Multivariate Density Regression for Censored, Constrained, and Binary Traits

S. Wade∗ R. Piccarreta † A. Cremaschi‡ I. Antoniano-Villalobos §

July 23, 2019

Abstract

Women in Colombia face difficulties related to the patriarchal traits of their societies and well-known conflict afflicting the country since 1948. In this critical context, our aim is to study the relationship between baseline socio-demographic factors and variables associated to fertility, partnership patterns, and work activity. To best exploit the explanatory structure, we propose a Bayesian multivariate density regression model, which can accommodate mixed responses with censored, constrained, and binary traits. The flexible nature of the models allows for nonlinear regression functions and non-standard features in the errors, such as asymmetry or multi-modality. The model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of categorical and continuous covariates. Computational difficulties for inference are overcome through an adaptive truncation algorithm combining adaptive Metropolis-Hastings and sequential Monte Carlo to create a sequence of automatically truncated posterior mixtures. For our study on Colombian women’s life patterns, a variety of quantities are visualised and described, and in particular, our findings highlight the detrimental impact of family violence on women’s choices and behaviors.

Keywords: Bayesian nonparametrics, adaptive truncation, sequential Monte Carlo, life patterns, time-to-event

∗School of Mathematics, University of Edinburgh, UK
†BIDSA and Dept. of Decision Sciences, Bocconi University, Milan, Italy
‡Dept. of Cancer Immunology, Institute of Cancer Research, Oslo University Hospital, Norway
§Dept. of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, Italy
1 Introduction

Colombian women face difficulties that are quite typical in Latin American countries, particularly related to the patriarchal traits of their society. Nonetheless, the welfare of Colombian women is possibly more critical due to the conflict between state military forces, paramilitaries, and guerrilla groups that has afflicted the country since 1948. In such a critical context, we are interested in studying women’s life events, focusing on the interplay between sexual initiation (debut), fertility, partnership, and participation in the labor market. Thus, rather than focusing on a specific life event, as in previous relevant studies (e.g. Gimenez Duarte et al., 2013; Azevedo et al., 2012; Restrepo Martínez et al., 2017), we adopt a broader perspective, considering a collection of events describing transition to adulthood and their relation with a set of structural baseline characteristics of the women’s environment and family. Besides some of the well known relevant factors – such as cohort, region, and area (urban or rural) of residence – we also study whether a violent family context contributes to shape transition to adulthood and possibly impairs women’s agency.

To this purpose, we analyze data arising from the survey conducted in Colombia in 2010 as a part of the Demographic and Health Survey (DHS) Program (implemented by the Inner City Fund and funded by USAID, https://www.dhsprogram.com). The data are cross-sectional, thus, no follow-up information on the life events of interest is recorded. Specifically, information is available on the age when the considered focal events – sexual debut, marriage or cohabitation, motherhood – were experienced for the first time, whereas work information concerns only the employment status of the woman (working or not) at the moment of the interview. Thus, we jointly analyze response variables with different levels of measurements (times at event and binary variables). Additionally, the focal events may not have been experienced (right-censoring) and are subject to constraints, e.g. motherhood can only occur after sexual debut. Furthermore, the available set of baseline explanatory variables is limited, so that heterogeneity may be present which would not be properly captured by a parametric model. This encourages the use of a flexible model to best exploit the explanatory structure without imposing possibly penalizing constraints. Additionally, such a model can encompass competing sociological theories which may be relevant in different subpopulations.

The data present various features that challenge existing parametric and semiparametric
models (e.g., Korsgaard et al., 2003; Jara et al., 2010; Hanson and Johnson, 2004; Kottas and Gelfand, 2001). First, some women postpone the events to relatively late in life, which induces right-skewed distributions. Also, the joint relationships between the age-at-event variables show different patterns, with gaps of various lengths between events. Moreover, these behaviors change depending on the covariates. Modeling such dependence structure is an ambitious task, requiring a model that allows for i) non-linear response curves, ii) non-normal distributions whose features may change with the covariates, iii) multivariate response and covariates of mixed nature, and iv) censoring and constraints of the responses. To the best of our knowledge, a model that can simultaneously deal with these issues does not exist.

We propose a Bayesian multivariate density regression model that extends the univariate model of Antoniano-Villalobos et al. (2014) to the case of multiple mixed-type responses with censoring and constraints. This approach is promising for our data, due to its ability to capture their peculiar features. Our infinite mixture model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of categorical and numerical covariates. In addition, the multivariate approach permits to study the joint relationship between the response variables, for example by considering one response conditioned on the others. With data on over 10,000 women and a multivariate response and covariate, the Markov chain Monte Carlo (MCMC) algorithm originally proposed for the univariate model becomes unsuitable. We therefore propose an algorithm for posterior inference that extends the adaptive truncation scheme of Griffin (2016).

The paper is structured as follows. The model and posterior simulation algorithm are presented in Sections 2 and 3, respectively. The model’s performance is first assessed via a simulation study in Section 4. Then, the results for the data on Colombian women are analyzed in Section 5. Section 6 summarizes and concludes.

2 Bayesian Nonparametric Density Regression

We develop a Bayesian nonparametric mixture model that can capture the relationship between \( n \) conditionally independent \( d \)-dimensional response vectors, \( \mathbf{Z}_i \), and a vector \( \mathbf{x}_i \) of predictors. To simplify notation, whenever possible we drop the sub-index \( i \) indicating indi-
vidual observations. The predictors $x^* = (x_1, \ldots, x_p, x_{p+1}^*, \ldots, x_q^*)$ may be of mixed nature. Without loss of generality, we assume that the first $p$ are numerical while the rest are categorical. As is common in regression models, we expand the categorical predictors with binary dummy variables and let $x = (x_1, \ldots, x_p, x_{p+1}, \ldots, x_q)$, where $q = p + \sum_{k=p+1}^{q} (R_k - 1)$ and $R_k$ denotes the number of categories of $x_k^*$. The observed responses are also of mixed nature. For example, in our application, we consider two types of responses: three positive integer-valued variables with possible censoring and constraints, representing the ages at events, and one binary variable indicating work status. In this case, we refer to the density of the mixed response $Z = (Z_1, \ldots, Z_d)$ with respect to the appropriate measure, e.g. Lebesgue or counting measure, for each response type.

To frame our model within existing literature, we review some related contributions. Bayesian nonparametric mixture models (Lo, 1984) are useful tools for density estimation, due to their attractive balance between flexibility and smoothness and ability to recover a wide range of densities (Ghosh and Ramamoorthi, 2003, Chapter 5). Extensions for conditional density estimation, also known as density regression, can be found in the pioneering works of Müller et al. (1996) and MacEachern (1999). In the latter, the Bayesian nonparametric mixture model is extended by allowing the mixing measure to depend on the covariates. This yields flexible density regression, where the entire density and not only the mean is regressed on the covariates. Several approaches exist in literature to specify the covariate-dependent mixing measure, but it is not clear how to choose between them. Examples include single-$p$ dependent Dirichlet processes (MacEachern, 2000; De Iorio et al., 2004), with covariate-dependent component parameters but single weights. However, Wade et al. (2014b) demonstrate the lack of robustness due to the fact that the local, cluster-specific predictions are averaged across the covariate space, possibly resulting in poor prediction. Alternatively, numerous proposals have been introduced for covariate-dependent weights exist (Griffin and Steel, 2006; Dunson and Park, 2008; Rodriguez and Dunson, 2011, to name a few). In this work, we build on the interpretable construction of the covariate-dependent weights developed by Antoniano-Villalobos et al. (2014), which allows for combinations of continuous and discrete covariates.

We require extending the model to multivariate responses of mixed type with possible
censoring and constraints. An appealing approach for this relies on a latent Gaussian representation, which provides a simple construction for dependence of the multivariate mixed-type data through the full covariance matrix of the latent Gaussian variables. Moreover, Bayesian inference can be carried out through Gibbs sampling and data augmentation techniques. A Bayesian parametric model based on this idea was proposed by Korsgaard et al. (2003) for multivariate data combining Gaussian, right-censored Gaussian, ordinal, and binary traits. To increase model flexibility, Bayesian nonparametric versions were proposed by De Yoreo and Reiter (2017) for mixed ordinal and nominal data, and by De Yoreo and Kottas (2018) for multivariate ordinal regression. Due to the increased flexibility of nonparametric mixtures, the cut-offs used to define the discrete data from the latent Gaussian variables can be fixed and not estimated or inferred. Moreover, Canale and Dunson (2011) show that Bayesian nonparametric mixtures for discrete data (specifically counts) based on latent Gaussian variables can approximate and consistently estimate a wider range of distributions than mixtures based on discrete distributions, e.g. Poisson or multinomial. Another relevant extension is the Bayesian semiparametric model of Jara et al. (2010) for multivariate doubly-censored data indicating time to event, based on a log transformation linking the observed responses to the latent Gaussian variables. When modeling time-to-event data, the log transformation is more appropriate than others, notably truncation. This allows recovering the underlying structure with fewer and more interpretable components with possibly heavy right tails. Recently, Norets and Pelenis (2018) demonstrated that optimal adaptive estimation of mixed discrete-continuous distributions can be achieved via the latent Gaussian mixture approach.

