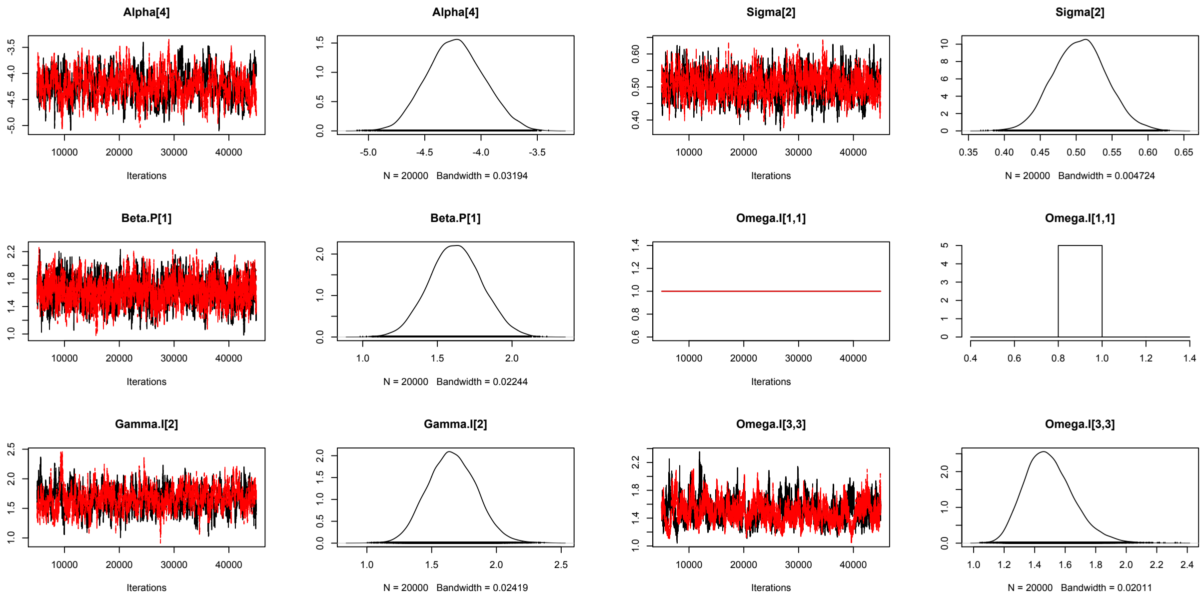

Figure 2: Traces and densities considering constraints on one variance of $\mathbf{\Omega}^P$ and $\mathbf{\Omega}^I$.

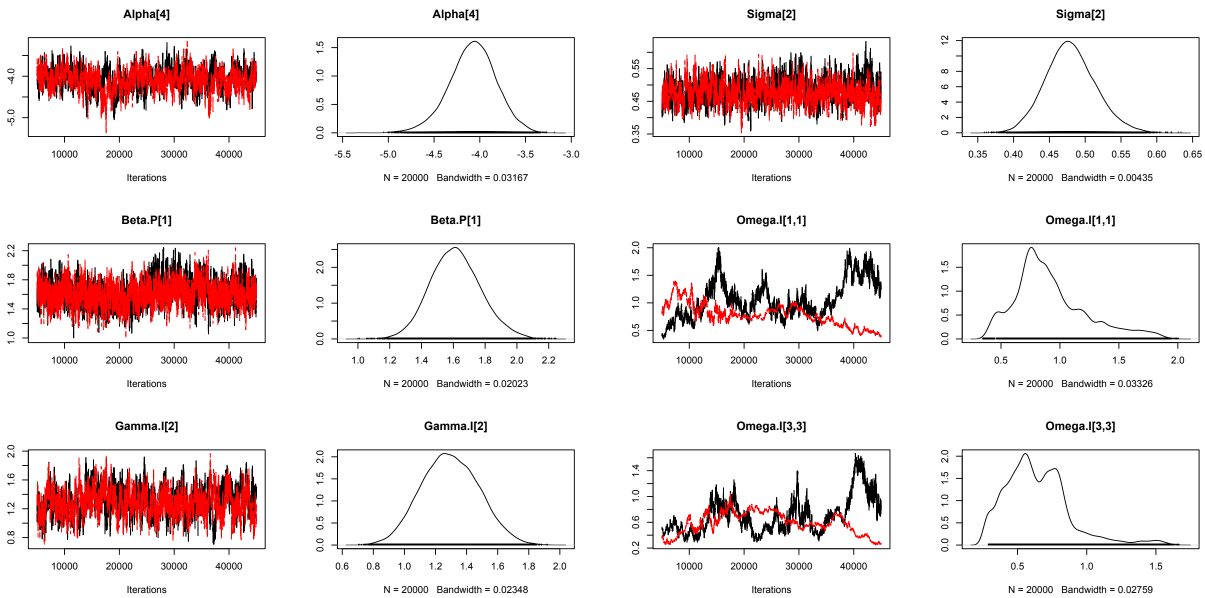

Figure 3: Traces and densities considering $\mathbf{\Omega}^P$ and $\mathbf{\Omega}^I$ are covariances, without constraints.

