Supplementary Material for
"A Bayesian Model for Co-clustering Ordinal Data with Informative Missing Entries"

Alice Giampino*[1], Antonio Canale†[2], and Bernardo Nipoti‡[1]

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, Milan, 20126, Italy.
[2]Department of Statistics, University of Padova, Via C. Battisti, 241, Padova, 35121, Italy.

This file is organized as follows. In Section A we report additional information on the full conditional distributions involved in the Gibbs sampling algorithm summarized in Algorithm 1. Section C and D provide additional results on the simulation study of Section 4 and the data analyses of Section 5, respectively,

Appendix A Full conditional distributions

We provide additional details on the full conditional distributions already discussed in Section 3. To this end, introducing additional notation is going to be useful. For $r = 1, 2$, we denote the r-th column of $\boldsymbol{M}_l$ as $\boldsymbol{m}_{lr} = (m_{lr1}, \ldots, m_{lrd})$; moreover, we let $v_{ls_1s_2}$ denote the element in position (s_1, s_2) of matrix $\boldsymbol{V}_l$ and $u_{lr_1r_2}$ denote the element in position (r_1, r_2) of matrix $\boldsymbol{U}_l$.

The weights in (7) are given, up to a proportionality constant, by

$$\pi_{1i0} \propto \frac{\alpha_1}{\alpha_1 + n - 1} |\tilde{\boldsymbol{V}}_1|^{-\frac{1}{2}} \frac{1}{\sqrt{u_{122}}} |\boldsymbol{V}_1|^{-\frac{1}{2}} |(\frac{1}{\sigma_1^2} \boldsymbol{\theta}_2^{(1)\intercal} \boldsymbol{\theta}_2^{(1)} + \tilde{\boldsymbol{V}}_1^{-1})|^{-\frac{1}{2}}$$
$$\times |(\frac{1}{\sigma_2^2} \boldsymbol{\theta}_2^{(2)\intercal} \boldsymbol{\theta}_2^{(2)} + \frac{1}{u_{122}} \boldsymbol{V}_1^{-1})|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left[\frac{1}{\sigma_1^2} \boldsymbol{z}_i \boldsymbol{z}_i^\intercal + \frac{1}{\sigma_2^2} \boldsymbol{w}_i \boldsymbol{w}_i^\intercal \right] \right\}$$

*alice.giampino@unimib.it
†antonio.canale@unipd.it
‡bernardo.nipoti@unimib.it

$$
\begin{aligned}
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\tilde{\boldsymbol{V}}_1^{-1}\tilde{\boldsymbol{m}}_{11}\tilde{\boldsymbol{m}}_{11}^{\intercal} + \frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\boldsymbol{m}_{12}\right]\right\}\\
&\times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{z}_i + \tilde{\boldsymbol{V}}_1^{-1}\tilde{\boldsymbol{m}}_{11}\right)\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{z}_i + \tilde{\boldsymbol{V}}_1^{-1}\tilde{\boldsymbol{m}}_{11}\right)^{\intercal}\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{\theta}_2^{(1)} + \tilde{\boldsymbol{V}}_1^{-1}\right)^{-1}\right]\right\}\\
&\times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{w}_i + \frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\right)\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{w}_i + \frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\right)^{\intercal}\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{\theta}_2^{(2)} + \frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\right)^{-1}\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
\pi_{2i0} \propto\ &\frac{\alpha_2}{\alpha_2+p-1}|\tilde{\boldsymbol{V}}_2|^{-\frac{1}{2}}\frac{1}{\sqrt{u_{222}}}|\boldsymbol{V}_2|^{-\frac{1}{2}}|(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{\theta}_1^{(1)} + \tilde{\boldsymbol{V}}_2^{-1})|^{-\frac{1}{2}}|(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{\theta}_1^{(2)} + \frac{1}{u_{222}}\boldsymbol{V}_2^{-1})|^{-\frac{1}{2}}\\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\boldsymbol{z}_j\boldsymbol{z}_j^{\intercal} + \frac{1}{\sigma_2^2}\boldsymbol{w}_j\boldsymbol{w}_j^{\intercal}\right]\right\}\\
&\times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\tilde{\boldsymbol{V}}_2\tilde{\boldsymbol{m}}_{21}\tilde{\boldsymbol{m}}_{21}^{\intercal} + \frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\boldsymbol{m}_{22}\right]\right\}\\
&\times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{z}_j + \tilde{\boldsymbol{V}}_2^{-1}\tilde{\boldsymbol{m}}_{21}\right)\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{z}_j + \tilde{\boldsymbol{V}}_2^{-1}\tilde{\boldsymbol{m}}_{21}\right)^{\intercal}\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{\theta}_1^{(1)} + \tilde{\boldsymbol{V}}_2^{-1}\right)^{-1}\right]\right\}\\
&\times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{w}_j + \frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\right)\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{w}_j + \frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\right)^{\intercal}\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{\theta}_1^{(2)} + \frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\right)^{-1}\right]\right\}.
\end{aligned}
$$

