Supplementary Material for "Colombian Women's Life Patterns: A Multivariate Density Regression Approach"

Sara Wade*, Raffaella Piccarreta†, Andrea Cremaschi‡, and Isadora Antoniano-Villalobos§

In this Supplementary Material, we describe how to compute predictive quantities of interest from the MCMC output in Section A, and in Sections B and C, we report additional results for the simulated data example and the application to the Colombian women dataset described in the paper.

Appendix A: Predictions

The weighted posterior samples obtained with the adaptive truncation algorithm can be used to produce various posterior and predictive quantities of interest. Here, we describe how to compute the predictive densities, medians, *probability of censoring* for the (undiscretized) ages at event and probability of success for binary variables. Full details and implementation for other quantities are provided in the accompanying software and documentation: `https://github.com/sarawade/BNPDensityRegression_AdaptiveTruncation`.

Focusing on the application in Section 6, for $\ell = 1, 2, 3$, we denote by $\tilde{Z}_\ell$ the (undiscretized) age at sexual debut, the (undiscretized) age at union, and the time from sexual debut to first child, respectively. These are linked to our model by the relation $\tilde{Z}_\ell = \exp(Y_\ell)$, and the corresponding ages are obtained through discretization. The (undiscretized) age at first child is denoted as $\check{Z}_3 = \tilde{Z}_1 + \tilde{Z}_3$. For *Work Status*, we have $Z_4 = \mathbb{1}_{(0,\infty)}(Y_4)$. In the following, let J denote the final truncation level, with corresponding weighted particles $(\mathbf{w}^m_{1:J}, \boldsymbol{\theta}^m_{1:J}, \boldsymbol{\psi}^m_{1:J}, \mathbf{y}^m_{1:n})$ and unnormalized particle weights $\tilde{\vartheta}^m$, for $m = 1, \ldots, M$ (without loss of generality, we drop the subscript J). We indicate with ϑ^m, for $m = 1, \ldots, M$, the normalized particle weights.

We begin with marginal predictive quantities of interest. First, the predictive probability of success for a binary response given $\mathbf{x}_*$ (e.g. for $\ell = 4$, shown in Figure 5, bottom row) is:

$$\mathbb{P}(Z_{*,\ell} = 1 | \mathbf{x}, \mathbf{z}, \mathbf{x}_*) = \mathbb{P}(Y_{*,\ell} > 0 | \mathbf{x}, \mathbf{z}, \mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \Phi\left(\frac{\mathbf{x}_* \boldsymbol{\beta}^m_{j,(\cdot,\ell)}}{\sqrt{\boldsymbol{\Sigma}^m_{j,(\ell,\ell)}}} \right).$$

*School of Mathematics, University of Edinburgh, Edinburgh , UK sara.wade@ed.ac.uk

†Department of Decision Sciences, Dondena Research Centre, Bocconi Institute for Data Science and Analytics (BIDSA), Bocconi University, Milan, Italy raffaella.piccarreta@unibocconi.it

‡Yale-NUS College, Singapore andrea.cremaschi@yale-nus.edu.sg

§Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice and BIDSA, Italy isadora.antoniano@unive.it

For $\ell = 1, 2, 3$, the marginal predictive density of $\tilde{Z}_{*,\ell}$ given $\mathbf{x}_*$, shown in Figure 6 for some values of $\mathbf{x}_*$, is given by:

$$\begin{aligned} f(\tilde{z}_{*,\ell}|\mathbf{x}, \mathbf{z}, \mathbf{x}_*) &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) f(\tilde{z}_{*,\ell}|\boldsymbol{\theta}_j^m, \mathbf{x}_*) \\ &= \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \text{logN}(\tilde{z}_{*,\ell}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m, \boldsymbol{\Sigma}_{j,(\ell,\ell)}^m), \end{aligned} \tag{1}$$

for $\tilde{z}_{*,\ell} > 0$, where $\boldsymbol{\beta}_{j,(\cdot,\ell)}^m$ denotes the ℓ-th column of $\boldsymbol{\beta}$ in component j and particle m; $\boldsymbol{\Sigma}_{j,(\ell,\ell)}^m$ denotes element (ℓ,ℓ) of the matrix $\boldsymbol{\Sigma}$ in component j and particle m; and $\text{logN}(\cdot|\mu, \sigma^2)$ denotes the log-normal density with parameters μ and σ^2. Due to the skewness of the predictive densities in our application (Section 6), we focus on the predictive median over the predictive mean to better represent the central tendencies and summarize the predictive densities. The marginal predictive median (Figure 5) can be computed numerically by evaluating the marginal predictive density (1) on a sufficiently dense grid of $\tilde{z}_{*,\ell}$ values.

For $\ell = 1, 2$ corresponding to age at sexual debut and union, an interesting quantity is the predictive probability that the indexed event has not yet occurred for a new individual with $x_{*,1}$ years of age (Figure C.8), computed as:

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{*,\ell} \geq (x_{*,1}+1)|\mathbf{x}, \mathbf{z}, \mathbf{x}_*) &= \mathbb{P}(Y_{*,\ell} > \log(x_{*,1}+1)|\mathbf{x}, \mathbf{z}, \mathbf{x}_*) \\ &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \left(1 - \Phi\left(\frac{\log(x_{*,1}+1) - \mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m}{\sqrt{\boldsymbol{\Sigma}_{j,(\ell,\ell)}^m}}\right)\right). \end{aligned} \tag{2}$$

This can be interpreted as the *predictive probability of censoring* of the event for a new individual and corresponds to the mass above the dashed line of Figure 6, given $x_{*,1}$.

Our model also recovers the joint relationship between responses, which allows inference on conditional properties. Specifically, when ℓ indexes a binary response and ℓ' indexes an age at event response, the conditional predictive probability of success given $\tilde{z}_{*,\ell'}$ and $\mathbf{x}_*$ (Figure 8) is:

$$\mathbb{P}(Z_{*,\ell} = 1|\tilde{z}_{*,\ell'}, \mathbf{x}, \mathbf{z}, \mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \Phi\left(\frac{\mu_{j,\ell|\ell'}^m}{\sqrt{\sigma_{j,\ell|\ell'}^{2\,m}}}\right) \frac{\text{logN}(\tilde{z}_{*,\ell'}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell')}^m, \boldsymbol{\Sigma}_{j,(\ell',\ell')}^m)}{f(\tilde{z}_{*,\ell'}|\mathbf{x}, \mathbf{z}, \mathbf{x}_*)},$$

where

$$\begin{aligned} \mu_{j,\ell|\ell'}^m &= \mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell)}^m + \boldsymbol{\Sigma}_{j,(\ell,\ell')}^m (\boldsymbol{\Sigma}_{j,(\ell',\ell')}^m)^{-1} (\log(\tilde{z}_{*,\ell'}) - \mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell')}^m), \\ \sigma_{j,\ell|\ell'}^{2\,m} &= \boldsymbol{\Sigma}_{j,(\ell,\ell)}^m - (\boldsymbol{\Sigma}_{j,(\ell,\ell')}^m)^2 (\boldsymbol{\Sigma}_{j,(\ell',\ell')}^m)^{-1}, \end{aligned}$$

and the density in the denominator is the marginal predictive of equation (1). For $\ell \neq \ell'$ both indexing ages at event, the conditional predictive density of $\tilde{Z}_{*,\ell}$ given $\tilde{z}_{*,\ell'}$ and $\mathbf{x}_*$ takes the form:

$$f(\tilde{z}_{*,\ell}|\tilde{z}_{*,\ell'}, \mathbf{x}, \mathbf{z}, \mathbf{x}_*) =$$

$$\sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \text{logN}(\tilde{z}_{*,\ell}|\mu_{j,\ell|\ell'}^m, \sigma_{j,\ell|\ell'}^{2\,m}) \frac{\text{logN}(\tilde{z}_{*,\ell'}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,\ell')}^m, \boldsymbol{\Sigma}_{j,(\ell',\ell')}^m)}{f(\tilde{z}_{*,\ell'}|\mathbf{x},\mathbf{z},\mathbf{x}_*)}. \quad (3)$$

Figure C.11 shows the conditional predictive density of $\tilde{Z}_{*,2} - \tilde{z}_{*,1}$ given $\tilde{z}_{*,1}$ and $\mathbf{x}_*$, which can be easily computed from (3). The corresponding predictive medians (Figure 7) can be obtained numerically from evaluations of this density on an adequate, dense grid of values. The conditional density plot for $\tilde{Z}_{*,3}$ given $\tilde{z}_{*,1}$ (Figure C.12) and the corresponding median (Figure C.9) can be obtained directly from equation (3).