We combine some of these ideas to build a model which can deal with the challenges presented by the data. We adopt the latent Gaussian approach, associating to each response variable $Z_\ell$ a latent real-valued $Y_\ell$. Specifically, an observed value $z_\ell$ of the response $Z_\ell$ is linked to the realization $y = (y_1, \ldots, y_d)$ of the latent $Y = (Y_1, \ldots, Y_d)$, through a function $h_\ell$ whose characteristics depend on the nature of the observable. Examples of transformations
for different response types include:

\[ z_\ell = h_\ell(y, x) = y_\ell, \quad \text{for } z_\ell \in \mathbb{R}, \]
\[ z_\ell = h_\ell(y, x) = \lfloor \exp(y_\ell) \rfloor, \quad \text{for } z_\ell \in \mathbb{N}, \]
\[ z_\ell = h_\ell(y, x) = \sum_{a=1}^{A_\ell - 1} 1_{[\alpha_{\ell,a}, \infty)}(y_\ell), \quad \text{for } z_\ell \in \{0, 1, 2, \ldots, A_\ell - 1\}, \]

where the last case considers an ordinal response with \( A_\ell \) categories and fixed cutoffs of \( \alpha_{\ell,1} < \ldots < \alpha_{\ell,A_\ell-1} \), and \( 1_B(y) \) denotes the indicator function taking the value one when \( y \in B \). In these examples, the functions \( h_\ell \) do not depend on \( x \) or \( y'_\ell \) for \( \ell' \neq \ell \), but they may, for example when accounting for censored or constrained responses, as is the case for the simulated and case studies described in Sections 4 and 5.

The basic building block for our model is the multivariate multiple linear regression:

\[ Y|x, \beta, \Sigma \overset{\text{ind}}{\sim} N_d(y|x\beta, \Sigma), \]

where \( \beta \) is a \((q+1) \times d\) matrix of regression parameters and \( \Sigma \) is a \(d \times d\) covariance matrix. Slightly abusing notation, \( x = (1, x_1, \ldots, x_q) \) denotes the vector of observed covariate values extended by a unitary entry. As previously discussed, this parametric model is not flexible enough to capture the complex dependence structures contained in the data. We therefore extend the nonparametric density regression framework introduced by Antoniano-Villalobos et al. (2014) to model the \( \mathbb{R}^d \)-valued latent variable \( Y \):

\[ f_{P_X}(y|x) = \sum_{j=1}^{\infty} w_j(x) N_d(y|x\beta_j, \Sigma_j), \quad \text{with } w_j(x) = \frac{w_j g(x|\psi_j)}{\sum_{j'=1}^{\infty} w_{j'} g(x|\psi_{j'})}. \quad (1) \]

This model results from considering a mixture

\[ f_{P_X}(y|x) = \int N_d(y|x\beta, \Sigma) dP_X(\theta), \]

where \( \theta = (\beta, \Sigma) \) and a nonparametric prior is assigned to the set of covariate-dependent
mixing measures $P_x$, which places mass one on the set of discrete probability measures:

$$P_x = \sum_{j=1}^{\infty} w_j(x) \delta_{\theta_j}.$$  

Here, $\delta_{\theta}$ denotes the Dirac-delta function with unit mass at $\theta$. For computational purposes and to ensure convergence of the normalizing constant in $w_j(x)$, it is convenient to adopt a stick-breaking representation for the weights, setting $w_1 = v_1$ and $w_j = v_j \prod_{j' < j}(1 - v_{j'})$, for $j > 1$, where $v_j \sim \text{Beta}(\zeta_{j,1}, \zeta_{j,2})$. The parameters of the local linear regression components, $\theta_j$, and of the covariate-dependent weights, $\psi_j$, are assumed to be independent and identically distributed according to a base measure $P_0$ and independent of the weights. Together with the functions $h_\ell$ linking the latent variables with the responses, this defines the likelihood structure for the observed data.

In this model, the regression parameters $\beta_j$ and $\Sigma_j$ capture the local linear relation between the latent response and covariates, with normal errors; whereas the $\psi_j$ determine, through $g$, how the influence of each local component to the overall model changes across the covariate space. This deals with situations when the stochastic relation between $y$ and $x$ is too complicated to be captured by a single parametric model. It can also be used when the population is assumed to be constituted by an unknown number of (covariate-dependent) groups such that, within each group, a linear regression model provides a good description of the data. While identifiability issues may prevent the individuation of such groups, this intuition can help in understanding the elements of the model.

Note that the Bayesian nonparametric model for the joint density of $y$ and $x$ introduced by Müller et al. (1996) for density regression, taking the form

$$f_{P}(y, x) = \sum_{j=1}^{\infty} w_j g(x|\psi_j) N_d(y|x\beta, \Sigma), \text{ with } P = \sum_{j=1}^{\infty} w_j \delta_{(\theta_j, \psi_j)},$$  

results in a conditional density coinciding with equation (1). However, an important difference is that in the joint mixture model, posterior inference for the parameters $(w_j, \theta_j, \psi_j)$ is based on the joint likelihood in (2); whereas, for our model, it is based directly on the conditional likelihood of interest. As stated by Müller and Quintana (2004, pp. 101–102), the joint model-
ing approach “wrongly introduces an additional factor” for the marginal of \( x \) in the likelihood “and thus provides only approximate inference”. Indeed, as shown by Wade et al. (2014a), when including this additional factor, extra components are required to fit the marginal of \( x \), which can degrade the performance of the conditional density estimate. Instead, since posterior inference is based only on the conditional likelihood, the model developed here is able to overcome this problem, but it still maintains the same natural and interpretable structure for the weights of the joint mixture model. Furthermore, we emphasize that the converse is not true: our conditional density model in (1) does not imply the joint density model in (2). This can be easily seen by constructing a joint density model as the product of (1) and any, say parametric, marginal density model for \( x \). This is a valid construction, which nonetheless recovers the joint model in (2) only when the marginal has the form:

\[
    f_P(x) = \sum_{j=1}^{\infty} w_j g(x|\psi_j).
\]

This is an important concept, as it highlights that the form chosen for \( g \) does not imply a modeling of the distribution for covariates, which may indeed be fixed. The choice and shape of this kernel, however, defines how the conditional distribution changes as \( x \) varies (given the parameters \( \psi \)). Thus, it determines the amount of information borrowed when making inference at unobserved points in the space of covariates.

The covariate-dependent weight \( w_j(x) \) represents the probability that an observation with a covariate value \( x \) is allocated to the \( j \)-th regression component. Such probability can be decomposed into the unconditional probability \( w_j \) that the parametric model \( j \) fits an individual observation, and the likelihood \( g(x|\psi_j) \) that an individual allocated to the \( j \)-th component is characterized by a covariate value \( x \). The \( g(\cdot|\psi) \) can be defined to accommodate different types of covariates. We adopt a factorizable structure:

\[
    g(x|\psi) = \prod_{k=1}^{q} g(x_k|\psi_k), \quad \text{where} \quad g(x_k|\psi_k) = \begin{cases} 
    N(x_k|\mu_k, \tau_k^{-1}) & \text{for } k = 1, \ldots, p, \\
    \text{Bern}(x_k|\rho_k) & \text{for } k = p + 1, \ldots, q,
\end{cases}
\]

with \( \psi_k = (\mu_k, \tau_k) \) for \( k = 1, \ldots, p \), and \( \psi_k = \rho_k \) for \( k = p + 1, \ldots, q \). The use of distribution kernels guarantees convergence, for all \( x \), of the denominator in equation (1). For the unco-
ditional probability \( w_j \), different choices of the stick-breaking parameters \((\zeta_{j,1}, \zeta_{j,2})\) result in different nonparametric priors (see Ishwaran and James, 2001). For instance, if \((\zeta_{j,1}, \zeta_{j,2}) = (1, \zeta)\), the prior on the weights \( w_j \) corresponds to that obtained from a Dirichlet process prior. The base measure is chosen as \( P_0(\beta, \Sigma, \psi) = P_0(\beta|\Sigma)P_0(\Sigma)P_0(\mu|\tau)P_0(\tau)P_0(\rho) \).

We use the conjugate matrix-variate Normal-Inverse Wishart for the regression parameters: \( P_0(\beta|\Sigma) = MN_{(q+1) \times d}(\beta_0, U, \Sigma) \), where \( \beta_0 \) is a \((q+1) \times d\) matrix and \( U \) is a \((q+1) \times (q+1)\) positive definite matrix; \( P_0(\Sigma) = IW(\Sigma_0, \nu) \), where \( \Sigma_0 \) is a \( d \times d \) positive definite matrix and \( \nu > 0 \). Notice that the Inverse Wishart assigns prior mass to full covariance matrices. Other prior specifications can be used to allow for other types of covariance structures, e.g. product of Inverse Gammas for diagonal covariance matrices or G-Wishart for sparse precision matrices. As for the \( \beta \) coefficients, we are assuming a structured dependence, allowing for efficient computations through Kronecker products and a reduced number of hyperparameters compared to a full Gaussian distribution. Alternatively, a multivariate Gaussian distribution could be used, assuming independence between columns. To complete the specification of the base measure, we set: \( P_0(\mu|\tau) = \prod_{k=1}^{p} N(\mu_k|\mu_{0,k}, (u_k \cdot \tau_k)^{-1}) \), \( P_0(\tau) = \prod_{k=1}^{p} \text{Gamma}(\tau_k|\alpha_k, \gamma_k) \), and \( P_0(\rho) = \prod_{k=p+1}^{q} \text{Beta}(\rho_k| \varrho_k) \), where \( \varrho_k = (\varrho_{k,1}, \varrho_{k,2}) \).

In the next section, we describe an adaptive truncation algorithm allowing posterior inference for our model. The algorithm is general and only requires specific adjustments depending on the \( h_{\ell} \) functions linking the observed responses with their latent counterparts.