Sampling from G_{1i} in (7) is straightforward after observing that G_{1i} is the distribution of a $(d \times 2)$-dimensional matrix $\boldsymbol{T}$, with columns $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, for which

$$
\begin{aligned}
\boldsymbol{t}_1 \mid \boldsymbol{t}_2 &\sim \mathrm{N}_d\left(\tilde{\boldsymbol{m}}_{11}, \tilde{\boldsymbol{V}}_1\right)\\
\boldsymbol{t}_2 &\sim \mathrm{N}_d\left(\boldsymbol{m}_{12}, u_{122}\boldsymbol{V}_1\right).
\end{aligned}
$$

Similarly, the distribution G_{2j} in (8) coincides with the distributions of a $(d \times 2)$-dimensional matrix $\boldsymbol{T}$, with columns $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$, for which

$$
\begin{aligned}
\boldsymbol{t}_1 \mid \boldsymbol{t}_2 &\sim \mathrm{N}_d\left(\tilde{\boldsymbol{m}}_{21}, \tilde{\boldsymbol{V}}_2\right)\\
\boldsymbol{t}_2 &\sim \mathrm{N}_d\left(\boldsymbol{m}_{22}, u_{222}\boldsymbol{V}_2\right).
\end{aligned}
$$

Next, we show how the weights in (7) and (8) simplify when the base measures H_1 and H_2 have independent components, which is achieved, respectively, when $u_{112} = u_{121} = 0$ and $u_{212} = u_{221} = 0$.

$$
\begin{aligned}
\pi_{1i0} \propto & \frac{\alpha_1}{\alpha_1+n-1}\frac{1}{\sqrt{u_{111}}}|\boldsymbol{V}_1|^{-\frac{1}{2}}\frac{1}{\sqrt{u_{122}}}|\boldsymbol{V}_1|^{-\frac{1}{2}}|(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{\theta}_2^{(1)}+\frac{1}{u_{111}}\boldsymbol{V}_1^{-1})|^{-\frac{1}{2}}|(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{\theta}_2^{(2)}+\frac{1}{u_{122}}\boldsymbol{V}_1^{-1})|^{-\frac{1}{2}} \\
& \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\boldsymbol{z}_i\boldsymbol{z}_i^\intercal+\frac{1}{\sigma_2^2}\boldsymbol{w}_i\boldsymbol{w}_i^\intercal\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\frac{1}{u_{111}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{11}\boldsymbol{m}_{11}^\intercal+\frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\boldsymbol{m}_{12}\right]\right\} \\
& \times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{z}_i+\frac{1}{u_{111}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{11}\right)\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{z}_i+\frac{1}{u_{111}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{11}\right)^\intercal\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)\intercal}\boldsymbol{\theta}_2^{(1)}+\frac{1}{u_{111}}\boldsymbol{V}_1^{-1}\right)^{-1}\right]\right\} \\
& \times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{w}_i+\frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\right)\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{w}_i+\frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12}\right)^\intercal\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)\intercal}\boldsymbol{\theta}_2^{(2)}+\frac{1}{u_{122}}\boldsymbol{V}_1^{-1}\right)^{-1}\right]\right\} \\
\pi_{2i0} \propto & \frac{\alpha_2}{\alpha_2+p-1}\frac{1}{\sqrt{u_{211}}}|\boldsymbol{V}_2|^{-\frac{1}{2}}\frac{1}{\sqrt{u_{222}}}|\boldsymbol{V}_2|^{-\frac{1}{2}}|(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{\theta}_1^{(1)}+\frac{1}{u_{211}}\boldsymbol{V}_2^{-1})|^{-\frac{1}{2}}|(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{\theta}_1^{(2)}+\frac{1}{u_{222}}\boldsymbol{V}_2^{-1})|^{-\frac{1}{2}} \\
& \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\boldsymbol{z}_j\boldsymbol{z}_j^\intercal+\frac{1}{\sigma_2^2}\boldsymbol{w}_j\boldsymbol{w}_j^\intercal\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\frac{1}{u_{211}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{21}\boldsymbol{m}_{21}^\intercal+\frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\boldsymbol{m}_{22}\right]\right\} \\
& \times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{z}_j+\frac{1}{u_{211}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{21}\right)\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{z}_j+\frac{1}{u_{211}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{21}\right)^\intercal\left(\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)\intercal}\boldsymbol{\theta}_1^{(1)}+\frac{1}{u_{211}}\boldsymbol{V}_2^{-1}\right)^{-1}\right]\right\} \\
& \times \exp\left\{+\frac{1}{2}\mathrm{tr}\left[\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{w}_j+\frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\right)\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{w}_j+\frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22}\right)^\intercal\left(\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)\intercal}\boldsymbol{\theta}_1^{(2)}+\frac{1}{u_{222}}\boldsymbol{V}_2^{-1}\right)^{-1}\right]\right\}.
\end{aligned}
$$