Finally, we note that for the (undiscretized) age at first child, $\breve{Z}_{*,3} = \tilde{Z}_{*,1} + \tilde{Z}_{*,3}$, and more generally constrained responses, the corresponding marginal and conditional predictive quantities may require integration over $\tilde{Z}_{*,1}$. For example, the conditional predictive density of $\breve{Z}_{*,3}$ given $\tilde{z}_{*,1}$ is simply the conditional predictive density of equation (3), evaluated at $\breve{z}_{*,3} - \tilde{z}_{*,1}$. While the marginal predictive density of $\breve{Z}_{*,3}$ given $\mathbf{x}_*$ (Figure 6) is obtained as:

$$\begin{aligned} f(\breve{z}_{*,3}|\mathbf{x},\mathbf{z},\mathbf{x}_*) &= \int f(\tilde{z}_{*,3}|\tilde{z}_{*,1},\mathbf{x},\mathbf{z},\mathbf{x}_*) f(\tilde{z}_{*,1}|\mathbf{x},\mathbf{z},\mathbf{x}_*) d\tilde{z}_{*,1} \\ &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \int_{-\infty}^{\breve{z}_{*,3}} \text{logN}(\tilde{z}_{*,3}|\mu_{j,3|1}^m, \sigma_{j,3|1}^{2\,m}) \text{logN}(\tilde{z}_{*,1}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,1)}^m, \boldsymbol{\Sigma}_{j,(1,1)}^m) d\tilde{z}_{*,1}, \end{aligned} \quad (4)$$

where $\tilde{z}_{*,3} = \breve{z}_{*,3} - \tilde{z}_{*,1}$. We evaluate the integral stochastically, via a Monte Carlo approximation, and compute the marginal predictive median of the undiscretized age at first child (Figure 5) numerically from the marginal predictive density in (4). Also, the predictive probability that the woman has not yet had a child at $x_{*,1}$ years of age (Figure C.8) takes the form:

$$\begin{aligned} \mathbb{P}(\breve{Z}_{*,3} > x_{*,1}|\mathbf{x},\mathbf{z},\mathbf{x}_*) &= \mathbb{P}(\tilde{Z}_{*,3} + \tilde{Z}_{*,1} \geq x_{*,1} + 1|\mathbf{x},\mathbf{z},\mathbf{x}_*) \\ &\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) \int \left(1 - \Phi\left(\frac{l(x_{*,1}+1) - \mu_{j,3|1}^m}{\sqrt{\sigma_{j,3|1}^{2\,m}}}\right)\right) \text{logN}(\tilde{z}_{*,1}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,1)}^m, \boldsymbol{\Sigma}_{j,(1,1)}^m) d\tilde{z}_{*,1}, \end{aligned}$$

where $l(z) = \log(\max(0, z - \tilde{z}_{*,1}))$. The conditional predictive density of the (undiscretized) age at first child $\breve{Z}_{*,3}$ given the (undiscretized) age at union $\tilde{z}_{*,2}$ and $x_{*,1}$ is:

$$f(\breve{z}_{*,3}|\tilde{z}_{*,2},\mathbf{x},\mathbf{z},\mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(\mathbf{x}_*) f(\breve{z}_{*,3}|\tilde{z}_{*,2},\boldsymbol{\theta}_j^m,\mathbf{x}_*) \frac{\text{logN}(\tilde{z}_{*,2}|\mathbf{x}_*\boldsymbol{\beta}_{j,(\cdot,2)}^m, \boldsymbol{\Sigma}_{j,(2,2)}^m)}{f(\tilde{z}_{*,2}|\mathbf{x},\mathbf{z},\mathbf{x}_*)}. \quad (5)$$

Notice that this expression differs from equation (3) in that

$$f(\breve{z}_{*,3}|\tilde{z}_{*,2},\boldsymbol{\theta}_j^m,\mathbf{x}_*) = \int_{-\infty}^{\breve{z}_{*,3}} \text{logN}(\breve{z}_{*,3} - \tilde{z}_{*,1}|\mu_{j,3|(1,2)}^m, \sigma_{j,3|(1,2)}^{2\,m}) \text{logN}(\tilde{z}_{*,1}|\mu_{j,1|2}^m, \sigma_{j,1|2}^{2\,m}) d\tilde{z}_{*,1},$$

where

$$\begin{aligned}\mu^m_{j,3|(1,2)} &= \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,3)} + \boldsymbol{\Sigma}^m_{j,(3,1:2)}\boldsymbol{\Sigma}^{-1\,m}_{j,(1:2,1:2)}(\log(\tilde{z}_{*,1:2}) - \mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,1:2)}),\\ \sigma^{2\,m}_{j,3|(1,2)} &= \boldsymbol{\Sigma}^m_{j,(3,3)} - \boldsymbol{\Sigma}^m_{j,(3,1:2)}\boldsymbol{\Sigma}^{-1\,m}_{j,(1:2,1:2)}\boldsymbol{\Sigma}^m_{j,(1:2,3)}.\end{aligned} \tag{6}$$

Figure C.13 shows the conditional predictive density of $\breve{Z}_{*,3} - \tilde{z}_{*,2}$ given $\tilde{z}_{*,2}$ and $\mathbf{x}_*$, which can be easily computed from (5), with the corresponding predictive medians in Figure C.10. Lastly, the conditional predictive probability of success for a binary response, e.g. $\ell = 4$ in our application, given $\breve{z}_{*,3}$ and $\mathbf{x}_*$ is:

$$\mathbb{P}(Z_{*,4} = 1|\breve{z}_{*,3}, \mathbf{x}, \mathbf{z}, \mathbf{x}_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w^m_j(\mathbf{x}_*)\mathbb{P}(Y_{*,4} > 0|\breve{z}_{*,3}, \boldsymbol{\theta}^m_j, \mathbf{x}_*)\frac{f(\breve{z}_{*,3}|\boldsymbol{\theta}^m_j, \mathbf{x}_*)}{f(\breve{z}_{*,3}|\mathbf{x}, \mathbf{z}, \mathbf{x}_*)},$$

where

$$\begin{aligned}&\mathbb{P}(Y_{*,4} > 0|\breve{z}_{*,3}, \boldsymbol{\theta}^m_j, \mathbf{x}_*)f(\breve{z}_{*,3}|\boldsymbol{\theta}^m_j, \mathbf{x}_*)\\ &\quad= \int_{-\infty}^{\log(\breve{z}_{*,3})} \Phi\left(\frac{\mu^m_{j,4|(1,3)}}{\sqrt{\sigma^{2\,m}_{j,4|(1,3)}}}\right) \mathrm{logN}(\breve{z}_{*,3} - \tilde{z}_{*,1}|\mu^m_{j,3|1}, \sigma^{2\,m}_{j,3|1})\mathrm{logN}(\tilde{z}_{*,1}|\mathbf{x}_*\boldsymbol{\beta}^m_{j,(\cdot,1)}, \boldsymbol{\Sigma}^m_{j,(1,1)})d\tilde{z}_{*,1},\end{aligned}$$

where $\mu_{j,4|(1,3)}$ and $\sigma^2_{j,4|(1,3)}$ are calculated analogously to expression (6).

Appendix B: Simulation study

We assess the performance of the proposed procedure on a simulated data set including $q^* = 3$ covariates and $d = 3$ responses. The first covariate mimics *Age* and, as such, is assumed to be registered at a discrete level: $x_1 = \lfloor \tilde{x}_1 \rfloor$, where $\tilde{x}_1 \sim \mathrm{U}(15, 30)$. The remaining covariates are categorical: x^*_2 has three levels with probabilities $(0.5, 0.3, 0.2)$ while x^*_3 has two levels with probabilities $(0.4, 0.6)$.

We generate two positive discretized responses and one binary response. The first response, Z_1, is a discretized noisy observation of a nonlinear function of x_1. To build Z_1, we first generate:

$$\tilde{Z}_{i,1} = \mu^t_1(\tilde{x}_{i,1}, x^*_{i,2}, x^*_{i,3}) + \epsilon_{i,1}, \text{ for } i = 1, \ldots, n,$$

where $\epsilon_{1,1}, \ldots, \epsilon_{n,1} \overset{i.i.d.}{\sim} 0.9\mathrm{N}(-15/90, 0.5^2) + 0.1\mathrm{N}(1.5, 0.75^2)$, and

$$\mu^t_1(\tilde{x}_{i,1}, x^*_{i,2}, x^*_{i,3}) = \begin{cases} -0.057\tilde{x}^2_{i,1} + 3.08\tilde{x}_{i,1} - 21.247 & \text{if } x^*_{i,2} \neq 1, x^*_{i,3} = 2\\ \frac{1}{3}\tilde{x}_{i,1} + 10 & \text{if } x^*_{i,2} \neq 1, x^*_{i,3} = 1\\ 0.0001\tilde{x}^3_{i,1} - 0.0695\tilde{x}^2_{i,1} + 3.83\tilde{x}_{i,1} - 30.584 & \text{if } x^*_{i,2} = 1, x^*_{i,3} = 2\\ \frac{8}{15}\tilde{x}_{i,1} + 7 & \text{if } x^*_{i,2} = 1, x^*_{i,3} = 1 \end{cases}.$$