3 Adaptive Truncation Algorithm

To scale appropriately with the sample size and data dimensions, we devise an algorithm for posterior inference based on a finite truncation of the mixture, where the number of components is allowed to increase adaptively to obtain a good approximation of the infinite-dimensional posterior. The truncated latent model with \( J \) components is:

\[
 f_{P_J}(y|x) = \sum_{j=1}^{J} w_j(x)N_d(y|x\beta_j, \Sigma_j). \] (3)
where the weights follow the re-normalized stick breaking construction:

\[
w^j_j(x) = \frac{w_j g(x|\psi_j)}{\sum_{j=1}^J w_j} = \frac{w_j g(x|\psi_j)}{\sum_{j'=1}^J w_{j'} g(x|\psi_{j'})}.
\] (4)

Notice that the normalizing constant \(\sum_{j=1}^J w_j\) in (4) cancels out. To ease notation, we use \(w_j(x)\) to denote the truncated covariate-dependent weights, dropping the superscript \(J\) when the truncation level is clear. Due to the exponential decay of the weights, for large enough \(J\), the truncated model (3) provides a close approximation to the infinite mixture model.

Alternative truncation methods could be considered, notably the popular truncated stick breaking method (Ishwaran and James, 2001) where \(v_J = 1\). However, re-normalized stick-breaking may provide a better finite-dimensional approximation by evenly distributing the remaining mass across components, as opposed to assigning all remaining mass to the last component in truncated stick-breaking.

The proposed algorithm is based on the adaptive truncation scheme developed by Griffin (2016), extended for density regression and mixed type responses. It consists of two main steps, namely an MCMC step for a fixed truncation level, \(J_0\), followed by a sequential Monte Carlo (SMC) step used to increase the number of components of the mixture. The first step produces \(M\) posterior draws \((w^m_{1:J_0}, \theta^m_{1:J_0}, \psi^m_{1:J_0}, y^m_{1:n})_{m=1}^M\), which are then used as particles in the SMC step. We provide a concise summary below, with software and full details provided through the authors’ GitHub repository and accompanying documentation (https://github.com/sarawade/BNPDensityRegression_AdaptiveTruncation).

**MCMC for fixed truncation.** Since the truncation level \(J_0\) is fixed throughout this step, we omit it from the notation, writing \(w = w_{1:J_0}, \theta = \theta_{1:J_0},\) and \(\psi = \psi_{1:J_0}\). Similarly, the observed response is denoted by \(z = (z_1, \ldots, z_n)\), with \(z_i = (z_{i,1}, \ldots, z_{i,d})\), and analogously for the covariates \(x\) and the latent \(y\). The approximate posterior given the sample \((x, z)\) of size \(n\), using the truncated likelihood (3), takes the form:

\[
P_{J_0}^n(w, \psi, \theta, y|z, x) \propto P_{J_0}(w, \psi, \theta) \prod_{i=1}^n \sum_{j=1}^{J_0} w_j(x_i|\psi_j)N_d(y_i|x_i\beta_j, \Sigma_j) \prod_{\ell=1}^d 1_{\{z_{i,\ell}\}}(h_{i,\ell}),
\]
where $P_{J_0}(w, \psi, \theta)$ indicates the restriction of the prior (as detailed in Section 2) to the parameters in the truncated space. Dependence $w_j(x) = w_j(x|\psi_j)$ of the weights on the parameters has been made explicit. Moreover, the functions $h_{i,\ell} = h_{\ell}(y_i, x_i)$ linking the latent variables to the observed responses are specifically defined for the simulated and case studies in Sections 4 and 5.

Due to lack of conjugacy, we resort to a generic Metropolis-within-Gibbs scheme to perform posterior sampling, that updates blocks of parameters adaptively. The adaptive random walk algorithm used here, based on Algorithm 6 in Griffin and Stephens (2013), adapts the proposal covariance matrix to achieve both a specified average acceptance rate ($a_0 = 0.234$) and a proposal covariance matrix equal to $2.4^2/p$ times the posterior covariance matrix, $p$ being the dimension of the parameter block of interest. These criteria have been shown to be optimal in many settings (Roberts et al., 1997; Roberts and Rosenthal, 2001). In more detail, suppose that we want to sample a block of parameters $\phi$ of dimension $p$ from a distribution with probability density function $Q$. In order to utilise the adaptive random walk algorithm, we first consider a transformation $t(\phi)$ that has full support on $\mathbb{R}^p$. At each iteration $m$, we propose a new $\phi^*$ such that:

$$t^* \equiv t(\phi^*) = t(\phi^{m-1}) + \epsilon, \quad \text{with} \quad \epsilon \sim N(0, \xi^{m-1}),$$

(5)

where $\xi^{m-1}$ is the adaptive covariance matrix. We accept $\phi^m = \phi^*$ with probability equal to the minimum between 1 and the ratio:

$$a(\phi^*, \phi^{m-1}) = \frac{Q(\phi^*)}{Q(\phi^{m-1})} \frac{|J_t(\phi^{m-1})|}{|J_t(\phi^*)|},$$

(6)

with $|J_t(\phi)|$ denoting the determinant of the Jacobian of the transformation.

Transformations of $\beta_j, \mu_j, \tau_j, \rho_j$, and $v_j$ are straightforward through identity, log, and logit functions. Instead, transformations of $\Sigma_j$ and $y_i$ are more involved. For each $\Sigma_j$, we consider a vectorization of a decomposition of the matrix, $\Sigma_j = L_j D_j L_j^\top$, where $L_j$ is a lower triangular matrix with unit entries on the diagonal and $D_j$ is a diagonal matrix with positive entries, and we take the log of the diagonal entries. In this case, the proposed $\Sigma_j^*$ can be obtained from the proposed $t^*$ in equation (5) by inverting this transformation. In addition, the determinant of
the Jacobian, which is required in the acceptance ratio in (6), depends only on the diagonal elements $D_{j,\ell,\ell}$ of the matrix $D_j$, specifically, $|J_t(\Sigma_j)| = \prod_{\ell=1}^{d} 1 / D_{j,\ell,\ell}^{d+1-\ell}$. For each latent vector $y_i = (y_{i,1}, \ldots, y_{i,d})$, the terms $h_{\ell}(y_i, x_i) = z_{i,\ell}$ define constrained regions for the latent $y_i$, such that $y_{i,\ell} \in (l_{i,\ell}, u_{i,\ell})$, which are provided for the simulated and case studies in Sections 4 and 5.

We assume that an appropriate ordering of the responses leads to bounds $(l_{i,\ell}, u_{i,\ell})$ that may in general depend on $y_{i,\ell'}$ for $\ell' < \ell$. This allows us to define a sequential logistic transformation $t(y_{i,\ell}; y_{i,1:\ell-1})$ for $\ell = 1, \ldots, d$, based on the bounds $(l_{i,\ell}, u_{i,\ell})$. From the proposed $t^*$ in equation (3), the inverse transformation can be applied to obtain the proposed $y_i^*$, sequentially for $\ell = 1, \ldots, d$, where the bounds may also be updated sequentially if they depend on $y_{1:(\ell-1)}^*$, e.g. for age at first child in our application. This ordering also guarantees that the Jacobian matrix is lower triangular, so its determinant is simply the product of the diagonal elements,

$$|J_t(y_i)| = \prod_{\ell=1}^{d} J_{t,\ell,\ell}(y_{i,\ell}; y_{i,1:\ell-1}) = (u_{i,\ell} - l_{i,\ell}) / [(y_{i,\ell} - l_{i,\ell})(u_{i,\ell} - y_{i,\ell})],$$

for $u_{i,\ell} \in \mathbb{R}, l_{i,\ell} \in \mathbb{R}$.

**SMC for adaptive truncation.** The second stage involves the selection of the truncation level $J$ by sequentially increasing it from the initial level $J_0$. The addition of a new component improves the quality of the approximation to the infinite-dimensional model but increases the computational burden, due to the considerable number of parameters added. Therefore, devising an algorithm that can select the level of truncation parsimoniously is crucial. To achieve this, possible approaches are presented in Norets (2017) and Griffin (2016). We focus on the latter, which adaptively increases the number of mixture components via SMC.

The MCMC draws from the previous step are used as the $M$ initial particles in the SMC. At each iteration of the SMC, a new component is added to the mixture, by sampling the additional set of parameters $(w_{J+1}, \psi_{J+1}^m, \theta_{J+1}^m)$ from a suitable importance distribution. We sample from the prior $\text{Beta}(v_{J+1}^m)P_0(\psi_{J+1}^m, \theta_{J+1}^m)$, independently for $m = 1, \ldots, M$, making use of the recursive stick-breaking relation $w_{J+1}^m = v_{J+1}^m [ (1 - v_{J+1}^m) / v_{J+1}^m ] w_{J}^m$. The particle weights $\tilde{\vartheta}_{J+1}^1 : M = (\tilde{\vartheta}_{J+1}^1, \ldots, \tilde{\vartheta}_{J+1}^M)$ are then updated as follows:

$$\tilde{\vartheta}_{J+1}^m = \tilde{\vartheta}_{J+1}^m \prod_{i=1}^{n} \frac{f_{P_{j_{i}^{J+1}}^{m}}(y_{i}^m|w_{1:j+1}^m, \psi_{1:j+1}^m, \theta_{1:j+1}^m)}{f_{P_{j_{i}^{J}}^{m}}(y_{i}^m|w_{1:j}^m, \psi_{1:j}^m, \theta_{1:j}^m)}.$$
When the effective sample size (ESS) is lower than a threshold, indicating poor mixing, the particle values are resampled according to such weights (Del Moral et al., 2006). Here, we resort to systematic resampling (Kitagawa, 1996) and perform a rejuvenating step (Gilks and Berzuini, 2001), where the particles are replaced with new values sampled through $m^*$ iterations of the adaptive MCMC with $J_0 = J + 1$. The SMC provides weighted samples from the sequence of truncated posteriors $P^n_j$, converging to the infinite posterior $P^n$. To decide when a sufficiently accurate approximation has been obtained, we follow Griffin (2016) and stop at the truncation level $J^*$, such that the discrepancy $D(P^n_j, P^n_{j+1}) = |ESS_j - ESS_{j+1}|$ is less than a specified $\delta > 0$, for a fixed number $I$ of consecutive increments, $J = J^* - I + 1, \ldots, J^*$. We use the suggested values of $\delta = 0.01M$, $I = 4$, and $m^* = 3$. As an alternative to the ESS, we also consider a discrepancy based on the conditional effective sample size (CESS), which was proposed by Zhou et al. (2016), in the context of model comparison via SMC.