5 Sensitivity analysis

A sensitivity analysis for some prior distributions in the model is given here. The aim of this sensitivity analysis is to assess if the results are affected when different prior distributions are assumed for the regression coefficients $\boldsymbol{\beta}^P$, $\boldsymbol{\beta}^I$, $\boldsymbol{\gamma}^P$ and $\boldsymbol{\gamma}^I$, and intercept parameters $\boldsymbol{\alpha}$ in the linear predictor, and for the variance parameters $\sigma_1^2, \ldots, \sigma_Q^2$. The remaining variables have the same prior distributions as described in Section 4.1. Four different scenarios have been considered, using the following prior distributions for $\boldsymbol{\beta}^P$, $\boldsymbol{\beta}^I$, $\boldsymbol{\gamma}^P$, $\boldsymbol{\gamma}^I$, $\boldsymbol{\alpha}$ and $\sigma_1^2, \ldots, \sigma_Q^2$:

(A) Normal and Inverse Gamma (as in Section 6). Specifically, we have taken as prior distributions: $\beta_{l_\beta}^P \sim \mathrm{N}(0, 100)$ and $\beta_{l_\beta}^I \sim \mathrm{N}(0, 100)$ for $l_\beta = 1, 2, 3, 4$, $\gamma_{l_\gamma}^P \sim \mathrm{N}(0, 100)$ and $\gamma_{l_\gamma}^I \sim \mathrm{N}(0, 100)$ for $l_\gamma = 1, 2, 3$, $\alpha_k \sim \mathrm{N}(0, 100)$ for $k = 1, \ldots, 6$, and for the variances $\sigma_{l_\sigma}^2 \sim \mathrm{IG}(0.01, 0.01)$ for $l_\sigma = 1, \ldots, 16$,

(B) Laplace and Inverse Gamma (as in Section 6). Specifically, Laplace prior distributions were considered for the regression coefficients and for the intercept parameters in the linear predictor, i.e., let θ denote each one of the parameters for the regression coefficients and for the intercept parameters in the linear predictor, $\theta \sim \mathrm{N}(0, 100)$ and $\theta \sim \mathrm{N}\left(0, \kappa_\theta^2\right)$, where $\kappa_\theta^2 = \tau_\theta^2 \rho_\theta^2$, $\tau_\theta^2 \sim \mathrm{Exp}\left(\lambda_\theta^2/2\right)$, $\lambda_\theta^2 \sim \mathrm{Gamma}(1, 1)$, $\rho_\theta^2 \sim \mathrm{IG}(1, 1)$. And for the variances $\sigma_{l_\sigma}^2 \sim \mathrm{IG}(0.01, 0.01)$ for $l_\sigma = 1, \ldots, 16$.

(C) Normal and Uniform. Specifically, Normal prior distributions were considered for the regression coefficients in the linear predictor (as in A), and for the standard deviations $\sigma_{l_\sigma} \sim \mathrm{Uniform}(0, 100)$ for $l_\sigma = 1, \ldots, 16$.

(D) Laplace and Uniform. Specifically, Laplace prior distributions were considered for the regression coefficients in the linear predictor (as in B), and for the standard deviations $\sigma_{l_\sigma} \sim \mathrm{Uniform}(0, 100)$ for $l_\sigma = 1, \ldots, 16$.

Tables 1 and 2 show the estimated posterior results for the regression coefficients and intercept parameters in the linear predictor. Note that all prior distributions lead to similar posterior results. No relevant changes have been observed, hence the effect of the choice of prior distributions for the parameters is minimal given the fact that different prior distributions have been used. Therefore, the model can be considered as robust against changes in the prior distributions for the parameters.

Figure 4 shows the estimates for the standar deviation parameters σ_ξ associated to the examiners ξ. No relevant changes have been observed.