Also sampling from G_{1i} and G_{2j} simplifies. G_{1i} is the distribution of a $(d \times 2)$-dimensional matrix $\boldsymbol{T}$, with independent columns $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ such that

$$
\begin{aligned}
\boldsymbol{t}_1 &\sim \mathrm{N}_d\left(\boldsymbol{m}_{11}, u_{111}\boldsymbol{V}_1\right) \\
\boldsymbol{t}_2 &\sim \mathrm{N}_d\left(\boldsymbol{m}_{12}, u_{122}\boldsymbol{V}_1\right).
\end{aligned}
$$

Similarly, for $\boldsymbol{T} \sim G_{2j}$ we have

$$
\boldsymbol{t}_1 \sim \mathrm{N}_d\left(\boldsymbol{m}_{21}, u_{211}\boldsymbol{V}_2\right)
$$

$$\boldsymbol{t}_2 \sim \mathrm{N}_d\left(\boldsymbol{m}_{22}, u_{222}\boldsymbol{V}_2\right).$$

We conclude this section by discussing how to implement the reshuffling step in (9). For any $\ell_1 = 1, \ldots, k_n$, we let $\bar{\boldsymbol{z}}_{\ell_1}$ and $\bar{\boldsymbol{w}}_{\ell_1}$ be, respectively, the component-wise averages of the vectors $\boldsymbol{z}_i$ and $\boldsymbol{w}_i$ for $i \in C_{\ell_1}$. By considering again the simplifying assumption $u_{112} = u_{121} = 0$, we obtain

$$\boldsymbol{\theta}_{1\ell_1}^{(1)*} \mid \ldots \overset{\text{ind}}{\sim} \mathrm{N}_d\Big((u_{111}^{-1}\boldsymbol{V}_1^{-1} + n_{\ell_1}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)}\boldsymbol{\theta}_2^{(1)\intercal})^{-1}(u_{111}^{-1}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{11} + n_{\ell_1}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)}\bar{\boldsymbol{z}}_{\ell_1}), (u_{111}^{-1}\boldsymbol{V}_1^{-1} + n_{\ell_1}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_2^{(1)}\boldsymbol{\theta}_2^{(1)\intercal})^{-1}\Big)$$
$$\boldsymbol{\theta}_{1\ell_1}^{(2)*} \mid \ldots \overset{\text{ind}}{\sim} \mathrm{N}_d\Big((u_{122}^{-1}\boldsymbol{V}_1^{-1} + n_{\ell_1}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)}\boldsymbol{\theta}_2^{(2)\intercal})^{-1}(u_{122}^{-1}\boldsymbol{V}_1^{-1}\boldsymbol{m}_{12} + n_{\ell_1}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)}\bar{\boldsymbol{w}}_{\ell_1}), (u_{122}^{-1}\boldsymbol{V}_1^{-1} + n_{\ell_1}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_2^{(2)}\boldsymbol{\theta}_2^{(2)\intercal})^{-1}\Big)$$