4 Simulation Study

We assess the performance of the proposed procedure on a simulated dataset with known structure. We consider $q^* = 3$ covariates; the first, denoted as $x_1$, is continuous and observed at a discrete scale (resembling the age at interview in our case study), while the remaining, denoted as $(x_2^*, x_3^*)$, are categorical with three and two levels, respectively. We generate two positive integer-valued responses and one binary response. The first response $Z_1$ is a discretized noisy observation of a nonlinear function of $x_1$. Similarly, $Z_2$ is a discretized noisy observation of a nonlinear function of $x_1$ and the realized $z_1$. In both cases, 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. Censoring is defined before discretization, when the responses are greater than the first covariate. The true curves and densities are depicted in Figure 1 (top row) for selected combinations of the covariates. Finally, a binary response is simulated from a linear probit model depending only on $x_1$ (Figure 2). Complete details of the data-generating distributions are provided in the Supplementary Material (SM).

We seek to recover the conditional distribution of the response variables given the covariates.
Figure 1: Simulation study. 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.

using our proposed model, from a sample of size $n = 700$. We define the link functions $h_\ell(y, x)$ as:

$$z_\ell = h_\ell(y, x) = c_\ell(y, x)[\exp(y_\ell)], \quad \text{for } \ell = 1, 2,$$

$$z_3 = h_3(y, x) = 1_{[0, \infty)}(y_3),$$

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

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

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

Specification of the prior parameters is detailed in the SM. 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. In the SM, we describe various posterior and predictive quantities that can be computed from the weighted particles to describe the relationship between the observed response $z$ and covariates $x$. 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 1 and 2, for a selected combinations of the categorical covariates. Overall, the model is able to recover the latent structure present in the data, despite the heavy censoring of $Z_2$ for lower levels of $x_1$, particularly when $x_3^* = 2$.

To provide further insight on the algorithm and model performance, we carry out a robustness analysis on the number of initial components $J_0$. Table 1 summarizes results regarding: the number of components inferred by the model ($J^*$); elapsed CPU time (in hours); and for each $Z_\ell$, the log-pseudo marginal likelihood (LPML, Geisser and Eddy, 1979) and percentage absolute errors with respect to the true mean and true density, denoted by ERR_{Mean} and ERR_{Dens} respectively; expressions for these quantities can be found in SM. Additionally, to compare mixing, we report the ESS of the log-likelihood for the MCMC stage (ESS_{MCMC}), and the ESS_{J*} of the final iteration of the SMC. 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 $h_\ell(y, x)$ in (7) and a prior given by the base measure $P_0$. For the sake of comparison, we use the Metropolis-within-Gibbs

![Figure 2: Simulation study. 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.](image)
Table 1: Simulation study. 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 $J_0$.

| $J_0$ | $J^*$ | CPU | ESS$_{MCMC}$ | ESS$_{J^*}$ | LPML $\times 10^3$ | ERR$_{Mean}$ | ERR$_{Dens}$ |
|-------|-------|-----|-------------|-------------|----------------|--------------|--------------|
|       |       |     |             |             | $Z_1$ | $Z_2$ | $Z_3$ | $Z_1$ | $Z_2$ | $Z_3$ | $Z_1$ | $Z_2$ | $Z_3$ |
| 1     | 1     | 0.66 | 533.8       | -1.17       | -0.82 | -0.34 | 4.54 | 2.40 | 5.06 | 328.66 | 41.11 | 5.69 |
| 2     | 13    | 1.90 | 195.7       | 1125.5      | -1.01 | -0.76 | -0.34 | 3.05 | 3.88 | 6.75 | 152.89 | 49.26 | 7.17 |
| 5     | 14    | 3.93 | 192.7       | 1966.7      | -0.91 | -0.71 | -0.34 | 2.53 | 3.92 | 5.23 | 129.03 | 44.28 | 5.49 |
| 10    | 19    | 5.29 | 202.6       | 1918.9      | -0.94 | -0.71 | -0.34 | 1.89 | 2.62 | 6.85 | 80.36 | 43.60 | 7.13 |
| 15    | 26    | 5.77 | 211.4       | 1266.0      | -0.93 | -0.73 | -0.35 | 2.28 | 2.84 | 6.89 | 90.57 | 47.20 | 7.04 |
| 20    | 24    | 5.43 | 205.3       | 1990.3      | -0.86 | -0.69 | -0.34 | 1.98 | 2.88 | 6.58 | 83.87 | 41.37 | 7.30 |
| 30    | 34    | 9.04 | 223.7       | 2000.0      | -0.85 | -0.69 | -0.34 | 2.10 | 3.11 | 6.56 | 86.82 | 41.65 | 6.96 |

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 is increasing with $J_0$, although this is not always the case, especially 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 (reported in the SM) 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.
5 Application: Life Patterns of Colombian Women

Our aim is to understand the relationship between life patterns and socio-demographic background for women in Colombia. To describe the characteristics of the fertility and partnership patterns, we consider the discrete variables recording the ages at *Sexual Debut* ($Z_1$), at *Union* ($Z_2$), referring to the first marriage or cohabitation, and at *First Child* ($Z_3$). The *Work Status* ($Z_4$) of the women is recorded as a binary variable indicating whether the respondent worked in the 12 months before the interview. The baseline socio-demographic factors considered are *Age* at interview ($X_1$); *Region* ($X_2$) indicating residence in Atlantica, Oriental, Central, Pacifica, Bogota, or Territorios Nacionales as defined in Ojeda et al. (2011); whether the respondent lives in an urban or rural *Area* ($X_3$); having (P) or not having (P) been disciplined using *Physical Punishment* ($X_4$) during childhood; and having (B) or not having (B) observed *Parental Domestic Violence* ($X_5$), referring to whether the respondent witnessed her father beating her mother. Our final dataset consists of $n = 10,740$ women, with full details on data filtering and descriptions provided in the SM.

To complete our model specification, we define the link functions:

$$z_\ell = h_\ell(y, x) = c_\ell(y, x)|\exp(y_\ell)|, \text{ for } \ell = 1, 2,$$

$$z_3 = h_3(y, x) = c_3(y, x)|\exp(y_1) + \exp(y_3)|,$$

$$z_4 = h_4(y_1, x) = 1_{[0,\infty)}(y_4),$$

with $c_\ell(y, x) = 1_{(0,x_1+1)}(\exp(y_\ell))$, for $\ell = 1, 2$, and $c_3(y, x) = 1_{(0,x_1+1)}(\exp(y_1) + \exp(y_3))$. In this case, $\exp(y_1)$ and $\exp(y_2)$ can be interpreted as the latent continuous ages at sexual debut and union, respectively. The constraint that age at first child must be greater than age at sexual debut is strictly enforced through the transformation, and we can interpret $\exp(y_3)$ as the latent continuous time between sexual debut and first child and $\exp(y_1) + \exp(y_3)$ as the latent continuous age at first child. The bounds required in the adaptive MCMC are obtained
from inverting $z_\ell = h_\ell(y, x)$; concretely,

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

$$
(l_3, u_3) = \begin{cases} 
\left( \log(\max[0, x_{i,1} + 1 - \exp(y_{i,1})]), \infty \right) & \text{for censored } z_3 = 0 \\
\left( \log(\max[0, z_{i,\ell} - \exp(y_{i,1})]), \log(z_{i,\ell} - \exp(y_{i,1}) + 1) \right) & \text{for uncensored } z_3 \neq 0 
\end{cases},
$$

$$
(l_4, u_4) = \begin{cases} 
(-\infty, 0) & \text{for } z_4 = 0 \\
(0, \infty) & \text{for } z_4 = 1 
\end{cases}.
$$

Prior specification is provided in the SM. We initialize the MCMC algorithm with a number of components, $J_0 = 35$, large enough to avoid a small ESS and subsequent resampling (interested readers are referred to the SM). The MCMC is run for 20,000 iterations after discarding the first 30,000 as burn-in, and one in every 10 iterations is saved to produce 2,000 particles. For the SMC, we choose the ESS-based stopping rule, due to the robustness observed in the simulation study. For the sake of conciseness, we display only a selection of predictive quantities, which offer some insights about the situation of Colombian women. Specifically, we compare women who were raised in violent family environments $(P, B)$ with those who were not $(\bar{P}, \bar{B})$. Figure 3 displays the predictive medians of the (undiscretized) ages at events and the posterior probability of working as functions of $Age$. More detailed information arises from the analysis of the predictive densities, some of which are reported in Figure 4. Notice that due to the clear asymmetry in the densities, the predictive median allows a better representation of the center, as opposed to the mean.
Figure 3: Predictive medians of the ages at sexual debut, union and first child, and posterior probability of working, as functions of Age, for women who grew up in violent (P, B) and non-violent families (P̅, B̅). Dotted lines indicate when the median exceeds Age.

