Orthogonalized Design Matrices Speed-ups of Bayesian Semiparametric Regression

NURUL FITRIYANI AND MATT P. WAND

School of Mathematical and Physical Sciences, University of Technology Sydney

9th July, 2026

Abstract

We explain how important classes of Bayesian semiparametric regression fitting and inference procedures can be sped up, significantly, via the use of orthogonalized design matrices. Typically, design matrices in semiparametric regression contain predictor observations and basis functions of such data. In *Bayesian* semiparametric regression, loop-type approaches such as Gibbs sampling and coordinate ascent variational inference typically are required. We show that pre-loop reformulation of Bayesian semiparametric regression models involving orthogonalized design matrices lead to two orders of magnitude, with respect to column dimension, computational reduction. Our computer experiments reveal that this simple paradigm results in approximately 5- to 60-fold speed-ups.

Keywords: Bayesian penalized splines; Generalized additive models; Group-specific curve models; Mean field variational Bayes; Markov chain Monte Carlo.

1 Introduction

Practical fitting and inference for Bayesian semiparametric regression models usually requires the running of loop-type procedures such as Gibbs sampling or coordinate ascent variational inference. The essence of this article is explaining and demonstrating how conversion of the model to an equivalent form involving orthogonalized design matrices provides two orders of magnitude reductions in computation. These orders of magnitude are with respect to the column dimensions of the design matrices, which can be in the dozens or even hundreds for contemporary Bayesian semiparametric models. Computer experiments reveal that practical speed-ups are in the approximate 5–60 factor range.

In Section 2 we commence our explanation of orthogonalized design matrices speed-ups for a Bayesian nonparametric regression model. This allows us to describe the essence of the approach with minimal notational overhead. Both Gaussian and Bernoulli response models are shown to benefit from the approach and our computer experiments in this section demonstrate speed-ups as high as a factor of around 60. We also corroborate these results with some order of magnitude comparisons. Algorithms 2 and 3 facilitate implementation.

Sections 3 and 4 explore the central theme applied to generalized additive models and group-specific curves models, respectively. For such arbitrarily large models the speed-ups offered by use orthogonalized design matrices can have noticeable practical benefits, as demonstrated by some simulated and actual data examples.

Even though speeding up Gibbs sampling is the main focus of this article, the general approach also applies to other Bayesian inference approach such as coordinate ascent variational inference. In Section 5 we provide illustration for the Bayesian nonparametric regression model from Section 2.

In Sections 2–5 we present five algorithms with orthogonalized design matrices speed-ups across three Bayesian semiparametric regression model types. These allow concrete illustration and evaluation of the approach. However, orthogonalized design matrices speed-ups is a general paradigm that applies to many other Bayesian regression-type models.

Section 6 contains some concluding discussion.

1.1 Notation

Scalar functions applied to a vector are evaluated in an element-wise fashion. For example, $\cosh([3\ 11]^T) \equiv [\cosh(3)\ \ \cosh(11)]^T$. If $\boldsymbol{v}$ is a column vector then $\|\boldsymbol{v}\| \equiv \sqrt{\boldsymbol{v}^T\boldsymbol{v}}$ is the Euclidean norm of $\boldsymbol{v}$. The notation diag($\boldsymbol{v}$) is used for the diagonal matrix containing the entries of $\boldsymbol{v}$ along its diagonal. If $\boldsymbol{v}$ and $\boldsymbol{w}$ are column vectors of the same size then $\boldsymbol{v} \odot \boldsymbol{w}$ and $\boldsymbol{v}/\boldsymbol{w}$ are the column vectors of element-wise products and quotients, respectively. Also, $\mathbf{1}$ is a column vector of ones. The symbol $\stackrel{\text{ind.}}{\sim}$ is shorthand for "independently distributed as". The random variable x has a Gamma distribution with shape parameter $\kappa > 0$ and rate parameter $\lambda > 0$, written $x \sim \text{Gamma}(\kappa, \lambda)$, if and only if the density function of x is $\mathfrak{p}(x) = \{\lambda^\kappa/\Gamma(\kappa)\}\, x^{\kappa-1}\exp(-\lambda\, x),\ x > 0$. The random variable x has an Inverse Gamma distribution with shape parameter $\kappa > 0$ and rate parameter $\lambda > 0$, written $x \sim \text{Inverse-Gamma}(\kappa, \lambda)$, if and only if the density function of x is $\mathfrak{p}(x) = \{\lambda^\kappa/\Gamma(\kappa)\}\, x^{-\kappa-1}\exp(-\lambda/x)$, $x > 0$. The symbol Φ is used for cumulative distribution function of the $N(0,1)$ distribution. For a logical proposition $\mathcal{P}$, $I(\mathcal{P}) = 1$ if $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ does not hold.

2 Nonparametric Regression

Bayesian nonparametric regression via penalized splines (e.g. Harezlak *et al.*, 2018, Chapter 2) is one of the simplest model types that benefits from the use of orthogonalized design matrices. In this section we start with the simplest Gaussian response case and then, later, discuss other response situations.

2.1 Gaussian Responses

For univariate and continuous predictor/response pairs (x_i, y_i), $1 \le i \le n$, the Gaussian responses nonparametric regression model has the generic form

$$y_i \stackrel{\text{ind.}}{\sim} N(f(x_i), \sigma_\varepsilon^2), \quad 1 \le i \le n. \tag{1}$$