Similarly, Z_2 is a discretized noisy observation of a nonlinear function of x_1 and the realized z_1 and it is built by generating:

$$\tilde{Z}_{i,2} = \begin{cases} -0.056\tilde{x}^2_{i,1} + 3.08\tilde{x}_{i,1} - 18 + 0.75\left[\tilde{z}_{i,1} - \mu^t_1(\tilde{x}_{i,1}, x^*_{i,2}, x^*_{i,3})\right] + \epsilon_{i,2} & \text{if } x^*_{i,3} = 2\\ 0.5\tilde{x}_{i,1} + 8 + 0.75\left[\tilde{z}_{i,1} - \mu^t_1(\tilde{x}_{i,1}, x^*_{i,2}, x^*_{i,3})\right] + \epsilon_{i,2} & \text{if } x^*_{i,3} = 1 \end{cases},$$

where the errors are assumed to depend also on $\tilde{x}_1$ and x_3^*:

$$\epsilon_{i,2} \sim \begin{cases} 0.9\mathrm{N}(-\frac{1}{6}, 0.4^2) + 0.1\mathrm{N}(1.5, 0.75^2) & \text{if } x_{i,3}^* = 2 \\ 0.9\mathrm{N}\left(-\frac{1}{6}, \left(\frac{7.5}{\tilde{x}_{i,1}}\right)^2\right) + 0.1\mathrm{N}\left(1.5, \left(\frac{7.5}{\tilde{x}_{i,1}}\right)^2\right) & \text{if } x_{i,3}^* = 1 \end{cases}.$$

Note that the response curves are the same for $x_2^* = 2, 3$ and differ for other categorical combinations, while the errors are not normal but right skewed, additionally depending on x_1 and x_3^* for the second response. Observed responses are set to missing for censored observations, defined as individuals with $\tilde{z}_{1,i} > \tilde{x}_{1,i}$ or $\tilde{z}_{2,i} > \tilde{x}_{1,i}$. Since the age-related variables in our motivating application are registered at a discrete level, the observed responses were rounded down to the nearest integer, i.e. $z_1 = \lfloor \tilde{z}_1 \rfloor$, $z_2 = \lfloor \tilde{z}_2 \rfloor$. The true curves and densities are depicted in Figure B.1 (top row) for selected combinations of the covariates. The behavior of Z_1 and Z_2 may at a first sight seem a simple function of x_1 alone, however it depends also on the interactions between x_1 and the categorical covariates. Moreover, this consideration holds both for the regression function and for the variance. Combined together these aspects, which are of the same nature as those present in our motivating data, pose challenges for parametric and semi-parametric models. Indeed, the relationship between each response and the *Age* at interview is relatively smooth, but interactions with the categorical covariates and changes in variability increase complexity. Finally, a binary response is simulated from a linear probit model depending only on x_1 (Figure B.2):

$$Z_{3,i} \sim \mathrm{Bern}\left(\Phi\left(\frac{\tilde{x}_{1,i} - 18}{6}\right)\right).$$

We seek to recover the conditional distribution of the response variables given the covariates using our proposed model, from a sample of size $n = 700$. We define the link functions $h_\ell(\mathbf{y}, \mathbf{x})$ as:

$$\begin{aligned} z_\ell &= h_\ell(\mathbf{y}, \mathbf{x}) = c_\ell(\mathbf{y}, \mathbf{x}) \lfloor \exp(y_\ell) \rfloor, \quad \text{for } \ell = 1, 2, \\ z_3 &= h_3(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{[0,\infty)}(y_3), \end{aligned} \tag{7}$$

where $c_\ell(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{(0, x_1+1)}(\exp(y_\ell))$. In this case, the bounds required in the adaptive MCMC are obtained from inverting $z_\ell = h_\ell(\mathbf{y}, \mathbf{x})$; concretely,

$$(l_\ell, u_\ell) = \begin{cases} (\log(x_1 + 1), \infty) & \text{for censored } z_\ell = 0 \\ (\log(z_\ell), \log(z_\ell + 1)) & \text{for uncensored } z_\ell \neq 0 \end{cases}, \text{ when } \ell = 1, 2,$$

$$(l_3, u_3) = \begin{cases} (-\infty, 0) & \text{for } z_3 = 0 \\ (0, \infty) & \text{for } z_3 = 1 \end{cases}.$$

Prior specification and computational details. Prior parameters for the linear coefficients and covariance matrix of each component are specified empirically based on multivariate linear regression fit to the data. Specifically, for $\ell = 1, 2$, we set $y_{i,\ell} =$

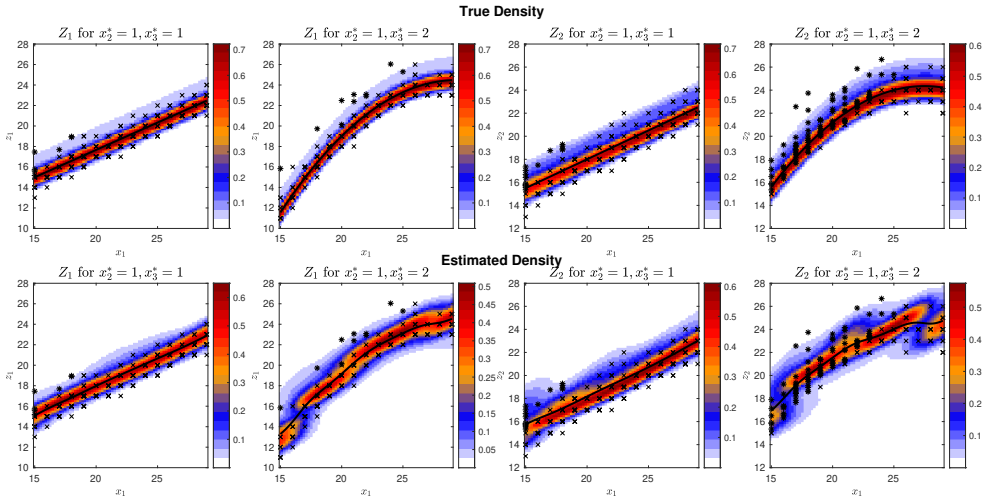

Figure B.1: Simulated data set. True data-generating density (top row) and estimated predictive density (bottom row) of the (undiscretized) Z_1 and Z_2 as functions of x_1 for two combinations of the categorical covariates. The estimated/true mean function is depicted with a black solid line; crosses and stars mark respectively observed and censored points.

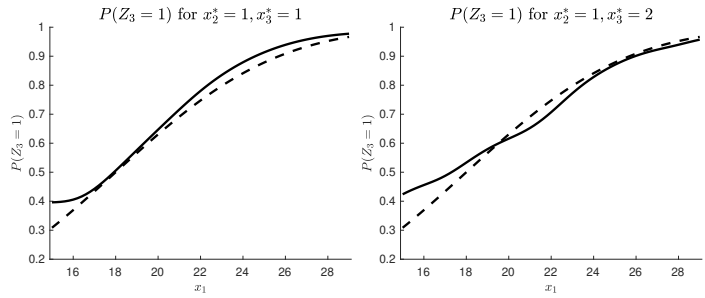

Figure B.2: Simulated data set. True (dashed line) and predictive (solid line) probability of $Z_3 = 1$ as a function of x_1 for two combinations of the categorical covariates.

$(l_{i,\ell} + u_{i,\ell})/2$ and $y_{i,\ell} = \log(x_{i,1} + 2)$ for uncensored and censored observations, respectively, where the bounds $l_{i,\ell}$ and $u_{i,\ell}$ are defined in Section 4. Additionally, we let $y_{i,3} = -1$ for $z_{i,3} = 0$ and $y_{i,3} = 1$ for $z_{i,3} = 1$. A multivariate linear regression fit on these auxiliary responses gives estimates $\widehat{\boldsymbol{\beta}}$ of the linear coefficients and $\widehat{\boldsymbol{\Sigma}}$ of the covariance matrix. We then define

$$\mathbb{E}[\boldsymbol{\beta}_j] = \boldsymbol{\beta}_0 = \widehat{\boldsymbol{\beta}} \quad \text{and} \quad \mathbb{E}[\boldsymbol{\Sigma}_j] = \frac{1}{\nu - b - 1}\boldsymbol{\Sigma}_0 = \widehat{\boldsymbol{\Sigma}}.$$

Together, $\mathbf{U}$ and $\boldsymbol{\Sigma}_j$ reflect the variability of $\boldsymbol{\beta}_j$ across components, and we set the matrix $\mathbf{U}$ such that $\min(\text{diag}(\widehat{\boldsymbol{\Sigma}}))\,\mathbf{U} = 10(\mathbf{X}^\intercal\mathbf{X})^{-1}$. The factor of this g-prior was selected to ensure reasonable ages (i.e. mostly lower than 100) in prior simulations. We explored more uninformative and vague prior choices but found that this could lead to quite large and unreasonable imputed ages for censored data. We further set $\nu = b + 3$, to ensure the existence of the first and second moments of $\boldsymbol{\Sigma}_j$ a-priori. Other specified hyperparameters include $\mu_{0,1} = \overline{x}_1$, $u_1 = 1/2$, $\alpha_1 = 2$, $\gamma_1 = u_1(\text{range}(x_{1:n,1})/4)^2$, $\boldsymbol{\varrho}_k = (1,1)$ for $k = p+1,\ldots,q$, and the parameters of the stick-breaking prior are $\zeta_{j,1} = 1$ and $\zeta_{j,2} = 1$. Here $\overline{x}_1$ and $\text{range}(x_{1:n,1})$ denote the sample mean and range of $(x_{1,1},\ldots,x_{n,1})$.