The data presents heavy censoring for younger cohorts (summarized in Table 2). This information is included by imputing, at each iteration of the algorithm, ages at events which must be higher than Age. Indeed, above the dashed lines of Figure 4, the density estimates are based on these imputed ages and borrowing of information at other covariate levels. Therefore, while we can reliably estimate the mass above the dashed line given Age, caution should be used when interpreting the shape of the right tail in this region, as this is not identifiable from the observed data. Moreover, when such mass exceeds 0.5, the predictive median is
Figure 4: Predictive densities of the ages at union and first child as functions of Age for women who grew up in violent (P, B) and non-violent families (P̅, B̅). 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.

affected by the imputed values and is therefore less reliable. This corresponds to median values of age at event which are higher than Age, represented as dotted lines in the figures. Further, censored data also arise from women who will never experience an event. This is the prevailing cause of censoring for the older cohorts, contributing to higher medians and heavier right tails. Accommodating censored cases is clearly useful; however, results arising from heavily censored data should be interpreted with caution.

Starting with Figure 3, observe that the shapes of the median curves change across combinations of the categorical covariates, which justifies the employment of a flexible model
that does not impose a single functional form. A clear difference is evident between urban and rural areas, the latter presenting lower ages at events, controlling for other covariates. This is expected since rural areas are generally characterized by lower levels of education and wealth indicators, both identified in the literature as factors related to anticipation of sexual activity and family formation. Comparing cohorts, we observe that younger women tend to anticipate sexual debut, a phenomenon largely recognized as a consequence of the better knowledge and the more diffuse use of contraceptive methods. Instead, the curves for the ages at union and at first child appear flatter, particularly for urban women with non-violent family environments and are even increasing for women from violent families. At first, this may seem counter-intuitive, because one would expect the younger generations to postpone family formation, particularly in urban areas, due to an expected prolonged education. However, an incorrect use of contraceptive methods, particularly among very young or less educated women, and the violent conditions linked to the armed conflict may result in early pregnancies [Ali et al., 2003; Núñez and Flórez, 2001; Daniels, 2015]. Indeed, an increase in teenage childbearing in Colombia has been observed since 1990, mainly among women from disadvantaged backgrounds (Batyra, 2016; Flórez and Soto, 2007, 2013).

Focusing on the predictive densities for the least and the most developed regions, Territorios Nacionales and the capital city Bogota (Figure 4), further justifies the use of a density regression model. In fact, the observed flat median curves correspond to rather different distributional behaviors of ages at union and child, across covariate values. Moving from the least to the most developed context (top to bottom in the figure) entails an increase of the median curves, dispersion, and probabilities of not having experienced the events by a given Age. An increased dispersion, with pronounced right-skewness, is more evident for older cohorts in urban environments. This is in line with the greater heterogeneity in urban contexts as well as with the wider range of opportunities offered, for example in terms of education. Such heterogeneity becomes more pronounced among the older cohorts who have had time to profit from such opportunities. The flexibility gained in urban contexts is offset in violent environments, thus resulting in more concentrated distributions. While our definition of a violent environment is not formal and refers only to the adoption of physical punishment methods and exposure to parental violence, the results signal the detrimental effect of family violence
on Colombian women life patterns, and provides quantitative support for psychosocial studies (Restrepo Martínez et al., 2017).

The joint modeling approach permits us to study also the conditional relation between responses. For example, Figure 5 shows the conditional predictive medians of the time from sexual debut to union given the age at sexual debut for women with Age = 20, 30, 40 (dotted lines indicate predicted ages at event higher than Age; the corresponding conditional densities are reported in the SM). Interesting differences can be observed across regions, likely related to their socio-demographic characteristics (Ojeda et al., 2011). We observe that women in Atlantica and Territorios Nacionales (and to a lesser extent Oriental) compared with Pacifica...
Figure 6: Predictive probability of working as function of Age conditional on different ages at first child, for women who grew up in violent $(P, B)$ and non-violent families $(\bar{P}, \bar{B})$. Dotted lines indicate ages at event higher than Age.

and Bogota tend to experience sexual debut and union closer in time, suggesting that sexual debut is possibly delayed until union. Such tendency is more pronounced, compared to the other regions, for rural women raised in violent families. Similar results are observed for the time from sexual debut to child (details in the SM).

Finally, the probability of working (Figure 3, bottom row) is, as expected, higher in urban areas. Moreover, women who grew up in violent environments show a higher propensity to work, more pronounced among younger women. These same women, as previously observed, show a tendency to anticipate events. A possible explanation is that young women who leave the parental house to escape violence may start cohabitation and decide to drop out of school,
entering the labor market to contribute to family income. This apparently contradicts studies (see e.g. Gimenez Duarte et al., 2015) pointing to the difficulties of young women, especially those with children, to participate in the labor market. However, this paradox is solved when analyzing the estimated predictive probabilities of working as functions of Age, conditional on having the first child at ages 15, 20 and 25 (Figure 6 top to bottom). Indeed, the probability of working at each Age increases with the age at first child. In particular, we observe a much lower probability of working for young mothers, that persists even when considering their labor market participation later in life. This suggests a scaring effect of teenage motherhood.

6 Concluding Remarks

In this work, we proposed a novel Bayesian nonparametric model for density regression, allowing for mixed responses with censored, constrained, and binary traits, that can flexibly change with combinations of the categorical and numerical covariates. We developed a general algorithm for posterior inference, that effectively scales to large datasets by adaptively determining the necessary truncation level to approximate the infinite-dimensional posterior. We customized the model and algorithm to a specific case study, but they can be applied in other contexts through minor modifications, by appropriate definition of the link functions. From a technical point of view, our results highlight the advantage of a flexible model, accounting for different shape, location, and dispersion of the response distribution across the covariate levels, as well as for censoring. Additionally, a variety of classic graphic tools and quantities of interest, such as survival curves and hazard functions, can be derived. Importantly, the joint analysis of the responses allows for a rich variety of conditional analyses, which can be conducted focusing on different aspects, a very useful feature when studying complex phenomena.

For our case study, the findings suggest interesting considerations regarding life patterns of Colombian women. In the first instance, we found a confirmation of the differences between rural and urban areas, which evidence the need of interventions towards a more balanced development of the country. Furthermore, our results signal that the regions with a higher risk of early transition to adulthood are those with the worse development and wellness indicators,
thus corroborating studies on the risks related to disadvantageous conditions. One of the most interesting results is the rather clear evidence of the impact of family violence on women’s choices and behaviors. An anticipation of the considered events is observed for women who were physically punished during childhood and witnessed parental domestic violence, two factors we used as proxies for a violent family environment. The relation between child abuse and neglect and the child’s future family choices has been discussed in the literature. Nonetheless, to our knowledge, this is the first attempt to study the possible relation between parental family violence and the events marking the transition to adulthood. Our findings confirm that a violent family environment can be regarded as a key risk factor that may nullify the positive influence of developed areas.

Overall, our case study may contribute to the planning of targeted interventions. Even if recent governments have shown an increased attention to the conditions of women and children, a formal statistical approach to systematically identify and quantify critical situations is crucial to support such a process. For example, teenage pregnancy is recognized as a priority issue in Colombia by the Government (Gimenez Duarte et al., 2015; Daniels, 2015), due to its hindering personal development and agency (Azevedo et al., 2012); our results confirm its scaring effect and quantify the risk of teenage pregnancy, identifying some of the most vulnerable groups. We conclude with the hope that the present work may stimulate further reflection, research and survey on the topic, and possibly lead to additional investigations exploiting the availability of DHS surveys on other developing countries and the flexibility and wide applicability of our model.

**Acknowledgements**

The work reported in this paper was funded by the University of Warwick Academic Returners Fellowship and the University of Oslo.
Note

The code will be publicly released after acceptance:
https://github.com/sarawade/BNPDensityRegression_AdaptiveTruncation,
along with the simulated data to reproduce results.

References

M.M. Ali, J. Cleland, and I.H. Shah. Trends in reproductive behaviour among young single women in Colombia and Peru: 1985-1999. *Demography*, 40:659–673, 2003.

I. Antoniano-Villalobos, S. Wade, and S.G. Walker. A Bayesian nonparametric regression model with normalized weights: A study of hippocampal atrophy in Alzheimer’s disease. *Journal of the American Statistical Association*, 109(506):477–490, 2014.

J.P. Azevedo, M. Favara, S.E. Haddock, L.F. Lopez-Calva, M. Müller, and E. Perova. *Teenage pregnancy and opportunities in Latin America and the Caribbean*. Washington, D.C.: World Bank Group, 2012.

E. Batyra. Fertility and the changing pattern of the timing of childbearing in Colombia. *Demographic Research*, 35:1343–1372, 2016.

A. Canale and D.B. Dunson. Bayesian kernel mixtures for counts. *Journal of the American Statistical Association*, 106:1528–1539, 2011.

J. P. Daniels. Tackling teenage pregnancy in colombia. *The Lancet*, 385(9977), 2015.

M. De Iorio, P. Müller, G.L. Rosner, and S.N. MacEachern. An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, 99:2205–215, 2004.

M. De Yoreo and A. Kottas. Bayesian nonparametric modeling for multivariate ordinal regression. *Journal of Computational and Graphical Statistics*, 27:71–84, 2018.

M. De Yoreo and D.S. Reiter, J.P.and Hillygus. Bayesian mixture models with focused clustering for mixed ordinal and nominal data. *Bayesian Analysis*, 12:679–703, 2017.
P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

D.B. Dunson and J.H. Park. Kernel stick-breaking processes. *Biometrika*, 95:307–323, 2008.