Similarly, for any $\ell_2 = 1, \ldots, k_p$, we let $\bar{\boldsymbol{z}}_{\ell_2}$ and $\bar{\boldsymbol{w}}_{\ell_2}$ be, respectively, the component-wise averages of the vectors $\boldsymbol{z}_j$ and $\boldsymbol{w}_j$ for $j \in C_{\ell_2}$. By considering the simplifying assumption $u_{212} = u_{221} = 0$, we obtain

$$\boldsymbol{\theta}_{2\ell_2}^{(1)*} \mid \ldots \overset{\text{ind}}{\sim} \mathrm{N}_d\Big((u_{211}^{-1}\boldsymbol{V}_2^{-1} + n_{\ell_2}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)}\boldsymbol{\theta}_1^{(1)\intercal})^{-1}(u_{211}^{-1}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{21} + n_{\ell_2}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)}\bar{\boldsymbol{z}}_{\ell_2}), (u_{211}^{-1}\boldsymbol{V}_2^{-1} + n_{\ell_2}\frac{1}{\sigma_1^2}\boldsymbol{\theta}_1^{(1)}\boldsymbol{\theta}_1^{(1)\intercal})^{-1}\Big)$$
$$\boldsymbol{\theta}_{2\ell_2}^{(2)*} \mid \ldots \overset{\text{ind}}{\sim} \mathrm{N}_d\Big((u_{222}^{-1}\boldsymbol{V}_2^{-1} + n_{\ell_2}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)}\boldsymbol{\theta}_1^{(2)\intercal})^{-1}(u_{222}^{-1}\boldsymbol{V}_2^{-1}\boldsymbol{m}_{22} + n_{\ell_2}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)}\bar{\boldsymbol{w}}_{\ell_2}), (u_{222}^{-1}\boldsymbol{V}_2^{-1} + n_{\ell_2}\frac{1}{\sigma_2^2}\boldsymbol{\theta}_1^{(2)}\boldsymbol{\theta}_1^{(2)\intercal})^{-1}\Big)$$

Appendix B Log Pseudo Marginal Likelihood (LPML)

In this section, we provide additional details on the computation of the Log Pseudo Marginal Likelihood (LPML) (Geisser and Eddy, 1979), which is the measure of fit used to determine the latent dimension d throughout the paper. The LPML is obtained by computing the logarithm of the pseudo-marginal likelihood, which is calculated as the product over all observations of the conditional predictive ordinates. Gelfand and Dey (1994) conveniently show how to compute such cross-validated measure from the output of a single run of a full-sample MCMC. Specifically, if B indicates the total number of the MCMC iterations, the LPML for a model with latent dimension d is computed as

$$\mathrm{LPML}_d = \log L_d = \sum_{i=1}^{n}\sum_{j=1}^{p} \log \mathrm{CPO}_{ijd},$$

where

$$\mathrm{CPO}_{ijd}^{-1} = \frac{1}{B}\sum_{s=1}^{B}\frac{1}{f_d(y_{ij}, \delta_{ij}, \mid \boldsymbol{\theta}_{1i}, \boldsymbol{\theta}_{2j})},$$

with $f_d(\cdot)$ representing the probability mass function of (y_{ij}, δ_{ij}) for a given value of d. Specifically,

$$f_d(y_{ij}, \delta_{ij} \mid \boldsymbol{\theta}_{1i}, \boldsymbol{\theta}_{2j}) \;=\left[\Phi\left(\frac{\gamma_{ij}^{(u)} - \boldsymbol{\theta}_{1i}^{(1)\intercal}\boldsymbol{\theta}_{2j}^{(1)}}{\sigma_1}\right) - \Phi\left(\frac{\gamma_{ij}^{(l)} - \boldsymbol{\theta}_{1i}^{(1)\intercal}\boldsymbol{\theta}_{2j}^{(1)}}{\sigma_1}\right)\right] + \left[\Phi\left(\frac{\eta_{ij}^{(u)} - \boldsymbol{\theta}_{1i}^{(2)\intercal}\boldsymbol{\theta}_{2j}^{(2)}}{\sigma_2}\right) - \Phi\left(\frac{\eta_{ij}^{(l)} - \boldsymbol{\theta}_{1i}^{(2)\intercal}\boldsymbol{\theta}_{2j}^{(2)}}{\sigma_2}\right)\right],$$

where the values $(\gamma_{ij}^{(l)}, \gamma_{ij}^{(u)})$ and $(\eta_{ij}^{(l)}, \eta_{ij}^{(u)})$ are determined on the basis of the observed pair (y_{ij}, δ_{ij}) as follows:

- if $\delta_{ij} = 0$, then $\gamma_{ij}^{(l)} = \gamma_0$, $\gamma_{ij}^{(u)} = \gamma_c$, $\eta_{ij}^{(l)} = -\infty$ and $\eta_{ij}^{(u)} = 0$;
- if $\delta_{ij} = 1$ and $y_{ij} = \kappa$, with $\kappa = 1, \ldots, c$, then $\gamma_{ij}^{(l)} = \gamma_{\kappa-1}$, $\gamma_{ij}^{(u)} = \gamma_\kappa$, $\eta_{ij}^{(l)} = 0$ and $\eta_{ij}^{(u)} = \infty$.

The notation $\Phi(\pm\infty)$ is to be intended as $\lim_{x\to\pm\infty}\Phi(x)$.

Appendix C Additional details on the simulation study

We provide additional details on the simulation study of Section 4.

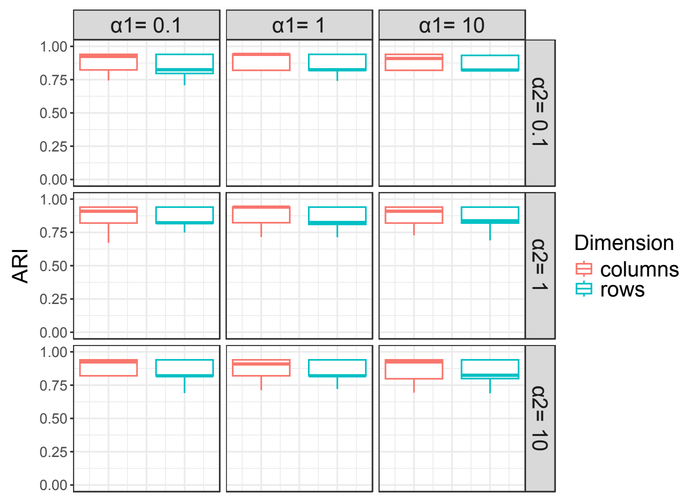

Figure C1: Sensitivity analysis. Boxplots of the ARI comparing true and inferred partitions of rows and columns, with both α_1 and α_2 ranging in $\{0.1, 1, 10\}$. Results are based on the analysis of 50 datasets of dimension 100×100 simulated from the same scenario with ordinal data considered in Section 4. Refer to Section 4 for further details.

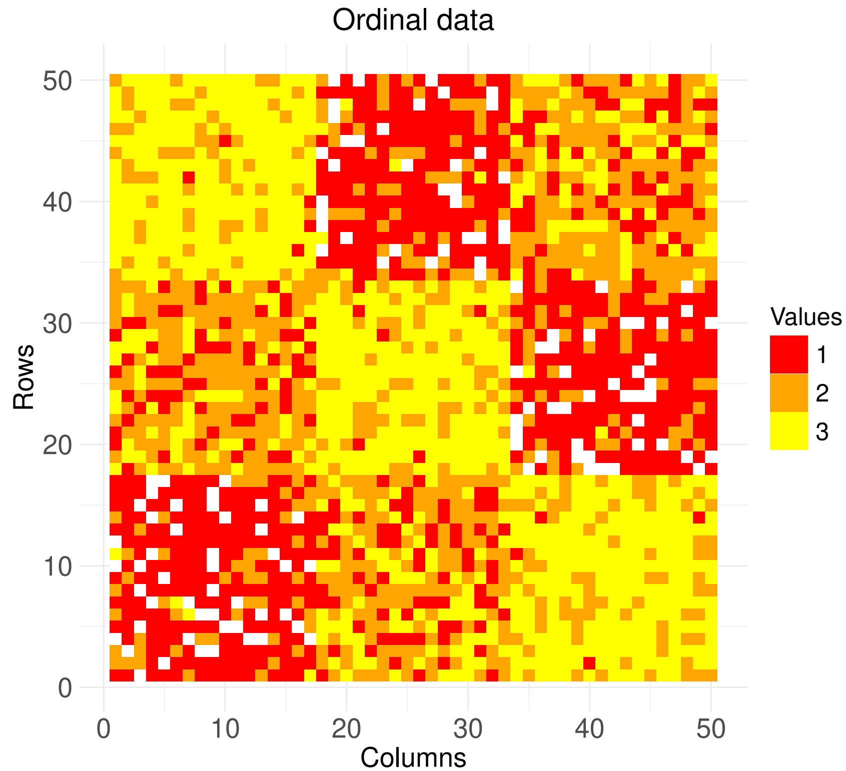

Figure C2: Graphical representation of a simulated dataset with ordinal observations. White pixels (5% of the total) indicate missing entries generated according to an informative censoring mechanism. Refer to Section 4 for further details.