The MCMC stage of the adaptive truncation algorithm, with $J_0 = 15$ components, is run for 20,000 iterations after discarding the first 10,000 as burn-in. Every 10-th iteration is saved to produce M= 2,000 initial values for the particles in the SMC stage.

Results In Appendix A, we describe various posterior and predictive quantities that can be computed from the weighted particles to describe the relationship between the observed response $\mathbf{z}$ and covariates $\mathbf{x}$. Here, we focus on the marginal predictive mean and density functions for (undiscretized) Z_1 and Z_2, as well as on the marginal predictive probability of success for Z_3, and compare them with the true data-generating functions in Figures B.1 and B.2, for a selected combinations of the categorical covariates. We also show the conditional mean of Z_2 as a function of x_1 given different values of z_1 in Figure B.3. Overall, the model is able to recover the underlying structure present in the data, despite the heavy censoring of Z_2 for lower levels of x_1, particularly when $x_3^* = 2$. The true conditional structure is well recovered in areas where data (crosses) is available, i.e. in the left plot of Figure B.3, for $15 \leq x_1 \leq 18$ given $z_1 = 15$ (magenta) and for $27 \leq x_1 \leq 29$ given $z_1 = 23$ (black). However, as can be expected, the model struggles when predicting at values far from the observed data, i.e in the left plot of Figure B.3, for $25 \leq x_1 \leq 30$ given $z_1 = 15$ (magenta) and for $15 \leq x_1 \leq 20$ given $z_1 = 23$ (black). We highlight that interpretation of the conditional dependence structure in the latent scale as well as the latent covariance matrices of the mixture components and its relation to the dependence structure on the observed scale is an open and interesting direction of research, which would expand the work of García-Zattera et al. (2007) in the parametric setting.

Robustness analysis. We perform a robustness analysis comparing several initialization specifications, namely by setting $J_0 = 2, 3, 5, 10, 15, 20, 30$. We also compare with a parametric version of the model that is similar in nature to the parametric model of Korsgaard et al. (2003), i.e. a multivariate Gaussian regression model with the link functions $h_\ell(\mathbf{y}, \mathbf{x})$ in (7) and a prior given by the base measure $\mathbf{P}_0$. For the sake of comparison, we use the Metropolis-within-Gibbs scheme for inference. In all scenarios, the adaptive MCMC algorithm is run for 30,000 iterations, discarding the first 10,000 as burn-in, and saving only every 10th iteration for a total of M = 2,000 particles to be used in the SMC step. A summary of the analysis is reported in Table B.1. Besides the number of components inferred by the model (J^*), and the elapsed CPU time (in hours), the table reports the ESS for the log-likelihood in the MCMC stage (ESS_{MCMC})

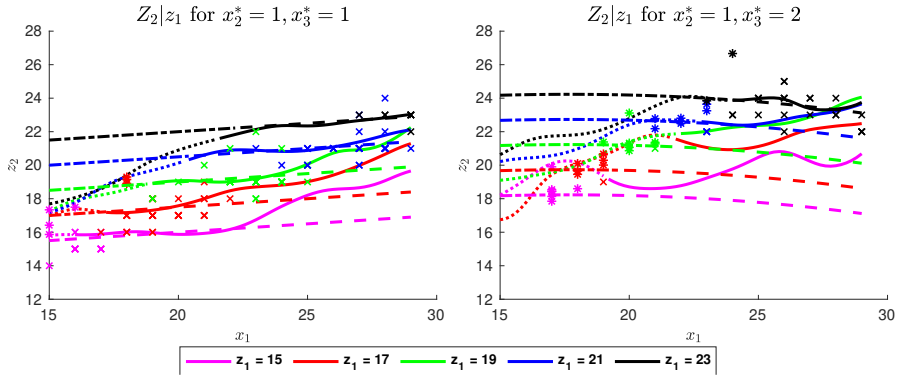

Figure B.3: Simulated data set. Conditional mean of Z_2 given z_1 as a function of x_1 with colors representing the different values of z_1, for two combinations of the categorical covariates. The estimated and true conditional mean functions are depicted with solid and dashed lines respectively, and dotted lines indicate when the mean exceeds x_1. Crosses and stars mark respectively the observed and censored points, colored by the observed value of z_1.

computed with the *mcmcse* package in **R** (Flegal et al., 2017), and the ESS_{J^*} of the final iteration of the SMC, that can be used to compare mixing. Results are reported for two different discrepancy measures used to define the stopping rule of SMC, namely the ESS and the CESS (Zhou et al., 2016). In order to assess the fit of the model, we compute for each Z_ℓ the log-pseudo marginal likelihood (LPML, Geisser and Eddy, 1979), and the percentage absolute errors with respect to the true mean and true density at a set of new test covariates, $\mathbf{x}_i^*$, for $i = 1, \ldots, n^*$, denoted by ERR_{Mean} and ERR_{Dens}:

$$\text{LPML}^\ell = \sum_{i=1}^{n} \log(\text{CPO}_i^\ell) \quad \text{with} \quad \text{CPO}_i^\ell = \left(\frac{1}{M} \sum_{m=1}^{M} \frac{1}{f(z_{i,\ell}|\mathbf{w}^m, \boldsymbol{\psi}^m, \boldsymbol{\theta}^m, \mathbf{x}_i)} \right)^{-1},$$

$$\text{ERR}_{\text{Mean}}^\ell = \frac{100}{n^*} \sum_{i=1}^{n^*} \frac{|\mu_\ell^t(\mathbf{x}_i^*) - \widehat{\mu}_\ell(\mathbf{x}_i^*)|}{|\mu_\ell^t(\mathbf{x}_i^*)|},$$

$$\text{ERR}_{\text{Dens}}^\ell = \frac{100}{n^*} \sum_{i=1}^{n^*} \frac{\int |f^t(z_\ell^*|\mathbf{x}_i^*) - \widehat{f}(z_\ell^*|\mathbf{x}_i^*)| dz_\ell^*}{\int |f^t(z_\ell^*|\mathbf{x}_i^*)| dz_\ell^*} \approx \frac{100}{n^*} \sum_{i=1}^{n^*} \sum_{g=1}^{G} |f^t(z_{g,\ell}^*|\mathbf{x}_i^*) - \hat{f}(z_{g,\ell}^*|\mathbf{x}_i^*)| \Delta,$$

where for each response $\ell = 1, \ldots, d$, $\mu_\ell^t(\mathbf{x}_i^*)$ and $\widehat{\mu}_\ell(\mathbf{x}_i^*)$ indicate the true and estimated mean functions, and $f^t(\cdot|\mathbf{x}_i^*)$ and $\widehat{f}(\cdot|\mathbf{x}_i^*)$ indicate the true and estimated densities. For each response, densities are evaluated on a grid of values, $z_{1,\ell}^*, \ldots, z_{G,\ell}^*$, with grid size Δ. The results show robustness with respect to the choice of the discrepancy measure.

We observe that for $J_0 \geq 20$ only a moderate number of components are added, suggesting that a sufficient approximation is obtained with around 20 components. Recall that the SMC is run for at least $I = 4$ cycles, i.e. at least four new components are added to the initial model. Therefore, if J_0 is large enough, we have $J^* = J_0 + I$. Generally, the computational time increases with J_0, although this is not always the case, especially

	J_0	J^*	CPU	ESS_{MCMC}	ESS_{J^*}	LPML (10^3)			ERR_{Mean}			ERR_{Dens}		
						Z_1	Z_2	Z_3	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
Parametric	1	1	0.66	767.3		-1.17	-0.82	-0.34	5.82	3.26	15.39	163.79	115.32	13.86
ESS_J	2	13	1.90	495.0	1,125.5	-1.01	-0.76	-0.34	3.93	5.38	6.99	132.55	119.69	7.23
	5	14	3.93	519.6	1,966.7	-0.91	-0.71	-0.34	2.94	4.89	7.41	97.80	104.74	6.59
	10	19	5.29	537.0	1,918.9	-0.94	-0.71	-0.34	2.41	3.12	8.42	90.47	98.56	7.99
	15	26	5.77	543.5	1,266.0	-0.93	-0.73	-0.35	2.94	3.75	9.30	82.99	109.44	8.32
	20	24	5.43	567.8	1,990.3	-0.86	-0.69	-0.34	2.31	3.56	7.92	78.89	95.45	7.88
	30	34	9.04	550.8	2,000.0	-0.85	-0.69	-0.34	2.59	3.67	8.37	78.06	97.60	7.96
CESS_J	2	14	3.67	495.0	1,989.8	-1.01	-0.76	-0.34	3.89	5.26	6.82	129.35	116.34	7.09
	5	14	3.90	519.6	1,978	-0.91	-0.71	-0.34	2.94	4.89	7.41	97.80	104.79	6.59
	10	17	5.18	537.0	1,905.9	-0.92	-0.71	-0.34	2.35	2.92	8.21	89.79	98.81	7.96
	15	23	6.13	543.5	1,974.9	-0.93	-0.73	-0.35	3.00	3.75	9.35	84.40	109.41	8.38
	20	24	5.51	567.8	1,994.1	-0.86	-0.69	-0.34	2.31	3.56	7.92	78.89	95.45	7.88
	30	34	11.17	550.8	2,000	-0.85	-0.69	-0.34	2.59	3.67	8.37	78.06	97.60	7.96