Throughout this section we consider the following Bayesian penalized spline model for f:

$$f(x) = \beta_0 + \beta_1\, x + \sum_{k=1}^{K} u_k z_k(x), \quad u_k|\sigma_u^2 \stackrel{\text{ind.}}{\sim} N(0, \sigma_u^2),$$

where $\{z_k(\cdot) : 1 \le k \le K\}$ is a suitable spline basis. Typically K is an integer between around 25 and 50, but may be higher if f is thought to be particularly wiggly (e.g. Section 2.4 of Harezlak *et al.*, 2018). A recommended default choice of the z_k is described in Section 4 of Wand & Ormerod (2008) and corresponds to low-rank smoothing splines. The full description of the model that we consider here is

$$\begin{gathered}\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \sigma_\varepsilon^2 \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\,\boldsymbol{u}, \sigma_\varepsilon^2\,\boldsymbol{I}_n), \quad \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2\boldsymbol{I}_2), \quad \boldsymbol{u}|\sigma_u^2 \sim N(\mathbf{0}, \sigma_u^2\boldsymbol{I}_K),\\ \sigma_u^{-2}|b_u \sim \text{Gamma}\big(\tfrac{1}{2}, b_u\big), \quad b_u \sim \text{Gamma}\big(\tfrac{1}{2}, s_u^{-2}\big),\\ \sigma_\varepsilon^{-2}|b_\varepsilon \sim \text{Gamma}\big(\tfrac{1}{2}, b_\varepsilon\big), \quad b_\varepsilon \sim \text{Gamma}\big(\tfrac{1}{2}, s_\varepsilon^{-2}\big)\end{gathered} \tag{2}$$

where $\sigma_\beta, s_u, s_\varepsilon > 0$ are user-specified hyperparameters,

$$\boldsymbol{y} \equiv [y_i]_{1 \le i \le n}, \quad \boldsymbol{X} \equiv [1\ x_i]_{1 \le i \le n} \quad \text{and} \quad \boldsymbol{Z} \equiv [z_k(x_i)]_{\substack{1 \le i \le n \\ 1 \le k \le K}}.$$

The vectors $\boldsymbol{\beta}$ (2×1) and $\boldsymbol{u}$ $(K \times 1)$ contain the β_k and u_k, respectively. The distributional structure of σ_ε in (2) involving the auxiliary variable b_ε is equivalent to imposition of the prior density function

$$\mathfrak{p}(\sigma_\varepsilon) = \frac{2}{\pi\{1 + (\sigma_\varepsilon^2/s_\varepsilon^2)\}s_\varepsilon}, \quad \sigma_\varepsilon > 0,$$

Inputs: $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times 2)$, $\boldsymbol{Z}$ $(n \times K)$, $\boldsymbol{X_g}$ $(N_{\text{grid}} \times 2)$, $\boldsymbol{Z_g}$ $(N_{\text{grid}} \times K)$,

$\sigma_\beta, s_u, s_\varepsilon > 0$, $N_{\text{burn}}, N_{\text{kept}} \in \mathbb{N}$.

Initialize: $\boldsymbol{u}^{[0]}$ $(K \times 1)$, $(\sigma_\varepsilon^{-2})^{[0]}, b_\varepsilon, (\sigma_u^{-2})^{[0]}, b_u > 0$.

XTy $\longleftarrow \boldsymbol{X}^T\boldsymbol{y}$; ZTy $\longleftarrow \boldsymbol{Z}^T\boldsymbol{y}$; XTX $\longleftarrow \boldsymbol{X}^T\boldsymbol{X}$; ZTX $\longleftarrow \boldsymbol{Z}^T\boldsymbol{X}$; ZTZ $\longleftarrow \boldsymbol{Z}^T\boldsymbol{Z}$

For $s = 1, \ldots, N_{\text{burn}} + N_{\text{kept}}$:

$$\boldsymbol{r}_\beta \longleftarrow \text{XTy} - \text{ZTX}^T\,\boldsymbol{u}^{[s-1]} \;\; ; \;\; \boldsymbol{\Psi}_\beta \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]}\text{XTX} + \sigma_\beta^{-2}\boldsymbol{I}_2$$

Decompose $\boldsymbol{\Psi}_\beta = \boldsymbol{U}_\beta \text{diag}(\boldsymbol{d}_\beta)\boldsymbol{U}_\beta^T$ where $\boldsymbol{U}_\beta^T\boldsymbol{U}_\beta = \boldsymbol{U}_\beta\boldsymbol{U}_\beta^T = \boldsymbol{I}_2$

$$\boldsymbol{z}_\beta \sim N(\boldsymbol{0}, \boldsymbol{I}_2) \;\; ; \;\; \boldsymbol{\beta}^{[s]} \longleftarrow \boldsymbol{U}_\beta\left(\frac{\boldsymbol{U}_\beta^T\boldsymbol{z}_\beta}{\sqrt{\boldsymbol{d}_\beta}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{U}_\beta^T\boldsymbol{r}_\beta}{\boldsymbol{d}_\beta}\right)$$

$$\boldsymbol{r}_u \longleftarrow \text{ZTy} - \text{ZTX}\,\boldsymbol{\beta}^{[s]} \;\; ; \;\; \boldsymbol{\Psi}_u \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]}\text{ZTZ} + (\sigma_u^{-2})^{[s-1]}\boldsymbol{I}_K$$

Decompose $\boldsymbol{\Psi}_u = \boldsymbol{U}_u \text{diag}(\boldsymbol{d}_u)\boldsymbol{U}_u^T$ where $\boldsymbol{U}_u^T\boldsymbol{U}_u = \boldsymbol{U}_u\boldsymbol{U}_u^T = \boldsymbol{I}_K$

$$\boldsymbol{z}_u \sim N(\boldsymbol{0}, \boldsymbol{I}_K) \;\; ; \;\; \boldsymbol{u}^{[s]} \longleftarrow \boldsymbol{U}_u\left(\frac{\boldsymbol{U}_u^T\boldsymbol{z}_u}{\sqrt{\boldsymbol{d}_u}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{U}_u^T\boldsymbol{r}_u}{\boldsymbol{d}_u}\right)$$

$$(\sigma_u^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(K+1), b_u + \tfrac{1}{2}\|\boldsymbol{u}^{[s]}\|^2\Big) \;\; ; b_u \longleftarrow \text{Gamma}\Big(1, (\sigma_u^{-2})^{[s]} + s_u^{-2}\Big)$$

$$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(n+1), b_\varepsilon + \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^{[s]} - \boldsymbol{Z}\boldsymbol{u}^{[s]}\|^2\Big)$$

$$b_\varepsilon \longleftarrow \text{Gamma}\Big(1, (\sigma_\varepsilon^{-2})^{[s]} + s_\varepsilon^{-2}\Big)$$

For $s = 1, \ldots, N_{\text{kept}}$:

$$\widehat{\boldsymbol{f}}_{\boldsymbol{g}}^{[s]} \longleftarrow \boldsymbol{X_g}\boldsymbol{\beta}^{[s+N_{\text{burn}}]} + \boldsymbol{Z_g}\boldsymbol{u}^{[s+N_{\text{burn}}]}$$

$$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow (\sigma_\varepsilon^{-2})^{[s+N_{\text{burn}}]} \;\; ; \;\; (\sigma_u^{-2})^{[s]} \longleftarrow (\sigma_u^{-2})^{[s+N_{\text{burn}}]}$$

Outputs: $\big\{\widehat{\boldsymbol{f}}_{\boldsymbol{g}}^{[s]}, (\sigma_\varepsilon^{-2})^{[s]}, (\sigma_u^{-2})^{[s]} : 1 \le s \le N_{\text{kept}}\big\}$

Algorithm 1: *A direct Gibbs sampling algorithm for the Bayesian nonparametric regression model (2).*

which corresponds to the Half-Cauchy distribution with scale parameter s_ε (e.g. Gelman, 2006). The prior on σ_u is analogous.