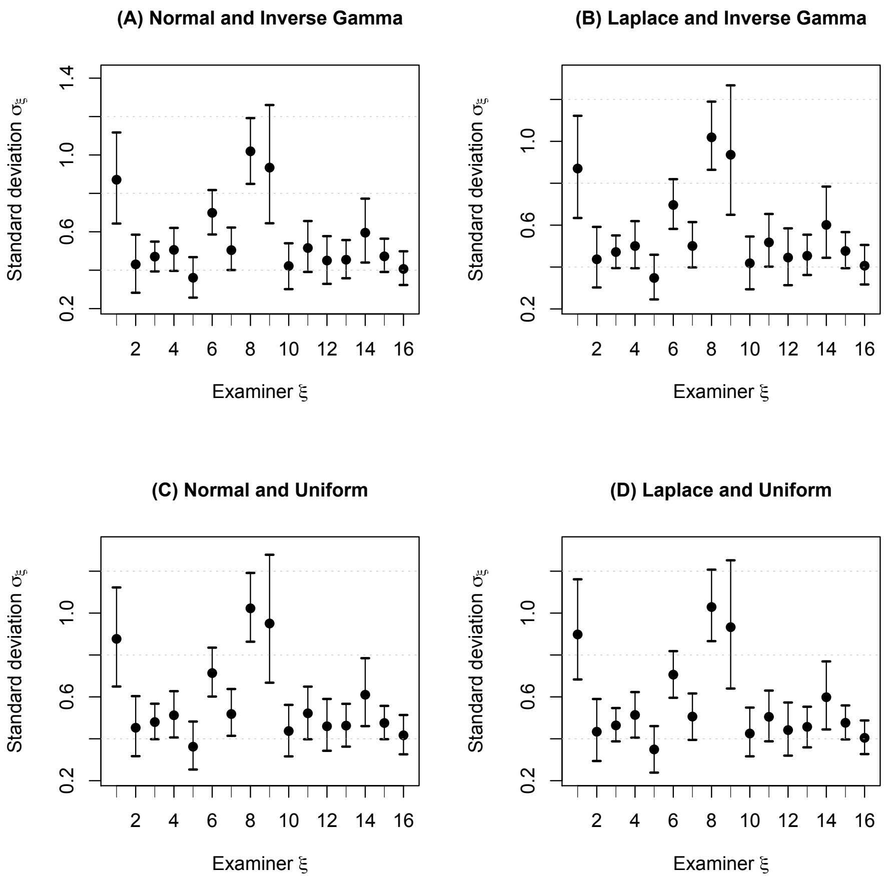

Figure 4: Estimated posterior medians, and 2.5% and 97.5% centiles for dental examiner's measurement error standard deviations σ_ξ.

Table 1: Estimated posterior means, medians, standard deviations (SD), and 2.5% and 97.5% centiles, for the regression coefficients associated to the prevalences and incidences for CE in permanent first molars.

Parameter	**Covariate**	**(A) Normal and Inverse Gamma**					**(B) Laplace and Inverse Gamma**				
		Mean	Median	SD	2.5%	97.5%	Mean	Median	SD	2.5%	97.5%
Prevalences	Gender	0.118	0.117	0.074	-0.023	0.264	0.107	0.105	0.067	-0.022	0.243
$\boldsymbol{\beta}^P$	Startbr	0.209	0.209	0.072	0.068	0.349	0.193	0.194	0.071	0.053	0.332
	x-ordinate	0.066	0.066	0.078	-0.091	0.216	0.060	0.060	0.070	-0.076	0.197
	y-ordinate	-0.174	-0.173	0.077	-0.329	-0.026	-0.129	-0.128	0.069	-0.267	0.004
Prevalences	Age	0.244	0.243	0.079	0.091	0.401	0.220	0.217	0.078	0.074	0.377
$\boldsymbol{\gamma}^P$	Meals	0.084	0.084	0.061	-0.037	0.202	0.077	0.076	0.064	-0.047	0.204
Incidences	Gender	0.145	0.144	0.082	-0.012	0.306	0.125	0.125	0.082	-0.031	0.285
$\boldsymbol{\beta}^I$	Startbr	0.204	0.204	0.080	0.052	0.359	0.166	0.166	0.080	0.011	0.326
	x-ordinate	0.257	0.255	0.084	0.094	0.426	0.223	0.223	0.096	0.032	0.421
	y-ordinate	-0.038	-0.038	0.086	-0.207	0.132	-0.045	-0.045	0.072	-0.188	0.093
Incidences	Age	-0.041	-0.039	0.081	-0.205	0.112	-0.031	-0.031	0.073	-0.176	0.112
$\boldsymbol{\gamma}^I$	Years-exam	0.214	0.216	0.091	0.031	0.387	0.178	0.180	0.086	0.013	0.345
	Meals	0.230	0.230	0.081	0.070	0.387	0.191	0.190	0.083	0.033	0.356
Intercepts	α_1	-3.394	-3.395	0.100	-3.592	-3.189	-3.341	-3.359	0.143	-3.574	-2.966
$\boldsymbol{\alpha}$	α_2	-4.026	-4.020	0.167	-4.376	-3.714	-3.979	-3.988	0.239	-4.416	-3.428
	α_3	-4.428	-4.416	0.207	-4.868	-4.063	-4.413	-4.403	0.240	-4.917	-3.964
	α_4	-6.020	-5.939	0.610	-7.447	-5.067	-5.999	-5.908	0.647	-7.569	-4.994
	α_5	-4.464	-4.442	0.298	-5.123	-3.945	-4.415	-4.400	0.309	-5.066	-3.867
	α_6	-16.597	-15.642	4.903	-28.261	-9.498	-15.568	-14.140	5.718	-30.921	-9.042