Table B.1: Simulated data set. Summaries of the performance: computational burden, mixing, goodness of fit, and predictive errors in mean and density obtained with the parametric model (first row) and the nonparametric model for different values of J_0. Results are reported for the adaptive truncation algorithm based on the ESS and CESS stopping rules.

if ESS_J becomes too low so that resampling and rejuvenation are required. Despite the increased number of parameters for large J_0, the mixing of the MCMC, reflected in the ESS_{MCMC}, does not deteriorate; however, note the improved mixing for the parametric model, which has the least number of parameters, due to the absence of the covariate-dependent weights. Focusing on the SMC, a larger J_0 generally results in less degeneracy of the particles, reflected in a higher ESS_{J^*}. Finally the LPML, measuring the goodness of fit of the model, increases with J_0, while the errors in predictive mean and density both decrease. This is particularly true for Z_1, the most nonlinear response, while there is little improvement in the binary response Z_3, which is indeed simulated from a linear probit model. Similar results are obtained when substituting the ESS with the CESS in the discrepancy measure of the SMC, confirming robustness to the choice of the stopping rule. To conclude, initializing the algorithm with a conservative number of components provides a good compromise between computational time, mixing, and accuracy.

Focusing on the model flexibility and its ability to recover the correct structure present in the data, we explore three additional scenarios: (i) longer number of iterations (25,000 burn-in; 10 thinning; $M = 2,500$); (ii) doubled sample size ($n = 1,400$); (iii) omitted censoring. In all cases, $J_0 = 15$. The results are reported in Table B.2, showing the same goodness-of-fit and error indicators used in Table B.1. As expected, mixing improves for scenario (i), due to the higher number of iterations, and modest improvements in the predictive power are observed. Increasing the sample size (scenario (ii)) without adjusting the algorithm settings (e.g., length of the chain, J_0), does not lead to consistent improvements. Note that this scenario yields approximately doubled LPML values, due to the definition of this quantity. The removal of censoring (scenario (iii)) yields faster and more precise computations. Estimates of the predictive densities for the undiscretized variable Z_1 as function of x_1 are reported in Figure B.4.

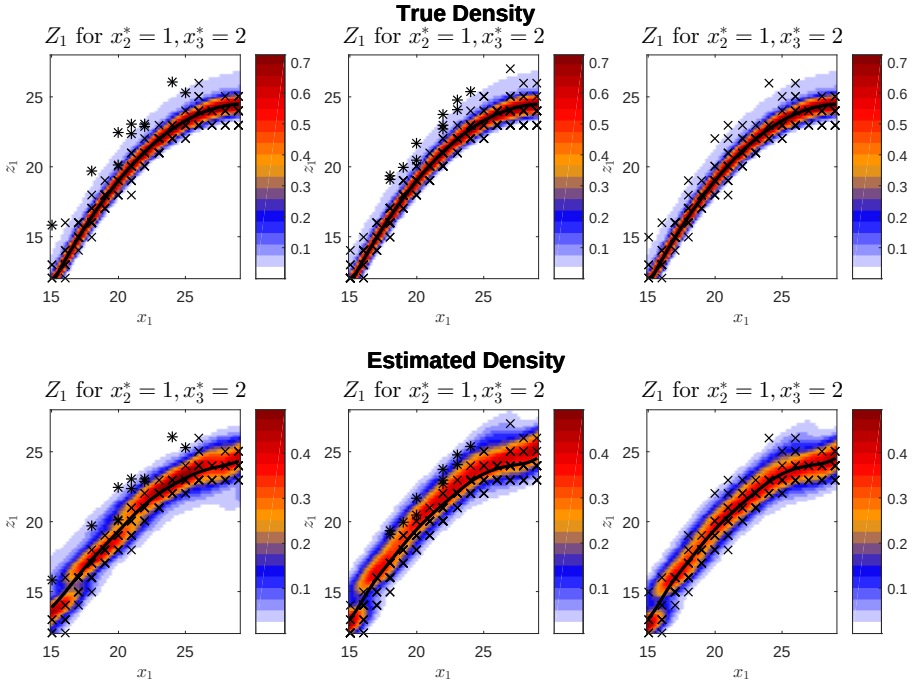

Figure B.4: Simulated data set. Alternative scenarios by column: (i) longer MCMC chain; (ii) $n = 1,400$; (iii) no censoring. True data-generating density (top) and estimated predictive density (bottom) of the (undiscretized) Z_1 as a function of x_1 for one combination of the categorical covariates. The true and estimated mean functions are depicted with black solid lines; crosses and stars mark observed and censored points, respectively.

	J^*	CPU	ESS_{MCMC}	ESS_{J^*}	LPML (10^3)			ERR_{Mean}			ERR_{Dens}		
					Z_1	Z_2	Z_3	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
(i)	23	10.24	619.1	1,352	-0.93	-0.74	-0.35	2.60	3.85	9.41	80.83	105.49	8.28
(ii)	23	12.15	516.0	1,117.2	-1.75	-1.4	-0.73	2.60	3.86	9.71	87.85	103.60	9.32
(iii)	23	4.42	560.5	1,070.9	-0.88	-0.96	-0.34	2.33	2.13	8.59	75.34	79.74	7.66

Table B.2: Simulated data set. Alternative scenarios: (i) longer MCMC chain; (ii) $n = 1,400$; (iii) no censoring. Summaries of the performance: computational burden, mixing, goodness of fit, and predictive errors in mean and density obtained with the algorithm initialized at $J_0 = 15$.

Appendix C: Application: life patterns of Colombian women

In the following sections, we provide some further discussion on the DHS data characteristics and report some additional figures, complementing and enriching the results reported in the main paper. We begin by providing a map of the regions of Colombia

used in the study in Figure C.5. Different territorial divisions of Colombia are used in different contexts. We follow Ojeda et al. (2011), considering six regions: Atlantica, Oriental, Central, Pacifica, Bogota, Territorios Nacionales. We point out that Bogota is in fact a part of the Oriental region but was treated separately because of its peculiar features in terms of social and economic development. Territorios Nacionales could be divided into smaller, more homogeneous regions (e.g., Orinoquia and Amazon), but this is the definition employed by the DHS for the 2010 survey.

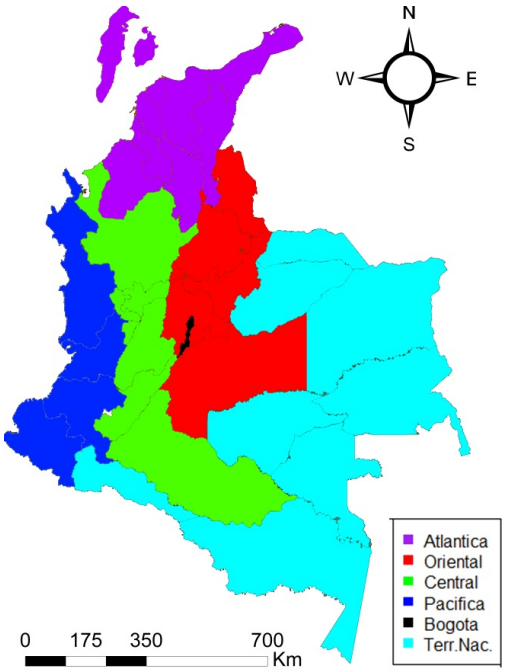

Figure C.5: Map of Colombia identifying the six regions considered by the DHS for the data collection and final report (Ojeda et al., 2011), and adopted in this work.

C.1 The data: some considerations on sample weights

The data arise from a complex survey design, and thus have associated weights with a rather complicated structure due both to survey design and to additional post-stratification carried out to adjust for various factors (e.g., the total number of women interviewed in each municipality, non-response, etc.; see Ojeda et al., 2011, for details). This creates difficulties in further adjusting the weights based on the filtering considered (see Section 2 in the main text). Specifically, we do not know the population proportions for each filter within each municipality (e.g. of ethnicity, or of residents that have lived in the same region from at least the age of 6). Thus, the weights for the filtered data

would need to be post-adjusted based on the observed sample proportions, leading to uncertainty in the weights. In addition, how this correction should relate to the post-stratification is unclear. We emphasize that even if we were able to obtain accurate weights for the filtered data, the use of survey weights in regression analyses is debated within the literature. From a frequentist perspective, Winship and Radbill (1994) point out that when ordinary least squares (OLS) and weighted OLS lead to similar estimates the former is preferred, because it leads to improved efficiency, power, coverage. In fact, the authors state that "where the sampling weights are solely a function of the independent variables" (as in our case), "unweighted OLS estimates are preferred because they are unbiased, consistent, and have smaller standard errors." Moreover, the authors suggest that when OLS and weighted OLS estimates are substantially different one should investigate more flexible models, with e.g. non-linearity or interactions. Thus, since our model is nonparametric and can flexibly recover non-linearity, interactions, multi-modality (that may be due to missing or unmeasured covariates), we prefer the unweighted median regression and density regression estimates.