Standard calculations show that the full conditional distribution of $\boldsymbol{u}$ is

$$\boldsymbol{u}|\text{rest} \sim N\big(\boldsymbol{\Psi}_u^{-1}\sigma_\varepsilon^{-2}\boldsymbol{r}_u, \boldsymbol{\Psi}_u^{-1}\big) \quad \text{where} \quad \boldsymbol{r}_u \equiv \boldsymbol{Z}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$
$$\text{and} \quad \boldsymbol{\Psi}_u \equiv \sigma_\varepsilon^{-2}\boldsymbol{Z}^T\boldsymbol{Z} + \sigma_u^{-2}\boldsymbol{I}_K$$

where 'rest' denotes all random variables in (2) other than $\boldsymbol{u}$. An analogous result holds for $\boldsymbol{\beta}|$rest. Derivations of the full conditionals of the scalar variables in (2) are particular simple, and lead to the direct Gibbs sampling scheme listed in Algorithm 1. Result 1 in the appendix justifies the forms of the $\boldsymbol{\beta}$ and $\boldsymbol{u}$ Multivariate Normal draws. The main output of Algorithm 1 is the kept Gibbs samples of fit vectors for inputted grid-wise design matrices $\boldsymbol{X_g}$ and $\boldsymbol{Z_g}$. These matrices are defined analogously to $\boldsymbol{X}$ and $\boldsymbol{Z}$ but with basis functions evaluated at grid points stored in an arbitrary vector N_{grid} of plotting abscissae, $\boldsymbol{g}$, rather than the x_is.

The main bottleneck in Algorithm 1 is obtaining the singular value decomposition of the $K \times K$ matrix $\boldsymbol{\Psi}_u$ for each iteration of the Gibbs sampling scheme, which requires $O(K^3)$ operations. The essence of the approach advocated in this article starts with the singular value decomposition of $\boldsymbol{Z}$:

$$\boldsymbol{Z} = \boldsymbol{U_Z}\text{diag}\big(\boldsymbol{d_Z}\big)\boldsymbol{V_Z}^T \quad \text{where} \quad \boldsymbol{U_Z}^T\boldsymbol{U_Z} = \boldsymbol{V_Z}^T\boldsymbol{V_Z} = \boldsymbol{V_Z}\boldsymbol{V_Z}^T = \boldsymbol{I}_K \tag{3}$$

and where $\boldsymbol{U_Z}$ is $n \times K$, $\boldsymbol{d_Z}$ is $K \times 1$ and $\boldsymbol{V_Z}$ is $K \times K$. Then observe that

$$\boldsymbol{Z}\boldsymbol{u} = \breve{\boldsymbol{Z}}\breve{\boldsymbol{u}} \quad \text{where} \quad \breve{\boldsymbol{Z}} \equiv \boldsymbol{U_Z}\text{diag}\big(\boldsymbol{d_Z}\big) \quad \text{and} \quad \breve{\boldsymbol{u}} \equiv \boldsymbol{V_Z}^T\boldsymbol{u}. \tag{4}$$

Inputs: $\boldsymbol{y}$ $(n\times 1)$, $\boldsymbol{X}$ $(n\times 2)$, $\boldsymbol{Z}$ $(n\times K)$, $\boldsymbol{X_g}$ $(N_{\text{grid}}\times 2)$, $\boldsymbol{Z_g}$ $(N_{\text{grid}}\times K)$,

$\sigma_\beta, s_u, s_\varepsilon > 0$, $N_{\text{burn}}, N_{\text{kept}} \in \mathbb{N}$.

Decompose $\boldsymbol{X} = \boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})\boldsymbol{V_X}^T$ where $\boldsymbol{U_X}^T\boldsymbol{U_X} = \boldsymbol{V_X}^T\boldsymbol{V_X} = \boldsymbol{V_X}\boldsymbol{V_X}^T = \boldsymbol{I}_2$

Decompose $\boldsymbol{Z} = \boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})\boldsymbol{V_Z}^T$ where $\boldsymbol{U_Z}^T\boldsymbol{U_Z} = \boldsymbol{V_Z}^T\boldsymbol{V_Z} = \boldsymbol{V_Z}\boldsymbol{V_Z}^T = \boldsymbol{I}_K$

$\breve{\boldsymbol{X}} \longleftarrow \boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})$; $\breve{\boldsymbol{Z}} \longleftarrow \boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})$; $\boldsymbol{d}^2_{\boldsymbol{X}} \longleftarrow \boldsymbol{d_X}\odot\boldsymbol{d_X}$; $\boldsymbol{d}^2_{\boldsymbol{Z}} \longleftarrow \boldsymbol{d_Z}\odot\boldsymbol{d_Z}$

$\breve{\text{X}}\text{Ty} \longleftarrow \breve{\boldsymbol{X}}^T\boldsymbol{y}$; $\breve{\text{Z}}\text{Ty} \longleftarrow \breve{\boldsymbol{Z}}^T\boldsymbol{y}$; $\breve{\text{Z}}\text{T}\breve{\text{X}} \longleftarrow \breve{\boldsymbol{Z}}^T\breve{\boldsymbol{X}}$

Initialize: $\breve{\boldsymbol{u}}^{[0]}$ $(K\times 1)$, $(\sigma_\varepsilon^{-2})^{[0]}, b_\varepsilon, (\sigma_u^{-2})^{[0]}, b_u > 0$

For $s = 1,\ldots,N_{\text{burn}}+N_{\text{kept}}$:

$\boldsymbol{r}_{\breve{\beta}} \longleftarrow \breve{\text{X}}\text{Ty} - (\breve{\text{Z}}\text{T}\breve{\text{X}})^T\,\breve{\boldsymbol{u}}^{[s-1]}$; $\boldsymbol{\psi}_{\breve{\beta}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{d}^2_{\boldsymbol{X}} + \sigma_\beta^{-2}\boldsymbol{1}_2$

$\boldsymbol{z}_{\breve{\beta}} \sim N(\boldsymbol{0},\boldsymbol{I}_2)$; $\breve{\boldsymbol{\beta}}^{[s]} \longleftarrow \dfrac{\boldsymbol{z}_{\breve{\beta}}}{\sqrt{\boldsymbol{\psi}_{\breve{\beta}}}} + \dfrac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{r}_{\breve{\beta}}}{\boldsymbol{\psi}_{\breve{\beta}}}$

$\boldsymbol{r}_{\breve{u}} \longleftarrow \breve{\text{Z}}\text{Ty} - \breve{\text{Z}}\text{T}\breve{\text{X}}\,\breve{\boldsymbol{\beta}}^{[s]}$; $\boldsymbol{\psi}_{\breve{u}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{d}^2_{\boldsymbol{Z}} + (\sigma_u^{-2})^{[s-1]}\boldsymbol{1}_K$

$\boldsymbol{z}_{\breve{u}} \sim N(\boldsymbol{0},\boldsymbol{I}_K)$; $\breve{\boldsymbol{u}}^{[s]} \longleftarrow \dfrac{\boldsymbol{z}_{\breve{u}}}{\sqrt{\boldsymbol{\psi}_{\breve{u}}}} + \dfrac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{r}_{\breve{u}}}{\boldsymbol{\psi}_{\breve{u}}}$

$(\sigma_u^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(K+1), b_u + \tfrac{1}{2}\|\breve{\boldsymbol{u}}^{[s]}\|^2\Big)$; $b_u \longleftarrow \text{Gamma}\Big(1, (\sigma_u^{-2})^{[s]} + s_u^{-2}\Big)$

$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(n+1), b_\varepsilon + \tfrac{1}{2}\Big\|\boldsymbol{y} - \breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}}^{[s]} - \breve{\boldsymbol{Z}}\breve{\boldsymbol{u}}^{[s]}\Big\|^2\Big)$

$b_\varepsilon \longleftarrow \text{Gamma}\Big(1, (\sigma_\varepsilon^{-2})^{[s]} + s_\varepsilon^{-2}\Big)$

$\breve{\boldsymbol{X}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{X_g}\boldsymbol{V_X}$; $\breve{\boldsymbol{Z}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{Z_g}\boldsymbol{V_Z}$

For $s = 1,\ldots,N_{\text{kept}}$:

$\widehat{\boldsymbol{f}}_{\boldsymbol{g}}^{[s]} \longleftarrow \breve{\boldsymbol{X}}_{\boldsymbol{g}}\breve{\boldsymbol{\beta}}^{[s+N_{\text{burn}}]} + \breve{\boldsymbol{Z}}_{\boldsymbol{g}}\breve{\boldsymbol{u}}^{[s+N_{\text{burn}}]}$

$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow (\sigma_\varepsilon^{-2})^{[s+N_{\text{burn}}]}$; $(\sigma_u^{-2})^{[s]} \longleftarrow (\sigma_u^{-2})^{[s+N_{\text{burn}}]}$

Outputs: $\big\{\widehat{\boldsymbol{f}}_{\boldsymbol{g}}^{[s]}, (\sigma_\varepsilon^{-2})^{[s]}, (\sigma_u^{-2})^{[s]} : 1\le s\le N_{\text{kept}}\big\}$

Algorithm 2: *An orthogonalized design matrices speed-up of Algorithm 1 for the Bayesian nonparametric regression model (2).*

Noting that

$$\text{Cov}(\breve{\boldsymbol{u}}|\sigma_u^2) = \text{Cov}(\boldsymbol{V_Z}^T\boldsymbol{u}|\sigma_u^2) = \boldsymbol{V_Z}^T\sigma_u^2\boldsymbol{I}_K\,\boldsymbol{V_Z} = \sigma_u^2\boldsymbol{I}_K$$

and applying the same logic to $\boldsymbol{X}$ and $\boldsymbol{\beta}$, model (2) is equivalent to an alternative formulation for which the first line is replaced by

$$\boldsymbol{y}|\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}},\sigma_\varepsilon^2 \sim N(\breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}} + \breve{\boldsymbol{Z}}\,\breve{\boldsymbol{u}}, \sigma_\varepsilon^2\,\boldsymbol{I}_n), \quad \breve{\boldsymbol{\beta}} \sim N(\boldsymbol{0},\sigma_\beta^2\boldsymbol{I}_2), \quad \breve{\boldsymbol{u}}|\sigma_u^2 \sim N(\boldsymbol{0},\sigma_u^2\boldsymbol{I}_K).$$

However, $\breve{\boldsymbol{Z}}$ is such that

$$\breve{\boldsymbol{Z}}^T\breve{\boldsymbol{Z}} = \big\{\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})\big\}^T\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z}) = \text{diag}(\boldsymbol{d_Z})\boldsymbol{U_Z}^T\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z}) = \text{diag}(\boldsymbol{d_Z}\odot\boldsymbol{d_Z}) \qquad (5)$$

which implies that the columns of $\breve{\boldsymbol{Z}}$ are orthogonal vectors in $\mathbb{R}^n$. Use of $\breve{\boldsymbol{Z}}$ is loosely related to the so-called *Demmler-Reinsch basis* version of smoothing splines (Demmler & Reinsch, 1975), but is simply a linear transformation of $\boldsymbol{Z}$ based on (3) and (4). The full conditional distribution of $\breve{\boldsymbol{u}}$ is

$$\breve{\boldsymbol{u}}|\text{rest} \sim N\big(\boldsymbol{\Psi}_{\breve{u}}^{-1}\sigma_\varepsilon^{-2}\boldsymbol{r}_{\breve{u}}, \boldsymbol{\Psi}_{\breve{u}}^{-1}\big) \quad \text{where} \quad \boldsymbol{r}_{\breve{u}} \equiv \breve{\boldsymbol{Z}}^T(\boldsymbol{y} - \breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}})$$

$$\text{and} \quad \boldsymbol{\Psi}_{\breve{u}} \equiv \text{diag}\Big(\sigma_\varepsilon^{-2}(\boldsymbol{d_Z}\odot\boldsymbol{d_Z}) + \sigma_u^{-2}\boldsymbol{1}_K\Big)$$