C.E. Flórez and V.E. Soto. Fecundidad adolescente y desigualdad en Colombia. *Notas de Población*, 83:41–74, 2007.

C.E. Flórez and V.E. Soto. Factores protectores y de riesgo del embarazo adolescente en Colombia. Estudios a profundidad 2010. Encuesta nacional de demografía y salud (ENDS 1990/2010), Bogotá: Profamilia, 2013.

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

J.K. Ghosh and R.V. Ramamoorthi. *Bayesian Nonparametrics*. Springer-Verlag New York, 2003.

W.R. Gilks and C. Berzuiini. Following a moving target-Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(1):127–146, 2001.

L.R. Gimenez Duarte, S.H. Van Wie, M. Muller, R.F. Schutte, M.Z. Roumseville, and M.C. Viveros Mendoza. Enhancing youth skills and economic opportunities to reduce teenage pregnancy in Colombia (English). World Bank Report 97822-CO, Washington, D.C.: World Bank Group, 2015.

J.E. Griffin. An adaptive truncation method for inference in Bayesian nonparametric models. *Statistics and Computing*, 26:423–441, 2016.

J.E. Griffin and M. Steel. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 10:179–194, 2006.

J.E. Griffin and D.A. Stephens. Advances in Markov chain Monte Carlo. In *Bayesian Theory and Applications*. Oxford University Press, 2013.
T. Hanson and W. O Johnson. A bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics*, 13(2):341–361, 2004.

H. Ishwaran and L.F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96:161–173, 2001.

A. Jara, E. Lesaffre, M. De Iorio, and F. Quintana. Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Annals of Applied Statistics*, 4(4):2126–2149, 12 2010.

G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

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

A. Kottas and A. E Gelfand. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456):1458–1468, 2001.

A.Y. Lo. On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics*, pages 351–357, 1984.

S.N. MacEachern. Dependent nonparametric processes. In ASA *Proceedings of the Section on Bayesian Statistical Science*, pages 50–55, Alexandria, VA, 1999. American Statistical Association.

S.N. MacEachern. Dependent Dirichlet processes. Technical report, Department of Statistics, Ohio State University, 2000.

P. Müller and F.A. Quintana. Nonparametric Bayesian data analysis. *Statistical Science*, 19: 95–110, 2004.

P. Müller, A. Erkanli, and M. West. Bayesian curve fitting using multivariate normal mixtures. *Biometrika*, 88:67–79, 1996.
A. Norets. Optimal auxiliary priors and reversible jump proposals for a class of variable
dimension models. Unpublished manuscript, Brown University, 2017.

A. Norets and J. Pelenis. Adaptive Bayesian estimation of mixed discrete-continuous distrib-
utions under smoothness and sparsity. arXiv preprint arXiv:1806.07484, 2018.

J. Núñez and C.E. Flórez. Teenage childbearing in Latin American countries. IDB Working
Paper No. 147, Inter-American Development Bank, Research Department, August 2001.
Available at http://dx.doi.org/10.2139/ssrn.1814694.

G. Ojeda, M. Ordóñez, and L. H. Ochoa. Colombia Encuesta Nacional de De-
mográﬁa y Salud 2010. Bogotá, Colombia: Proﬁmila, 2011. Available at
http://dhsprogram.com/pubs/pdf/FR246/FR246.pdf.

M. Restrepo Martínez, L. Trujillo Numa, D. Restrepo Bernal, Y. Torres de Galvis, and
G. Sierra. Sexual abuse and neglect situations as risk factors for adolescent pregnancy.
Revista Colombiana de Psiquiatría (English ed.), 46(2):74–81, 2017.

G.O. Roberts and J.S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms.
Statistical Science, 16:351–367, 2001.

G.O. Roberts, A. Gelman, and W.R. Gilks. Weak convergence and optimal scaling of random
walk Metropolis algorithms. Annals of Applied Probability, 7:110–120, 1997.

A. Rodriguez and D.B. Dunson. Nonparametric Bayesian models through probit stick-
breaking processes. Bayesian Analysis, 6:145–178, 2011.

S. Wade, D.B. Dunson, S. Petrone, and L. Trippa. Improving prediction from Dirichlet process
mixtures via enrichment. Journal of Machine Learning Research, 15:1041–1071, 2014a.

S. Wade, S. G. Walker, and S. Petrone. A predictive study of Dirichlet process mixture models
for curve fitting. Scandinavian Journal of Statistics, 41(3):580–605, 2014b.

Y. Zhou, A.M. Johansen, and J.A.D. Aston. Toward automatic model comparison: an adap-
tive sequential Monte Carlo approach. Journal of Computational and Graphical Statistics,
25(3):701–726, 2016.
Supplementary Material for “Multivariate Density Regression for Censored, Constrained, and Binary Traits”

S. Wade∗ R. Piccarreta † A. Cremaschi ‡ I. Antoniano-Villalobos §

July 23, 2019

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.

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 $\tilde{Z}_3 = \tilde{Z}_1 + \tilde{Z}_3$. For Work Status, we have $Z_4 = \mathbb{1}_{(0,\infty)}(Y_4)$. In the following, let $J$...

∗School of Mathematics, University of Edinburgh, UK
†BIDSA and Dept. of Decision Sciences, Bocconi University, Milan, Italy
‡Dept. of Cancer Immunology, Institute of Cancer Research, Oslo University Hospital, Norway
§Dept. of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, Italy
denote the final truncation level, with corresponding weighted particles \((w_{1,J}^m, \theta_{1,J}^m, \psi_{1,J}^m, y_{1,m}^m)\) and unnormalized particle weights \(\tilde{\varrho}^m\), for \(m = 1, \ldots, M\) (without loss of generality, we drop the subscript \(J\)). We indicate with \(\varrho^m\), 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 \(x_*\) (e.g. for \(\ell = 4\), shown in Figure 3) is:

\[
P(Z_{s,\ell} = 1| x, z, x_*) = P(Y_{s,\ell} > 0| x, z, x_*) \approx \sum_{m=1}^{M} \varrho^m \sum_{j=1}^{J} w_{j}^m(x_*) \Phi \left( \frac{x_* \beta_{\ell,J}^m}{\sqrt{\Sigma_{\ell,J}^m}} \right).
\]

For \(\ell = 1, 2, 3\), the marginal predictive density of \(\tilde{Z}_{s,\ell}\) given \(x_*\), shown in Figure 4 for some values of \(x_*\), is given by:

\[
f(\tilde{z}_{s,\ell}| x, z, x_*) \approx \sum_{m=1}^{M} \varrho^m \sum_{j=1}^{J} w_{j}^m(x_*) f(\tilde{z}_{s,\ell}| \theta_{\ell}^m, x_*)
= \sum_{m=1}^{M} \varrho^m \sum_{j=1}^{J} w_{j}^m(x_*) \logN(\tilde{z}_{s,\ell}| x_* \beta_{\ell,J}^m, \Sigma_{\ell,J}^m),
\]

for \(\tilde{z}_{s,\ell} > 0\), where \(\beta_{\ell,J}^m\) denotes the \(\ell\)-th column of \(\beta\) in component \(j\) and particle \(m\); \(\Sigma_{\ell,J}^m\) denotes element \((\ell, \ell)\) of the matrix \(\Sigma\) in component \(j\) and particle \(m\); and \(\logN(\cdot| \mu, \sigma^2)\) denotes the log-normal density with parameters \(\mu\) and \(\sigma^2\). 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 3) can be computed numerically by evaluating the marginal predictive density \((1)\) on a sufficiently dense grid of \(\tilde{z}_{s,\ell}\) 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 \(x_{s,1}\) years of age (Figure C.3), computed as:

\[
P(\tilde{Z}_{s,\ell} \geq (x_{s,1} + 1)| x, z, x_*) = P(Y_{s,\ell} > \log(x_{s,1} + 1)| x, z, x_*) \approx \sum_{m=1}^{M} \varrho^m \sum_{j=1}^{J} w_{j}^m(x_*) \left( 1 - \Phi \left( \frac{\log(x_{s,1} + 1) - x_* \beta_{\ell,J}^m}{\sqrt{\Sigma_{\ell,J}^m}} \right) \right).
\]
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 4, given $x_{*,1}$.

Our model also recovers the joint relationship between responses, which allows inference on conditional properties. Specifically, when $\ell$ indexes a binary response and $\ell'$ indexes an age at event response, the conditional predictive probability of success given $\tilde{z}_{*,\ell'}$ and $x_*$ (Figure 6) is:

$$
\mathbb{P}(Z_{*,\ell} = 1 | \tilde{z}_{*,\ell'}, x, z, x_*) \approx \sum_{m=1}^{M} \sum_{j=1}^{J} w_j^n(x_*) \Phi\left( \frac{\mu_{j,\ell,\ell'}^m}{\sigma_{j,\ell,\ell'}^2} \right) \frac{\log N(\tilde{z}_{*,\ell'} | x_*, \beta_{j,\ell,\ell'}^m, \Sigma_{j,\ell,\ell'}^m)}{f(\tilde{z}_{*,\ell'} | x, z, x_*)},
$$

where

$$
\mu_{j,\ell,\ell'}^m = x_* \beta_{j,\ell,\ell'}^m + \sum_{j=1}^{J} (\Sigma_{j,\ell,\ell'}^m)^{-1} (\log(\tilde{z}_{*,\ell'}) - x_* \beta_{j,\ell,\ell'}^m),
$$

$$
\sigma_{j,\ell,\ell'}^2 = \Sigma_{j,\ell,\ell'}^m - (\Sigma_{j,\ell,\ell'}^m)^2 \Sigma_{j,\ell,\ell'}^m - 1,
$$

and the density in the denominator is the marginal predictive of equation (4). For $\ell \neq \ell'$ both indexing ages at event, the conditional predictive density of $\tilde{Z}_{*,\ell}$ given $\tilde{z}_{*,\ell'}$ and $x_*$ takes the form:

$$
f(\tilde{z}_{*,\ell} | \tilde{z}_{*,\ell'}, x, z, x_*) = \sum_{m=1}^{M} \sum_{j=1}^{J} w_j^n(x_*) \log N(\tilde{z}_{*,\ell} | \mu_{j,\ell,\ell'}^m, \sigma_{j,\ell,\ell'}^2) \frac{\log N(\tilde{z}_{*,\ell'} | x_*, \beta_{j,\ell,\ell'}^m, \Sigma_{j,\ell,\ell'}^m)}{f(\tilde{z}_{*,\ell'} | x, z, x_*)}, \tag{3}
$$

Figure [C.6] shows the conditional predictive density of $\tilde{Z}_{*,2} - \tilde{z}_{*,1}$ given $\tilde{z}_{*,1}$ and $x_*$, which can be easily computed from (3). The corresponding predictive medians (Figure 5) 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.7]) and the corresponding median (Figure [C.4]) can be obtained directly from equation (3).