From a Bayesian perspective, although our model and algorithm could be adjusted to include the sample weights, for example through a pseudo (i.e. weighted) likelihood or via data augmentation, as outlined by Gunawan et al. (2020), this may lead to Bayesian interval estimates that do not have the correct frequentist coverage. Indeed, the authors show that this is the case with the pseudo likelihood approach, and provide empirical evidence that data augmentation can perform better. In summary, due to the issues involved in the determination of accurate weights for the filtered data and to the poor coverage and efficiency of the weighted estimates that has been identified in literature, we carry out inference based on the unweighted data, which is further supported by the flexibility of our model.

C.2 Additional figures

We display additional figures, enriching the results reported in the main text, and for convenience, comments on possibly relevant findings are reported in the figures' captions. Figure C.6 complements Figure 5 by reporting median ages at events for women who grew up in violent environments with only physical punishment or only parental domestic violence, i.e $(\mathbf{P}, \bar{\mathbf{B}})$ or $(\bar{\mathbf{P}}, \mathbf{B})$. Figure C.7 completes Figure 6 by displaying the predictive density of the age at sexual debut. Figure C.8 reports the predictive probability of censoring, that is the probability that a woman will experience the event after the given *Age*, as a function of *Age*. Turning to the conditional analysis, the conditional predictive medians for the time from sexual debut to first child given the age at sexual debut is shown in Figure C.9 and for the time from union to first child given the age at union is shown in Figure C.10. The underlying conditional predictive densities for selected covariate combinations are visualized in Figures C.11, C.12, and C.13. Finally, to explore the possible relation between anticipation of union on work activity, Figure C.14 reports the conditional predictive probability of working as function of *Age* given different ages at union.

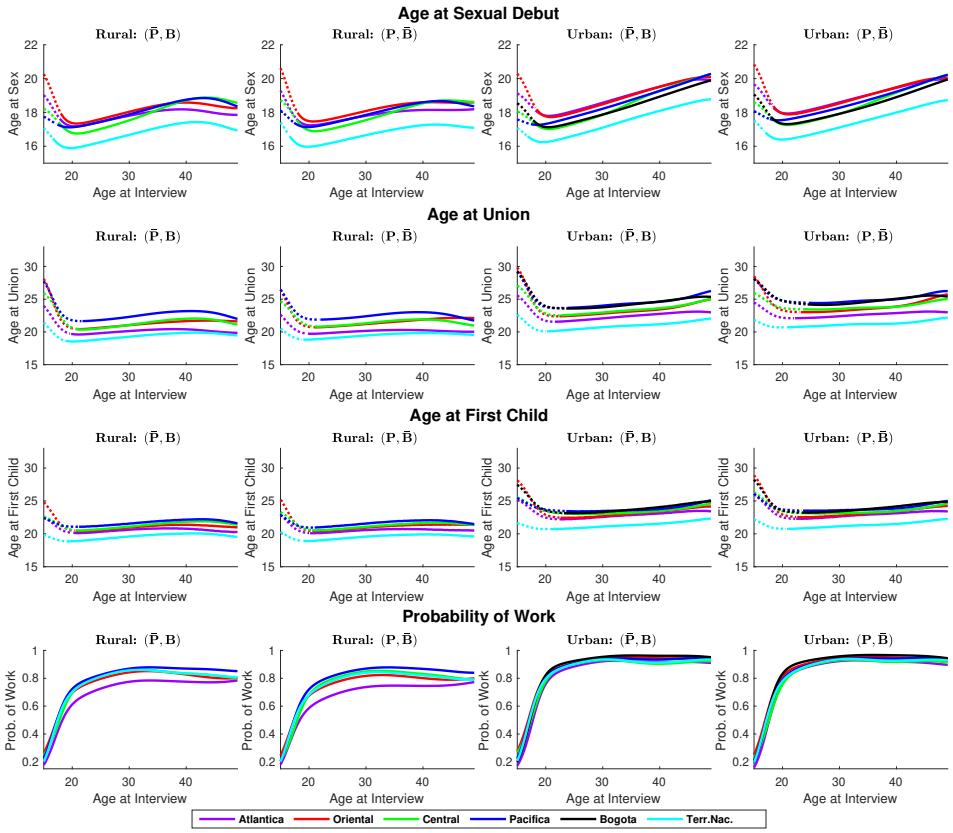

Figure C.6: Predictive medians of the ages at sexual debut, union and child, and posterior probability of working, as functions of *Age*, for women who grew up in violent environments with only physical punishment or only parental domestic violence, i.e $(\mathbf{P}, \bar{\mathbf{B}})$ or $(\bar{\mathbf{P}}, \mathbf{B})$. Dotted lines indicate when the median exceeds *Age*. Combined with Figure 5, observe that median ages increase as violence levels decrease, while the probability of working increases in younger cohorts for greater violence levels. This provides evidence for an anticipation of adulthood as violence levels increase.

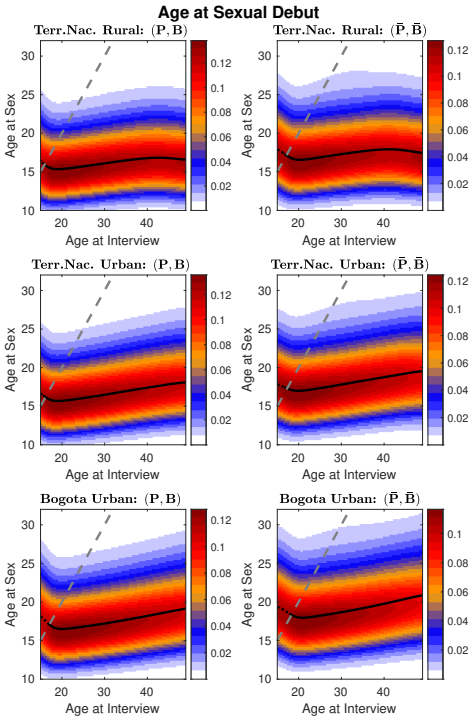

Figure C.7: Predictive density of the age at sexual debut as a function of *Age* for women who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent families $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$. Analogously to Figure 6, results are reported for urban and rural areas of the least developed region (Territorios nacionales) and for the capital (Bogota). The region above the dashed line indicates when age at event exceeds *Age*. The black line is the posterior median function. The median represents well the center of the distribution, and a decrease in both the median and dispersion of sexual debut is observed in younger cohorts, particularly in urban and developed regions.

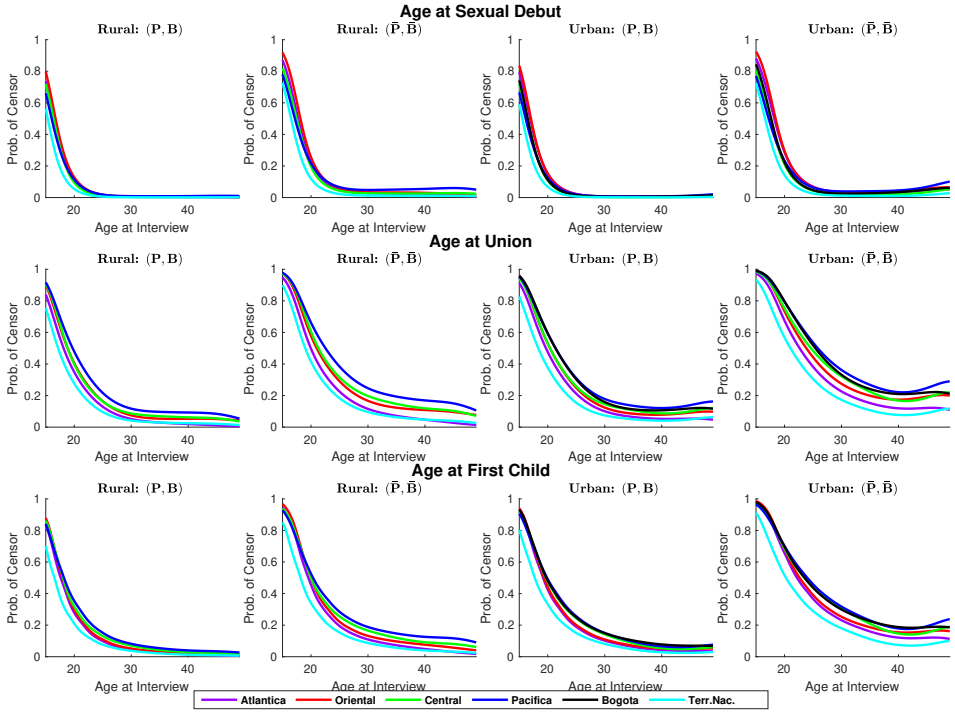

Figure C.8: The predictive probability of censoring represents the probability that a woman will experience the event after the specified *Age* and is depicted for the events of sexual debut, union and child as a function of *Age*, for women who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent families $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$. Equivalently, the censoring probability represents the mass above the dashed line for a given *Age* in the density plots of Figures 6 and C.7; when the right tail in the density exceeds the dashed line, interpreting the censoring probability is more reliable than focusing on the shape of the right tail. As expected, higher censoring probabilities are observed for younger cohorts and more developed regions and for the age at union and child over sexual debut.