Table C1: Simulated data with ordinal observations. Estimated ARI for the marginal partitions of the rows and columns of $\boldsymbol{Y}$, along with the estimated number of clusters, $\hat{k}_n$ and $\hat{k}_p$, for varying dataset sizes. Values in parentheses represent the estimated standard error. Refer to Section 4 for further details.

	Rows		Columns	
$n \times p$	ARI	$\hat{k}_n$	ARI	$\hat{k}_p$
50×50	0.880 (0.059)	2.006 (0.826)	0.932 (0.013)	2.451 (0.973)
100×100	0.895 (0.065)	3.257 (1.490)	0.961 (0.008)	2.556 (0.998)
200×200	0.957 (0.019)	3.067 (1.609)	0.970 (0.001)	3.809 (1.103)

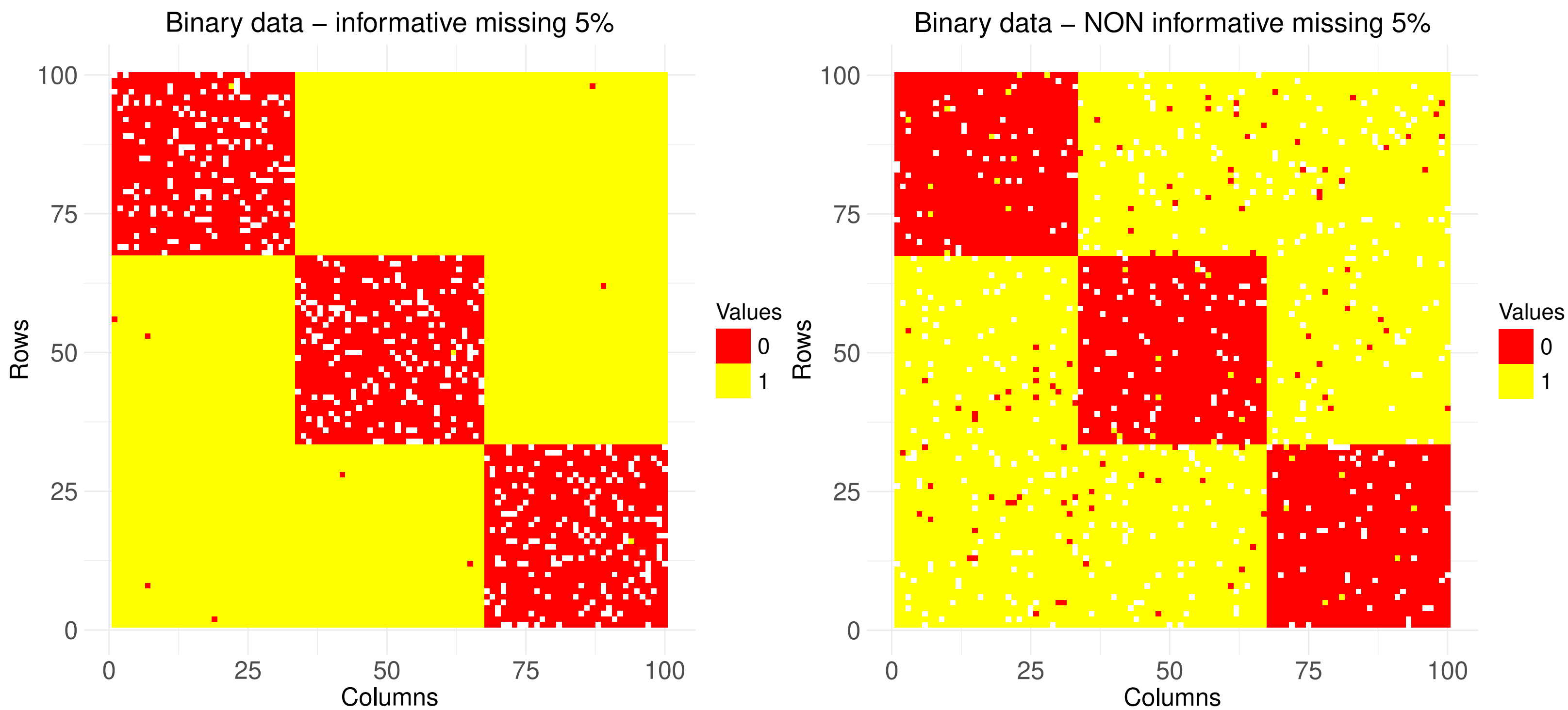

Figure C3: Graphical representation of a simulated dataset with binary observations. White pixels (5% of the total) indicate missing entries generated according to an informative censoring mechanism (left panel) or to a non-informative censoring mechanism (right panel). Refer to Section 4 for further details.

Appendix D Additional details on the real data analyses

In this section, we provide additional details of the analysis of datasets we presented in 5.