Finally, we note that for the (undiscretized) age at first child, $\tilde{Z}_{*,3} = \tilde{Z}_{*,1} + \tilde{Z}_{*,3}$, 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 $\tilde{Z}_{*,3}$ given $\tilde{z}_{*,1}$ is simply the conditional predictive density of equation (3), evaluated at
\( \hat{z}_{s,3} - \hat{z}_{s,1} \). While the marginal predictive density of \( \tilde{Z}_{s,3} \) given \( x_* \) (Figure 4) is obtained as:

\[
f(\tilde{z}_{s,3}|x, z, x_*) = \int f(\tilde{z}_{s,3}|\hat{z}_{s,1}, x, z, x_*) f(\hat{z}_{s,1}|x, z, x_*) d\hat{z}_{s,1}
\]

\[
\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(x_*) \int_{-\infty}^{\tilde{z}_{s,3}} \log N(\hat{z}_{s,3}|\mu_{j,3(1)}^m, \sigma_{j,3(1)}^m) \log N(\hat{z}_{s,1}|x_* \beta_{j(1)}^m, \Sigma_{j(1,1)}^m) d\hat{z}_{s,1}, \tag{4}
\]

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

\[
\mathbb{P}(\tilde{Z}_{s,3} > x_{s,1}|x, z, x_*) = \mathbb{P}(\tilde{Z}_{s,3} + \tilde{Z}_{s,1} > x_{s,1} + 1|x, z, x_*)
\]

\[
\approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(x_*) \int \left( 1 - \Phi \left( \frac{l(x_{s,1} + 1) - \mu_{j,3(1)}^m}{\sqrt{\sigma_{j,3(1)}^m}} \right) \right) \log N(\tilde{z}_{s,1}|x_* \beta_{j(1)}^m, \Sigma_{j(1,1)}^m) d\tilde{z}_{s,1},
\]

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

\[
f(\tilde{z}_{s,3}|\tilde{z}_{s,2}, x, z, x_*) \approx \sum_{m=1}^{M} \vartheta^m \sum_{j=1}^{J} w_j^m(x_*) \log N(\tilde{z}_{s,2}|x_* \beta_{j(2)}^m, \Sigma_{j(2,2)}^m) \log N(\hat{z}_{s,1}|x_* \beta_{j(1)}^m, \Sigma_{j(1,1)}^m) d\tilde{z}_{s,1}, \tag{5}
\]

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

\[
f(\tilde{z}_{s,3}|\tilde{z}_{s,2}, \theta_j^m, x_*) = \int_{-\infty}^{\tilde{z}_{s,3}} \log N(\tilde{z}_{s,3} - \tilde{z}_{s,1}|x_* \beta_{j(1,2)}^m, \sigma_{j,3(1,2)}^m) \log N(\hat{z}_{s,1}|x_* \beta_{j(1,1)}^m, \sigma_{j,3(1,1)}^m) d\tilde{z}_{s,1},
\]

where

\[
\mu_{j,3(1,2)}^m = x_* \beta_{j(1,2)}^m + \Sigma_{j,3(1,2)}^m \Sigma_{j,1,2(1,2)}^{-1} \log(\tilde{z}_{s,1}) - x_* \beta_{j(1,1)}^m,
\]

\[
\sigma_{j,3(1,2)}^2 = \Sigma_{j,3(1,2)}^m - \Sigma_{j,3(1,2)}^m \Sigma_{j,1,2(1,2)}^{-1} \Sigma_{j,1,2(1,2)}^m,
\tag{6}
\]

Figure C.8 shows the conditional predictive density of \( \tilde{Z}_{s,3} - \tilde{z}_{s,2} \) given \( \tilde{z}_{s,2} \) and \( x_* \), which can be easily computed from (5), with the corresponding predictive medians in Figure C.5.
Lastly, the conditional predictive probability of success for a binary response, e.g. \( \ell = 4 \) in our application, given \( \tilde{z}_{*,3} \) and \( x_* \) is:

\[
\Pr(Z_{*,4} = 1|\tilde{z}_{*,3}, x, z, x_*) \approx \sum_{m=1}^{M} \sum_{j=1}^{J} \sum_{i=0}^{q} w_j^{m}(x_*) \Pr(Y_{*,4} > 0|\tilde{z}_{*,3}, \theta_j^m, x_*) \frac{f(\tilde{z}_{*,3}|\theta_j^m, x_*)}{f(\tilde{z}_{*,3}|x, z, x_*)},
\]

where

\[
\Pr(Y_{*,4} > 0|\tilde{z}_{*,3}, \theta_j^m, x_*) f(\tilde{z}_{*,3}|\theta_j^m, x_*) = \int_{-\infty}^{\log(\tilde{z}_{*,3})} \Phi \left( \frac{\mu_{j,4}(1,3)}{\sqrt{\sigma_{j,4}^2(1,3)}} \right) \logN(\tilde{z}_{*,3} - z_{*,1}|\mu_{j,3}(1,1), \sigma_{j,3}^2(1,1)) \logN(z_{*,1}|x_*\beta_{j,1,1}, \Sigma_{j,1,1}) d\tilde{z}_{*,1},
\]

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

## B Simulation Study

We generate a dataset of size \( n = 700 \) with \( q^* = 3 \) covariates and \( d = 3 \) responses. The first covariate mimics \( Age \) and, as such, is assumed to be registered at a discrete level: \( x_1 = [\tilde{x}_1] \), where \( \tilde{x}_1 \sim \U(15, 30) \). 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. To build \( Z_1 \), we first generate:

\[
\tilde{Z}_{i,1} = \mu_1^t(\tilde{x}_{i,1}, x_{i,2}^*, x_{i,3}^*) + \epsilon_{i,1}, \text{ for } i = 1, \ldots, n,
\]

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

\[
\mu_1^t(\tilde{x}_{i,1}, x_{i,2}^*, x_{i,3}^*) = \begin{cases} 
-0.057\tilde{x}_{i,1}^2 + 3.08\tilde{x}_{i,1} - 21.247 & \text{if } x_{i,2}^* \neq 1, x_{i,3}^* = 2 \\
\frac{1}{3}\tilde{x}_{i,1} + 10 & \text{if } x_{i,2}^* \neq 1, x_{i,3}^* = 1 \\
0.0001\tilde{x}_{i,1}^3 - 0.0695\tilde{x}_{i,1}^2 + 3.83\tilde{x}_{i,1} - 30.584 & \text{if } x_{i,2}^* = 1, x_{i,3}^* = 2 \\
\frac{8}{15}\tilde{x}_{i,1} + 7 & \text{if } x_{i,2}^* = 1, x_{i,3}^* = 1
\end{cases}
\]
Similarly, to build $Z_2$, we generate:

$$
\tilde{Z}_{i,2} = \begin{cases} 
-0.056\tilde{x}_{i,1}^2 + 3.08\tilde{x}_{i,1} - 18 + 0.75\left[\tilde{z}_{i,1} - \mu_1^t(\tilde{x}_{i,1}, x_{i,2}^*, x_{i,3}^*)\right] + \epsilon_{i,2} & \text{if } x_{i,3}^* = 2 \\
0.5\tilde{x}_{i,1} + 8 + 0.75\left[\tilde{z}_{i,1} - \mu_1^t(\tilde{x}_{i,1}, x_{i,2}^*, x_{i,3}^*)\right] + \epsilon_{i,2} & \text{if } x_{i,3}^* = 1
\end{cases},
$$

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

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

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. $z_1 = \lfloor \tilde{z}_1 \rfloor$, $z_2 = \lfloor \tilde{z}_2 \rfloor$. Finally, a binary response variable is simulated as:

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

**Prior specification.** In the simulated study, 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 $y_{i,\ell} = (l_{i,\ell} + u_{i,\ell})/2$ 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 $\hat{\beta}$ of the linear coefficients and $\hat{\Sigma}$ of the covariance matrix. We then define

$$
E[\beta_j] = \beta_0 = \hat{\beta} \quad \text{and} \quad E[\Sigma_j] = \frac{1}{\nu - b - 1}\Sigma_0 = \hat{\Sigma}.
$$

Together, $U$ and $\Sigma_j$ reflect the variability of $\beta_j$ across components, and we set $U$ such that $\min(\text{diag}(\hat{\Sigma}))U = 10(X'X)^{-1}$. 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 $\Sigma_j$ a-priori. Other specified hyperparameters include $\mu_{0,1} = \bar{x}_1$, $u_1 = 1/2$, $\alpha_1 = 2$, $\gamma_1 =$
\( u_1(\text{range}(x_{1:n,1})/4)^2, \mathbf{q}_k = (1, 1) \) 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 \( \bar{x}_1 \) and \( \text{range}(x_{1:n,1}) \) denote the sample mean and range of \( (x_{1,1}, \ldots, x_{n,1}) \).