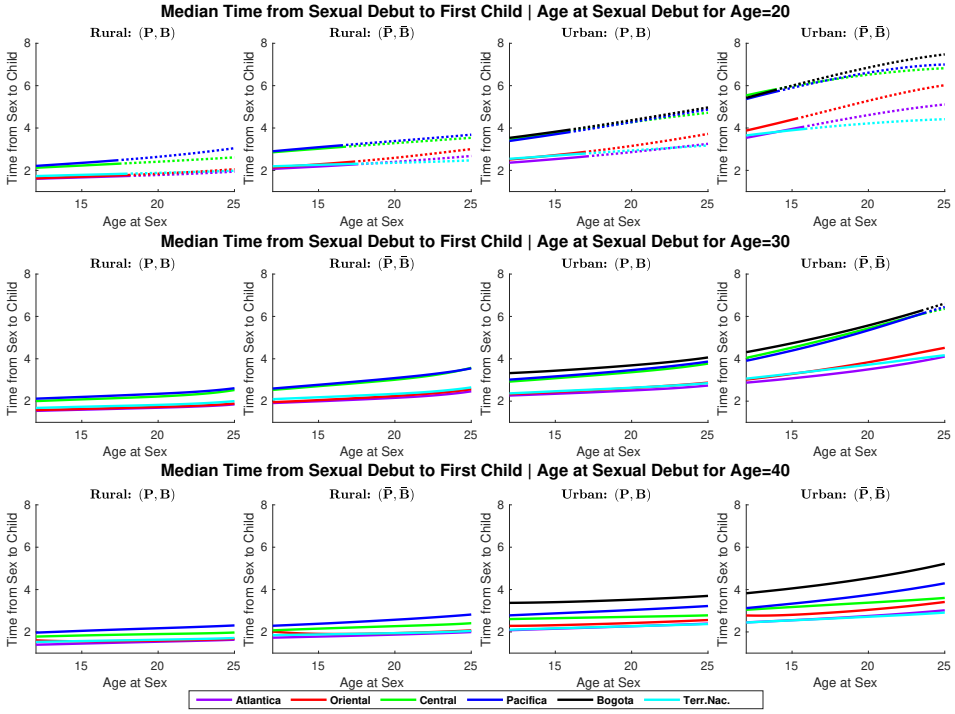

Figure C.9: Conditional predictive medians of the time from sexual debut to first child given the age at sexual debut, as a function of the latter, for women with $Age = 20, 30, 40$, who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent families $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$. Dotted lines indicate when the age at child is higher than the *Age*. Notice that medians are higher for younger cohorts; thus, although we observe an anticipation of sexual debut in younger generations in Figure 5, these women tend to wait longer between sexual debut and first child. We can also appreciate a polarization between Atlantica, Oriental, and Territorios Nacionales on one side and Central, Pacifica, and Bogota on the other, particularly as *Age* increases.

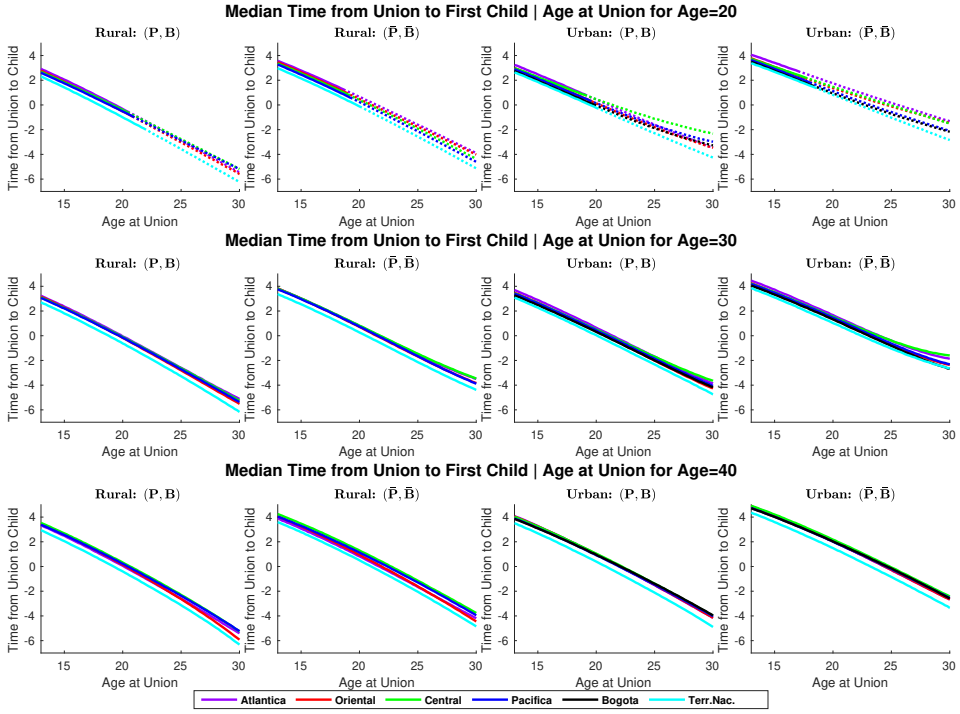

Figure C.10: Conditional predictive medians of the time from union to first child given the age at union, as a function of the latter, for women aged 20, 30, and 40 at interview and who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent families $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$. Dotted lines indicate when the age at child is higher than the *Age*. As can be expected, median time from union to child decreases with age at union. Indeed, it is negative for high values of age at union, particularly in rural areas and for violent family environments, suggesting a greater tendency to have children out of wedlock.

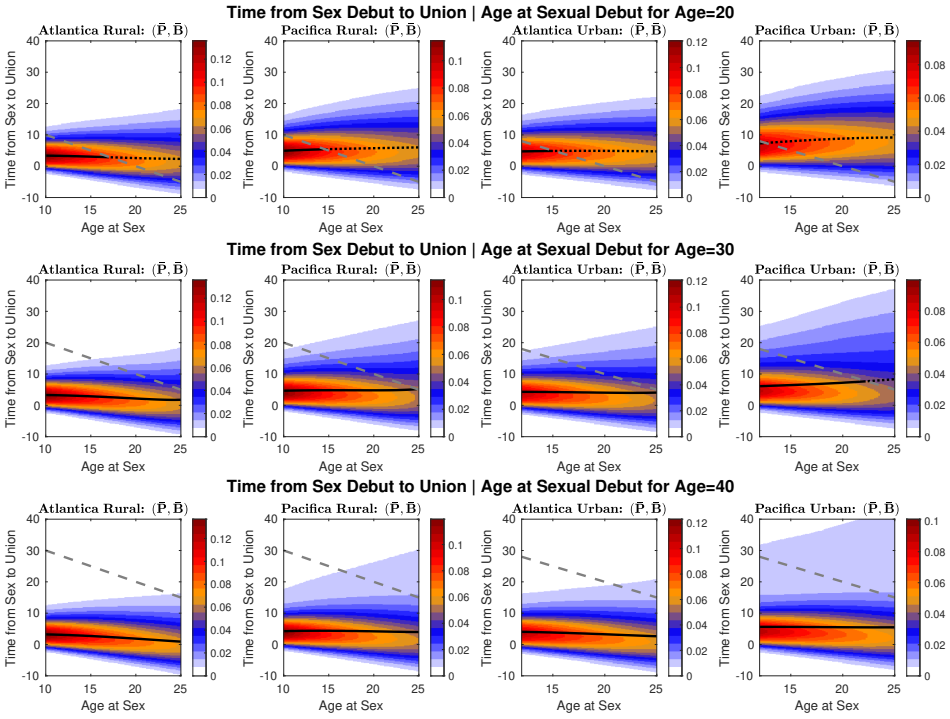

Figure C.11: Conditional predictive density of the time from sexual debut to union given age at sexual debut, as a function of the latter, for women with $Age = 20, 30, 40$. Results are shown for women who grew up in a non-violent family $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$ and for urban and rural areas of Atlantic and Pacifica. The region above the dashed line indicates when age at union exceeds *Age*. Combined with Figure 7, we observe that women in Pacifica and Bogota compared with Atlantica and Territorios Nacionales (and to a lesser extent Oriental) not only have a higher median time from sexual debut to union but also increased dispersion and a heavier right tail, reflecting a wider variety of choices for women to delay union after sexual debut in these regions. Additionally, a slight increase in median time and dispersion can be appreciated for decreasing *Age*, supporting a weaker relation between sexual debut and union in younger cohorts, that is more evident in developed urban areas.