The fact that $\Psi_{\breve{u}}$ is a diagonal matrix implies that the need for $K \times K$ matrix inversion is avoided when obtaining draws from the full conditional distribution of $\breve{\boldsymbol{u}}$. In view of Result 1, the following $O(K)$ steps lead to a draw from $\breve{\boldsymbol{u}}|\text{rest}$:

$$\boldsymbol{\psi}_{\breve{u}} \longleftarrow \sigma_\varepsilon^{-2}(\boldsymbol{d}_{\boldsymbol{Z}} \odot \boldsymbol{d}_{\boldsymbol{Z}}) + \sigma_u^{-2}\mathbf{1}_K \;\; ; \;\; \boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{I}_K) \;\; ; \;\; \breve{\boldsymbol{u}} \longleftarrow \frac{\boldsymbol{z}}{\sqrt{\boldsymbol{\psi}_{\breve{u}}}} + \frac{\sigma_\varepsilon^{-2}\boldsymbol{r}_{\breve{u}}}{\boldsymbol{\psi}_{\breve{u}}}.$$

Similar statements apply to the $\breve{\boldsymbol{\beta}}$ full conditional draws. Since the construction of $\breve{\boldsymbol{Z}}$ is done outside of the Gibbs sampling loop there is, as a function of K, a two orders of magnitude reduction in computation realized by working with the orthogonalized design matrices $\breve{\boldsymbol{X}}$ and $\breve{\boldsymbol{Z}}$.

Algorithm 2 describes this faster orthogonalized design matrices alternative to Algorithm 1.

2.2 Bernoulli Responses with Probit Link

Now suppose that the y_i are binary: $y_i \in \{0, 1\}$. Then an appropriate alternative to (1) is the probit link nonparametric regression model

$$y_i \stackrel{\text{ind.}}{\sim} \text{Bernoulli}\big(\Phi\big(f(x_i)\big)\big), \quad 1 \le i \le n, \tag{6}$$

Following Albert & Chib (1993), a useful adaptation of (2) for fitting (6), involving the auxiliary variables vector $\boldsymbol{\alpha}$, is

$$\begin{gathered} \boldsymbol{\alpha}\ (n \times 1) \text{ has } i\text{th entry } \alpha_i \text{ such that } y_i = I\big(\alpha_i \ge 0\big), \quad 1 \le i \le n, \\ \boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{u} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\,\boldsymbol{u}, \boldsymbol{I}_n), \quad \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_2), \quad \boldsymbol{u}|\sigma_u^2 \sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{I}_K), \\ \sigma_u^{-2}|b_u \sim \text{Gamma}\big(\tfrac{1}{2}, b_u\big), \quad b_u \sim \text{Gamma}\big(\tfrac{1}{2}, s_u^{-2}\big), \end{gathered} \tag{7}$$

Gibbs sampling for (7) proceeds similarly to that laid out in Algorithms 1 and 2. The main addition is due to the distributional result

$$(2y_i - 1)\,\alpha_i\big|\text{rest} \sim \text{Truncated-Normal}_+\Big((2y_i - 1)(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u})_i, 1\Big), \quad 1 \le i \le n,$$

where the random variable x has a Truncated-Normal$_+$ distribution with location parameter μ and scale parameter σ, written $x \sim \text{Truncated-Normal}_+(\mu, \sigma^2)$, if and only if the density function of x is

$$\mathfrak{p}(x) = \frac{\exp\{-(x-\mu)^2/(2\sigma^2)\}I(x > 0)}{\sigma\Phi(\mu/\sigma)\sqrt{2\pi}}.$$

Robert (1995) describes methodology for the efficient generation of Truncated-Normal$_+$ random variables. Algorithm 3 is a resultant orthogonalized design matrices Gibbs sampling algorithm.

2.3 Other Links and Response Types

For binary response Bayesian regression models with non-probit links and other response types, such as counts, the full conditional distributions involve weighted forms, which nullify the advantage of orthogonalized design matrices. Consider, for example, the Bayesian penalized spline model with logit link and Pólya-Gamma augmentation:

$$\begin{gathered} y_i|\boldsymbol{\beta}, \boldsymbol{u} \stackrel{\text{ind.}}{\sim} \text{Bernoulli}\big(\text{expit}\{(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\,\boldsymbol{u})_i\}\big), \quad \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_2), \quad \boldsymbol{u}|\sigma_u^2 \sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{I}_K), \\ \sigma_u^{-2}|b_u \sim \text{Gamma}\big(\tfrac{1}{2}, b_u\big), \quad b_u \sim \text{Gamma}\big(\tfrac{1}{2}, s_u^{-2}\big), \\ \alpha_i|\boldsymbol{\beta}, \boldsymbol{u} \stackrel{\text{ind.}}{\sim} \text{Pólya-Gamma}(1, (\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\,\boldsymbol{u})_i), \quad 1 \le i \le n. \end{gathered} \tag{8}$$

Inputs: $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times 2)$, $\boldsymbol{Z}$ $(n \times K)$, $\boldsymbol{X_g}$ $(N_{\text{grid}} \times 2)$, $\boldsymbol{Z_g}$ $(N_{\text{grid}} \times K)$

$\sigma_\beta, s_u, s_\varepsilon > 0$, $N_{\text{burn}}, N_{\text{kept}} \in \mathbb{N}$.