**Robustness analysis.** We perform a robustness analysis comparing several initialization specifications, namely by setting \( J_0 = 2, 3, 5, 10, 15, 20, 30 \), and show the results for two different discrepancy measures used to define the stopping rule of SMC, i.e. ESS and CESS. We also offer a comparison with a parameteric version of the proposed model. 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. The ESS of the log-likelihood for the MCMC stage (ESSMCMC) is computed with the \texttt{mcmcse} package in R \citep{flegal2017}.

A summary of the analysis is reported in Table B.1. The quantities used in this comparison include the LPML 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 \( i = 1, \ldots, n^* \):

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

\[
\text{ERR}_{\text{Mean}}^\ell = \frac{100}{n^*} \sum_{i=1}^{n^*} \frac{|\mu_i^\ell(\mathbf{x}_{i}^*) - \hat{\mu}_\ell(\mathbf{x}_{i}^*)|}{|\mu_i^\ell(\mathbf{x}_{i}^*)|},
\]

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

where for each response \( \ell = 1, \ldots, d \), \( \mu_i^\ell(\mathbf{x}_{i}^*) \) and \( \hat{\mu}_\ell(\mathbf{x}_{i}^*) \) indicate the true and estimated mean functions, and \( f^\ell(\cdot|\mathbf{x}_{i}^*) \) and \( \hat{f}(\cdot|\mathbf{x}_{i}^*) \) 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 \( \Delta \). The results show robustness with respect to the choice of the discrepancy measure.
Table B.1: Simulation study. 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.

### C Application: Life Patterns of Colombian Women

#### C.1 The Data

The DHS Program collects and disseminates data on random samples of households selected from random clusters from a national sampling frame.\(^1\) The 2010 survey in Colombia was conducted by the Profamilia association, and we refer to the final report for a detailed description of its features (Ojeda et al., 2011). Since all the women of childbearing potential (i.e. aged 13-49) in the same household were interviewed, we randomly select at most one case from each household to avoid unwanted dependencies.

To describe the characteristics of the fertility and partnership patterns, we consider the discrete variables recording the ages at Sexual Debut, at Union, referring to the first marriage or cohabitation, and at First Child. The Work Status is a binary variable indicating if the woman worked in the 12 months before the interview. We exclude women who gave inconsistent information, namely, those who report the birth of the first child as preceding the first sexual intercourse, and those who report union with a partner but for whom sexual intercourse never occurred. We also filter out women who experienced sexual violence or were

\(^1\)available through [https://www.dhsprogram.com/](https://www.dhsprogram.com/)
forced to have sex in exchange for money, as we consider that their choices concerning union
and childbirth may be related to the experienced violence. Following the same reasoning,
we remove women who were forced to use contraceptive methods. Thus, we attempt to focus
as much as possible on life choices and plans rather than on events imposed by circumstances,
even if the latter may be unknown and unmeasured, so that the observed events may not
necessarily reflect choices.

We are interested in the relationship between the responses and some baseline socio-
demographic factors. First, we consider the woman’s *Age* (in years) at the moment of inter-
view. We focus on women aged 15 or more, as most younger women had not yet experienced
any event at the time of the survey. Next, we include the *Region* (Atlantica, Oriental, Cen-
tral, Pacifica, Bogota, Territorios Nacionales) and the type of *Area* (urban or rural) where the
respondent lives. Since information is only available on the current region of residence and
on the age when she moved there, we limit attention to respondents who were raised in the
current region at least from the age of 6, to properly account for regional effects. Moreover, to
assess the respondent’s well-being in her original family, we refer to the disciplining methods
used by her parents in her childhood, considering if she was exposed to *Physical Punishment*
(spanking, hitting, pushing, throwing water) or not. Also, we account for the exposure of the
respondent to *Parental Domestic Violence*, indicating if she ever witnessed her father beating
her mother. All cases where a respondent chose not to report on at least one explanatory or
response variable are excluded.

We note that even if the DHS dataset is very rich, including other covariates is not straight-
forward. Most of the variables refer to the moment of interview, and thus cannot be considered
as antecedents of the focal events. For example, although it would be interesting to include
information on education and wealth, only the highest level of education attained and the
wellness of the respondent’s family at the moment of interview are available. Another rele-
vant aspect that could be taken into account concerns women’s ethnicity. However, most
(about 80%) of the women in the sample do not recognize themselves as part of an ethnic
minority. Furthermore, those who do, belong to a heterogeneous variety of ethnic groups,
none of which is sufficiently represented in the sample. We therefore exclude ethnic minorities
from our study. Our final dataset consists of *n = 10,740* women.
**Prior specification.** For our motivating application, the prior parameters for the linear coefficients and covariance matrix of each component are once again specified empirically based on a multivariate linear regression fit. Specifically, we set $y_{i,\ell} = (l_{i,\ell} + u_{i,\ell})/2$ for uncensored observations, where the bounds $l_{i,\ell}$ and $u_{i,\ell}$ are defined in Section 5. For $\ell = 3$, when the lower bound is $-\infty$, i.e. age at sexual debut is equal to age at first child, we set $y_{i,3} = u_{i,3} - 1$. For censored observations, we sample $y_{i,\ell}$ from a truncated normal distribution with mean and covariance computed from the uncensored observations. For the binary response, $y_{4,i} = -1$ for $z_{4,i} = 0$ and $y_{4,i} = 1$ for $z_{4,i} = 1$. A multivariate linear regression fit for this auxiliary response gives estimates $\hat{\beta}$ of the linear coefficients and $\hat{\Sigma}$ of the covariance matrix. We then define

$$E[\beta_j] = \beta_0 = \hat{\beta} \quad \text{and} \quad E[\Sigma_j] = \frac{1}{\nu - b - 1} \Sigma_0 = \hat{\Sigma}. $$

Together, $\mathbf{U}$ and $\Sigma_j$ reflect the variability of $\beta_j$ across components, and we set $\mathbf{U}$ such that $\min(\text{diag}(\hat{\Sigma})) \mathbf{U} = 20(X'X)^{-1}$. 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 $\nu = b + 3$. Other specified hyperparameters include $\mu_{0,1} = \bar{x}_1; \ u_1 = 1/2; \ \alpha_1 = 2; \ \gamma_1 = u_1(\text{range}(x_{1:n,1})/4)^2; \ \varrho_k = (1,1)$ for $k = 2, \ldots, q$; and the parameters of the stick-breaking prior are $\zeta_{j,1} = 1$ and $\zeta_{j,2} = 1$. 

**Algorithm details.** We initialize the MCMC algorithm with $J_0 = 35$ components, a number large enough to avoid a small ESS and subsequent resampling. Indeed, for large sample sizes, the parametric mixture likelihoods, unnormalized weights and normalizing constant can no longer be saved for every data point and particle, due to memory constraints. Thus, if resampling is required, we must recompute these terms at each block update of the MCMC rejuvenation step. In our example, this resulted in approximately a three-fold increase in computation time. In this case, a more computationally efficient approach is to initialize with a generous number of components. Due to the robustness of the algorithm with respect to the stopping rule based on ESS or CESS in simulations, we consider only ESS here.
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.1 complements Figure 3 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. (P, B) or (P, B). Figure C.2 completes Figure 4 by displaying the predictive density of the age at sexual debut. Figure C.3 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.4 and for the time from union to first child given the age at union is shown in Figure C.5. The underlying conditional predictive densities for selected covariate combinations are visualized in Figures C.6, C.7, and C.8. Finally, to explore the possible relation between anticipation of union on work activity, Figure C.9 reports the conditional predictive probability of working as function of Age given different ages at union.

References

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

G. Ojeda, M. Ordóñez, and L. H. Ochoa. *Colombia Encuesta Nacional de Demografía y Salud 2010*. Bogotá, Colombia: Profamilia, 2011. Available at [http://dhsprogram.com/pubs/pdf/FR246/FR246.pdf](http://dhsprogram.com/pubs/pdf/FR246/FR246.pdf).
Figure C.1: 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 \((P, \bar{B})\) or \((\bar{P}, B)\). Dotted lines indicate when the median exceeds Age. Combined with Figure 3, 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.2: Predictive density of the age at sexual debut as a function of Age for women who grew up in violent (P, B) and non-violent families (P, B). Analogously to Figure 4, 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.3: 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 ($P, B$) and non-violent families ($\bar{P}, \bar{B}$). Equivalently, the censoring probability represents the mass above the dashed line for a given Age in the density plots of Figures 4 and C.2 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.4: 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 (P, B) and non-violent families (P̅, B̅). 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 3, 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.5: 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 \((P_1, B_1)\) and non-violent families \((\bar{P}_1, \bar{B}_1)\). 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.6: 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 ($\overline{P}, \overline{B}$) 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 5, 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.7: 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 (\(\bar{P}, \bar{B}\)) 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.8: 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 (P, 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.9: Conditional predictive probability of working as function of Age given different ages at union, for women who grew up in violent (P, B) and non-violent families (P̅, B̅). 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 6, no scaring effect is visible, i.e. the probability of working in older cohorts is unaffected by the conditioned age at union.