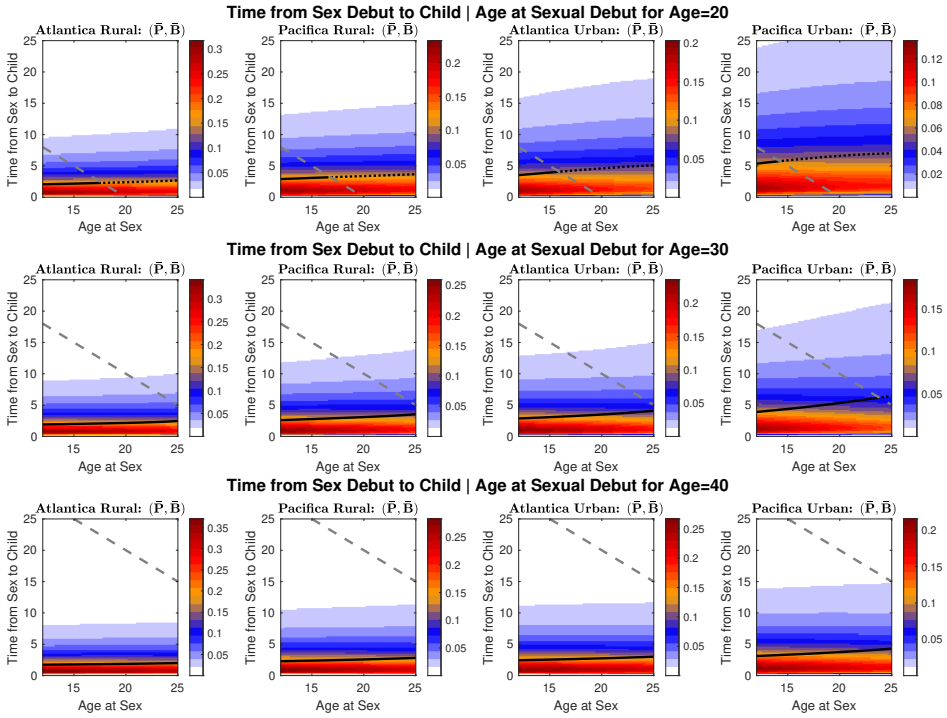

Figure C.12: Conditional predictive density of the time from sexual debut to child given age at sexual debut, as a function of the latter, for women with $Age = 20, 30, 40$. Results are shown for women who grew up in a non-violent family $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$ and for urban and rural areas of Atlantic and Pacifica. The region above the dashed line indicates when age at child exceeds *Age*. The heavier right tail, reflecting a wider variety of choices for women to delay motherhood after sexual debut, is evident as *Age* increases, particularly in developed urban areas. This supports the claim of a weaker relation between sexual debut and motherhood in younger cohorts.

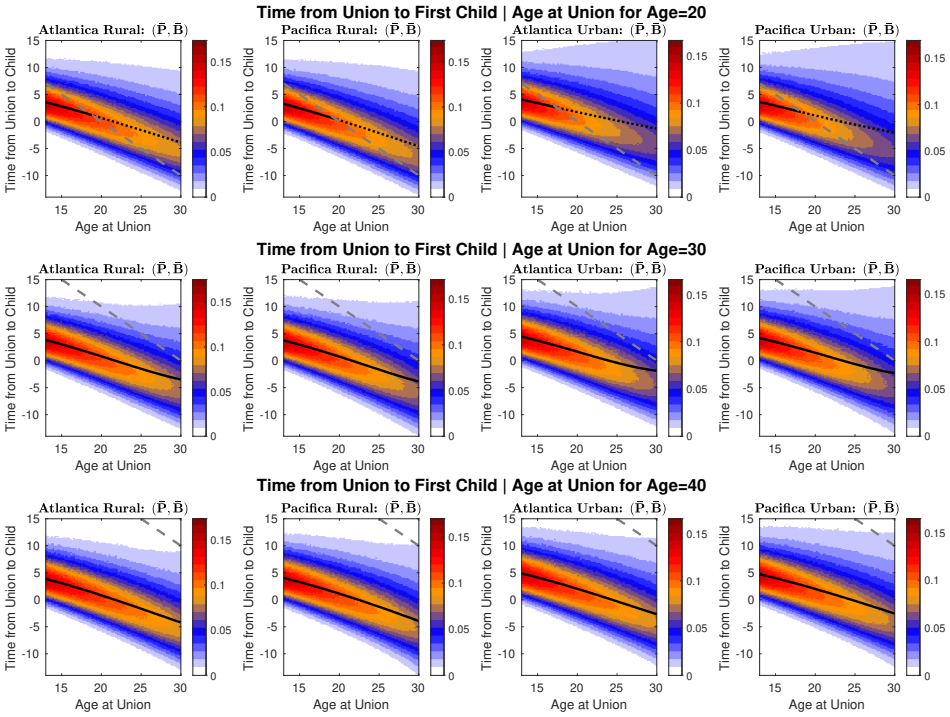

Figure C.13: Conditional predictive density of the time from union to first child given age at union, as a function of the latter, for women with $Age = 20, 30, 40$. Results are shown for women who grew up in a non-violent family $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$ and for urban and rural areas of Atlantic and Pacifica. The region above the dashed line indicates when age at first child exceeds *Age*.

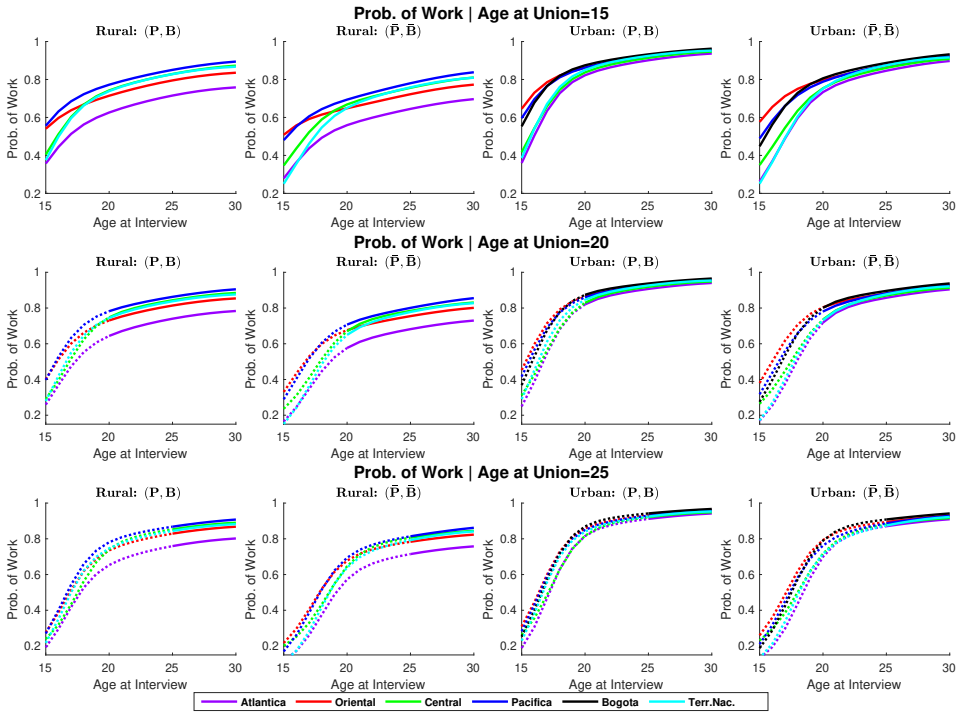

Figure C.14: Conditional predictive probability of working as function of *Age* given different ages at union, for women who grew up in violent $(\mathbf{P}, \mathbf{B})$ and non-violent families $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$. Dotted lines indicate when *Age* is less than the age at event. While we observe an increased probability of working for young cohorts that established an early union, in contrast to Figure 8, no scaring effect is visible, i.e. the probability of working in older cohorts is unaffected by the conditioned age at union.

References

Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2017). *mcmcse: Monte Carlo standard errors for MCMC*. R package version 1.3-2.

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 & Data Analysis*, 51(6): 3223–3234.

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

Gunawan, D., Panagiotelis, A., Griffiths, W., and Chotikapanich, D. (2020). "Bayesian weighted inference from surveys." *Australian & New Zealand Journal of Statistics*, 62(1): 71–94.

Korsgaard, I. R., Lund, M. S., Sorensen, D., Gianola, D., Madsen, P., and Jensen, J. (2003). "Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling." *Genetics Selection Evolution*, 35(2): 159–183.

Ojeda, G., Ordóñez, M., and Ochoa, L. H. (2011). *Colombia Encuesta Nacional de Demografía y Salud 2010*. Bogotá, Colombia: Profamilia. Available at `http://dhsprogram.com/pubs/pdf/FR246/FR246.pdf`.

Winship, C. and Radbill, L. (1994). "Sampling weights and regression analysis." *Sociological Methods & Research*, 23(2): 230–257.

Zhou, Y., Johansen, A. M., and Aston, J. A. D. (2016). "Toward automatic model comparison: an adaptive sequential Monte Carlo approach." *Journal of Computational and Graphical Statistics*, 25(3): 701–726.