Table 2: Estimated posterior means, medians, standard deviations (SD), and 2.5% and 97.5% centiles, for the regression coefficients associated to the prevalences and incidences for CE in permanent first molars.

Parameter	**Covariate**	**(C) Normal and Uniform**					**(D) Laplace and Uniform**				
		Mean	Median	SD	2.5%	97.5%	Mean	Median	SD	2.5%	97.5%
Prevalences	Gender	0.139	0.138	0.075	-0.004	0.288	0.113	0.113	0.069	-0.022	0.248
$\boldsymbol{\beta}^P$	Startbr	0.226	0.224	0.074	0.083	0.376	0.208	0.207	0.068	0.076	0.345
	x-ordinate	0.069	0.072	0.080	-0.099	0.221	0.064	0.064	0.073	-0.077	0.206
	y-ordinate	-0.191	-0.192	0.077	-0.346	-0.039	-0.143	-0.141	0.072	-0.292	-0.001
Prevalences	Age	0.252	0.252	0.072	0.110	0.392	0.220	0.221	0.075	0.071	0.363
$\boldsymbol{\gamma}^P$	Meals	0.082	0.081	0.064	-0.040	0.210	0.082	0.081	0.061	-0.034	0.203
Incidences	Gender	0.148	0.150	0.081	-0.021	0.304	0.119	0.117	0.078	-0.027	0.276
$\boldsymbol{\beta}^I$	Startbr	0.208	0.207	0.077	0.054	0.355	0.167	0.166	0.074	0.023	0.312
	x-ordinate	0.254	0.252	0.081	0.096	0.421	0.222	0.222	0.080	0.065	0.381
	y-ordinate	-0.039	-0.041	0.075	-0.182	0.107	-0.056	-0.052	0.082	-0.231	0.104
Incidences	Age	-0.046	-0.047	0.079	-0.205	0.104	-0.041	-0.039	0.078	-0.199	0.109
$\boldsymbol{\gamma}^I$	Years-exam	0.220	0.221	0.087	0.049	0.393	0.165	0.164	0.088	-0.009	0.338
	Meals	0.245	0.247	0.079	0.088	0.393	0.237	0.237	0.078	0.077	0.389
Intercepts	α_1	-3.440	-3.438	0.087	-3.618	-3.276	-3.394	-3.396	0.087	-3.563	-3.218
$\boldsymbol{\alpha}$	α_2	-4.083	-4.081	0.156	-4.392	-3.793	-4.024	-4.014	0.156	-4.350	-3.745
	α_3	-4.416	-4.407	0.214	-4.863	-4.023	-4.404	-4.390	0.227	-4.907	-3.980
	α_4	-6.061	-5.965	0.633	-7.613	-5.098	-6.064	-5.962	0.699	-7.779	-4.988
	α_5	-4.414	-4.389	0.273	-4.997	-3.923	-4.553	-4.550	0.288	-5.168	-4.024
	α_6	-16.087	-15.361	4.190	-25.436	-9.836	-14.223	-12.868	4.934	-27.024	-8.417

References

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85** (2), 347–361.

Curtis, S. M. (2010). BUGS code for item response theory. *Journal of Statistical Software, Code Snippet*, **36**(1), 1–34.

García-Zattera, M. J., Jara, A., Lesaffre, E., and Declerck, D. (2007). Conditional independence of multivariate binary data with an application in caries research. *Computational Statistics and Data Analysis*, **51**, 3223–3234.

García-Zattera, M. J., Jara, A., Lesaffre, E., and Marshall, G. (2012). Modeling of multivariate monotone disease processes in presence of misclassification. *Journal of the American Statistical Association*, **107**(499), 976–989.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* John Wiley, Chichester.