Decompose $\boldsymbol{X} = \boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})\boldsymbol{V_X}^T$ where $\boldsymbol{U_X}^T\boldsymbol{U_X} = \boldsymbol{V_X}^T\boldsymbol{V_X} = \boldsymbol{V_X}\boldsymbol{V_X}^T = \boldsymbol{I}_2$

Decompose $\boldsymbol{Z} = \boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})\boldsymbol{V_Z}^T$ where $\boldsymbol{U_Z}^T\boldsymbol{U_Z} = \boldsymbol{V_Z}^T\boldsymbol{V_Z} = \boldsymbol{V_Z}\boldsymbol{V_Z}^T = \boldsymbol{I}_K$

$\breve{\boldsymbol{X}} \longleftarrow \boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})$; $\breve{\boldsymbol{Z}} \longleftarrow \boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})$; $\boldsymbol{d_X^2} \longleftarrow \boldsymbol{d_X} \odot \boldsymbol{d_X}$; $\boldsymbol{d_Z^2} \longleftarrow \boldsymbol{d_Z} \odot \boldsymbol{d_Z}$

$\breve{\text{Z}}\text{T}\breve{\text{X}} \longleftarrow \breve{\boldsymbol{Z}}^T\breve{\boldsymbol{X}}$

Initialize: $\breve{\boldsymbol{u}}^{[0]}$ $(K \times 1)$, $(\sigma_u^{-2})^{[0]}, b_u > 0$, $\breve{\text{X}}\text{T}\alpha$ (2×1), $\breve{\text{Z}}\text{T}\alpha$ $(K \times 1)$

For $s = 1, \ldots, N_{\text{burn}} + N_{\text{kept}}$:

$\boldsymbol{\omega}_{\breve{\beta}} \longleftarrow \breve{\text{X}}\text{T}\alpha - (\breve{\text{Z}}\text{T}\breve{\text{X}})^T\,\breve{\boldsymbol{u}}^{[s-1]}$; $\boldsymbol{\psi}_{\breve{\beta}} \longleftarrow \boldsymbol{d_X^2} + \sigma_\beta^{-2}\mathbf{1}_2$

$\boldsymbol{z}_{\breve{\beta}} \sim N(\mathbf{0}, \boldsymbol{I}_2)$; $\breve{\boldsymbol{\beta}}^{[s]} \longleftarrow \dfrac{\boldsymbol{z}_{\breve{\beta}}}{\sqrt{\boldsymbol{\psi}_{\breve{\beta}}}} + \dfrac{\boldsymbol{\omega}_{\breve{\beta}}}{\boldsymbol{\psi}_{\breve{\beta}}}$

$\boldsymbol{\omega}_{\breve{u}} \longleftarrow \breve{\text{Z}}\text{T}\alpha - \breve{\text{Z}}\text{T}\breve{\text{X}}\,\breve{\boldsymbol{\beta}}^{[s]}$; $\boldsymbol{\psi}_{\breve{u}} \longleftarrow \boldsymbol{d_Z^2} + (\sigma_u^{-2})^{[s-1]}\mathbf{1}_K$

$\boldsymbol{z}_{\breve{u}} \sim N(\mathbf{0}, \boldsymbol{I}_K)$; $\breve{\boldsymbol{u}}^{[s]} \longleftarrow \dfrac{\boldsymbol{z}_{\breve{u}}}{\sqrt{\boldsymbol{\psi}_{\breve{u}}}} + \dfrac{\boldsymbol{\omega}_{\breve{u}}}{\boldsymbol{\psi}_{\breve{u}}}$

$(\sigma_u^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(K+1), b_u + \tfrac{1}{2}\|\breve{\boldsymbol{u}}^{[s]}\|^2\Big)$; $b_u \longleftarrow \text{Gamma}\Big(1, (\sigma_u^{-2})^{[s]} + s_u^{-2}\Big)$

For $i = 1, \ldots, n$:

$\zeta \sim \text{Truncated-Normal}_+\big((2y_i - 1)(\breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}}^{[s]} + \breve{\boldsymbol{Z}}\breve{\boldsymbol{u}}^{[s]})_i, 1\big)$; $\alpha_i \longleftarrow (2y_i - 1)\zeta$

$\breve{\text{X}}\text{T}\alpha \longleftarrow \breve{\boldsymbol{X}}^T\boldsymbol{\alpha}$; $\breve{\text{Z}}\text{T}\alpha \longleftarrow \breve{\boldsymbol{Z}}^T\boldsymbol{\alpha}$

$\breve{\boldsymbol{X}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{X_g}\boldsymbol{V_X}$; $\breve{\boldsymbol{Z}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{Z_g}\boldsymbol{V_Z}$

For $s = 1, \ldots, N_{\text{kept}}$:

$\widehat{\boldsymbol{\eta}}_{\boldsymbol{g}}^{[s]} \longleftarrow \breve{\boldsymbol{X}}_{\boldsymbol{g}}\breve{\boldsymbol{\beta}}^{[s+N_{\text{burn}}]} + \breve{\boldsymbol{Z}}_{\boldsymbol{g}}\breve{\boldsymbol{u}}^{[s+N_{\text{burn}}]}$; $(\sigma_u^{-2})^{[s]} \longleftarrow (\sigma_u^{-2})^{[s+N_{\text{burn}}]}$

Outputs: $\big\{\widehat{\boldsymbol{\eta}}_{\boldsymbol{g}}^{[s]}, (\sigma_u^{-2})^{[s]} : 1 \le s \le N_{\text{kept}}\big\}$

Algorithm 3: *An orthogonalized design matrices Gibbs sampling scheme for the Bayesian probit nonparametric regression model (7).*

Here $\text{expit}(x) \equiv 1/(1+e^{-x})$ and the Pólya-Gamma distribution is as defined in Polson *et al.* (2013). For model (8), the full conditional distribution of $\boldsymbol{u}$ is

$$\boldsymbol{u}|\text{rest} \sim N\big(\boldsymbol{\Psi}_u^{-1}\boldsymbol{r}_u, \boldsymbol{\Psi}_u^{-1}\big) \quad \text{where} \quad \boldsymbol{r}_u \equiv \boldsymbol{Z}^T\big\{\boldsymbol{y} - \tfrac{1}{2}\mathbf{1} - \boldsymbol{\alpha} \odot (\boldsymbol{X\beta})\big\},$$
$$\boldsymbol{\Psi}_u \equiv \boldsymbol{Z}^T\text{diag}(\boldsymbol{\alpha})\boldsymbol{Z} + \sigma_u^{-2}\boldsymbol{I}_K$$

and $\boldsymbol{\alpha}$ is the vector of α_i values. The generalization of (5) to this weighted case is

$$\breve{\boldsymbol{Z}}^T\text{diag}(\boldsymbol{\alpha})\breve{\boldsymbol{Z}} = \big\{\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})\big\}^T\text{diag}(\boldsymbol{\alpha})\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z}) = \text{diag}(\boldsymbol{d_Z})\boldsymbol{U_Z}^T\text{diag}(\boldsymbol{\alpha})\boldsymbol{U_Z}\text{diag}(\boldsymbol{d_Z})$$

which is not necessarily diagonal. Also, since the $\boldsymbol{\alpha}$ vector changes throughout the Gibbs sampling iterations there is no re-definition of $\breve{\boldsymbol{Z}}$ that leads to the precision matrix of $\boldsymbol{u}|\text{rest}$ having a fixed diagonal form as a function of σ_ε^2 and σ_u^2. Similar comments apply to non-Gibbsian approaches such as those involving Metropolis-Hastings schemes. Nevertheless, the Gaussian and Bernoulli response cases are ubiquitous in semiparametric regression and significantly speeding up their Bayesian analyses is worthwhile.