The code implementing CO^3 , along with a script to replicate the real-data analysis, is available at `https://github.com/AliceGiampino/CO3/`.

D.1 U.S. Senate data

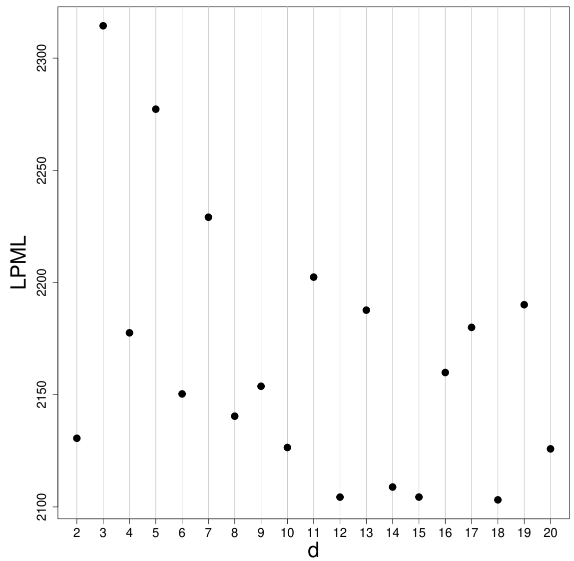

Figure D4: U.S. Senators Data. LPML as a function of the latent dimension d. Refer to Section 5.1 for further details.

Name	Party
BLUMENTHAL, Richard	Democrat
DURBIN, Richard Joseph	Democrat
FEINSTEIN, Dianne	Democrat
FISCHER, Debra (Deb)	Republican
JOHNSON, Ron	Republican
LEE, Mike	Republican
MARKEY, Edward John	Democrat
MORAN, Jerry	Republican
PAUL, Rand	Republican
RUBIO, Marco	Republican
SCHATZ, Brian Emanuel	Democrat
SCOTT, Tim	Republican
SHELBY, Richard C.	Republican
WICKER, Roger F.	Republican

Table D2: U.S. Senators Data. Names and party affiliations of the senators assigned to cluster 3 in the bottom plot of Figure 5.

D.2 Movielens data

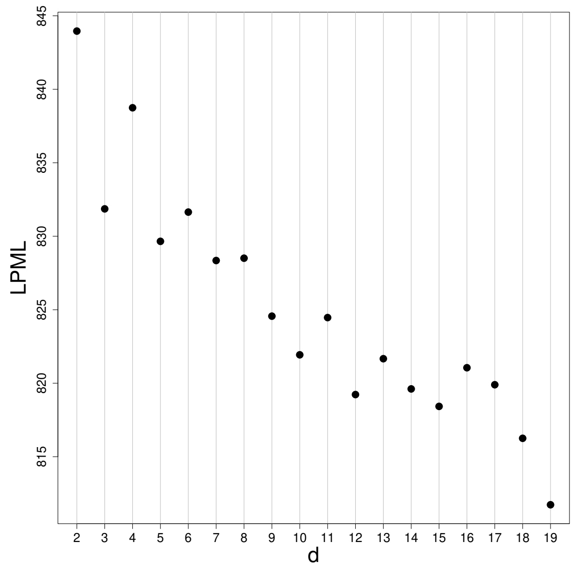

Figure D5: Movielens Data. LPML as a function of the latent dimension d. Refer to Section 5.2 for further details.

References

Geisser, S. and Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160.

Gelfand, A. E. and Dey, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):501–514.