2.4 Operations Order of Magnitude Comparisons

Let $N_{\text{Gibbs}} \equiv N_{\text{burn}} + N_{\text{kept}}$ and assume that

$$n, K, N_{\text{Gibbs}}, N_{\text{grid}} \gg 1 \quad \text{with} \quad n \gg K. \tag{9}$$

Under (9) the singular value decomposition of an $n \times K$ matrix requires $O(nK^2)$ operations and the singular value decomposition of a $K \times K$ matrix requires $O(K^3)$ operations. This leads to the following total operations order of magnitude statements for each of Algorithms 1 and 2:

	Gibbs loop operations	other operations
Algorithm 1:	$O(N_{\text{Gibbs}} K^3)$	$O(nK^2) + O(N_{\text{kept}} N_{\text{grid}}\, K)$
Algorithm 2:	$O(N_{\text{Gibbs}} K)$	$O(nK^2) + O(N_{\text{grid}} K^2) + O(N_{\text{kept}} N_{\text{grid}}\, K)$

For the Gibbs loops operations it is clear that Algorithm 2 provides a two orders of magnitude improvement as a function of K. This comes at the price of the extra $O(nK^2)$ operations required to decompose $\boldsymbol{X}$ and $\boldsymbol{Z}$ prior to the Algorithm 2 Gibbs loop and the $O(N_{\text{grid}} K^2)$ step to compute $\breve{\boldsymbol{Z}}_{\boldsymbol{g}}$. For typical values of N_{Gibbs} and n the Gibbs loop speed-ups will outweigh the cost of these non-loop steps. The constants associated with the orders of magnitude also impact the relative speeds of the two algorithms. Next, in Section 2.5, we investigate the actual speed-ups via computer experiments.

2.5 Computer Experiments

We coded Algorithms 1 and 2 in the `C++` computer programming language and generated 100 replications of (1) for each of $n \in \{100, 200, 400\}$. The number of basis functions was varied over $K \in \{25, 50\}$ and the Gibbs sample sizes were fixed at $N_{\text{burn}} = N_{\text{kept}} = 1000$. The number of grid points was $N_{\text{grid}} = 101$. All calculations were performed on the first author's MacBook Air computer which has 24 gigabytes of random access memory and a 3.5 gigahertz processor.

Figure 1 displays side-by-side boxplots of the effective sample size per second ratios for the kept Gibbs sample of four quantities of interest: f evaluated at each of the population quantiles, $f(Q_k)$, $k = 1, 2, 3$, and the error standard deviation σ_ε. The ratio numerator corresponds to Algorithm 2. Effective sample sizes of Markov chain Monte Carlo samples are based on established approaches that account for loss of information due to autocorrelation. The particular version used here corresponds to the `monitor()` function within the R package rstan (Stan Development Team, 2025) with details given in that package's reference manual.

The side-by-side boxplots in Figure 1 reveal the considerable practical benefits of use of orthogonalized design matrices. For the vast majority of replications there are at least 20-fold improvements in effective sample size per second when the number of basis functions is $K = 25$. When the number of basis functions doubles to $K = 50$ the improvement also approximately doubles and Algorithm 2 is usually 30–60 times faster than Algorithm 1 in delivering the same quality of Gibbs samples. For larger n the improvement lessens slightly, which can be attributed to the $O(nK^2)$ singular value decomposition of $\boldsymbol{Z}$ prior in Algorithm 2 prior to the Gibbs sampling phase. Nevertheless, the improvement factors are still around 50. A more formal exemplification of the improvements is provided by Table 1, which lists the medians and 95% confidence intervals for the basis size: 50 and sample size: 200 panel of Figure 1.

Even though Table 1 gives formal evidence of a big statistically significant improvement due to use of orthogonalized design matrices, we contend that that side-by-side boxplot summaries better convey the improvements. The remaining computer studies will only use such summaries.

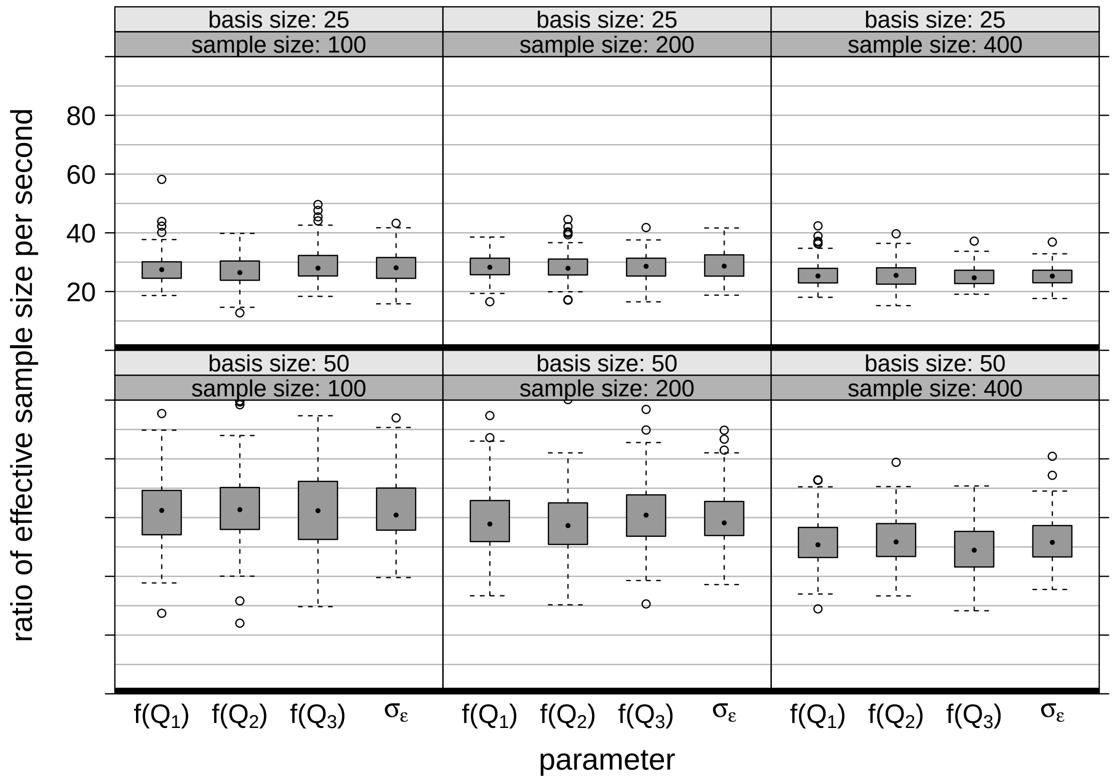

Figure 1: *Side-by-side boxplots of the effective sample size per second ratios for the computer experiment involving Bayesian nonparametric regression described in the text. The ratio numerator corresponds to Algorithm 2. The quantities of interest are $f(Q_k)$, $k = 1, 2, 3$, where Q_k is the kth population quantile of the predictor distribution and σ_ε is the error standard deviation. Each ratio corresponds to the effective sample size per second for the direct approach (Algorithm 2) divided by the same quantity for the orthogonalized design matrices approach (Algorithm 1). The thick horizontal lines correspond to a ratio of* 1.

3 Generalized Additive Models

We now consider the case of multiple predictors and generalized additive models extensions of Bayesian nonparametric regression. Throughout this section the data are assumed to be of the form

$$(\mathring{\boldsymbol{x}}_i, \dot{\boldsymbol{x}}_i, y_i), \quad 1 \le i \le n, \tag{10}$$

where y_i is the ith response observation. For each i, $\mathring{\boldsymbol{x}}_i$ is a vector of observations corresponding to the $d_\circ \times 1$ vector of predictors $\mathring{\boldsymbol{x}}$. The predictors in $\mathring{\boldsymbol{x}}$ are assumed to have linear impacts on the mean response which, for example, is appropriate for components of $\mathring{\boldsymbol{x}}$ that are indicator variables. The other predictors, corresponding to the $d_\bullet \times 1$ vector of predictors $\dot{\boldsymbol{x}}$, are such that its variables may have nonlinear impacts on the mean response. For each i, $\dot{\boldsymbol{x}}_i$ is a vector of observations corresponding to $\dot{\boldsymbol{x}}$.

To illustrate the notation defined in the previous paragraph, consider the Boston mortgages applications data described in Section 1.3.2 of Harezlak *et al.* (2018). These data consist of several variables for $n = 2,380$ mortgage applications in Boston, U.S.A. An example probit additive model for these data is

$$\texttt{deny}_i \overset{\text{ind.}}{\sim} \text{Bernoulli}\Big(\Phi\big(\beta_0 + \beta_1\,\texttt{self-employed}_i + \beta_2\,\texttt{single}_i + \beta_3\,\texttt{condominium}_i + f_1(\texttt{DIR}_i) + f_2(\texttt{LVR}_i)\big)\Big), \quad 1 \le i \le 2,380$$

	$f(Q_1)$	$f(Q_2)$	$f(Q_3)$	σ_ε
CI low.	56.2	55.6	58.7	57.6
median	57.8	57.3	60.8	58.2
CI upp.	60.7	60.2	63.2	61.4

Table 1: *Medians and 95% Wilcoxon confidence interval (CI) lower and upper limits based on ratio of effective sample sizes per second data corresponding to the basis size: 50 and sample size: 200 panel of Figure 1.*

where $\texttt{deny}_i$ equals 1 if the ith mortgage application was denied and 0 otherwise. The predictor observations $\texttt{self-employed}_i$, $\texttt{single}_i$ and $\texttt{condominium}_i$ correspond to similarly defined indicator variables for whether the applicant is self-employed, the applicant is single and the property is a condominium, respectively. Lastly, $\texttt{DIR}_i$ and $\texttt{LVR}_i$ are, respectively, the debt to income ratio and the loan to property value ratio for the ith application. For this example, $d_\circ = 3$, $d_\bullet = 2$ and the vectors in (10) are

$$\mathring{\boldsymbol{x}}_i = \begin{bmatrix} \texttt{self-employed}_i \\ \texttt{single}_i \\ \texttt{condominium}_i \end{bmatrix} \quad \text{and} \quad \dot{\boldsymbol{x}}_i = \begin{bmatrix} \texttt{DIR}_i \\ \texttt{LVR}_i \end{bmatrix}.$$

In addition, $y_i = \texttt{deny}_i$.

The generic forms of the generalized additive models considered in this section are

$$y_i \sim \begin{cases} N(\eta_i, \sigma_\varepsilon^2), & y_i \text{ continuous}, \\ \text{Bernoulli}\big(\Phi(\eta_i)\big), & y_i \text{ binary}, \end{cases}$$

where

$$\eta_i \equiv \beta_0 + \sum_{j=1}^{d_\circ} \beta_j \mathring{x}_{ji} + \sum_{j=1}^{d_\bullet} f_j(\dot{x}_{ji}) \quad \text{with} \quad f_j(\dot{x}_{ji}) \equiv \beta_{d_\circ + j}\, \dot{x}_{ji} + \sum_{k=1}^{K_j} u_{jk} z_{jk}(\dot{x}_{ji})$$

being a penalized spline model for the jth predictor in $\dot{\boldsymbol{x}}$. The u_{jk} and z_{jk} notation is analogous to that used in Section 2, with the addition of the j subscript to denote the jth predictor entering the model non-linearly.

The design matrices for this generalized additive model set-up are

$$\boldsymbol{X} \equiv \big[1\; \mathring{\boldsymbol{x}}_i^T\; \dot{\boldsymbol{x}}_i^T\big]_{1 \le i \le n} \quad \text{and} \quad \boldsymbol{Z}_j \equiv \big[\, z_{jk}(\dot{x}_{ji})\big]_{\substack{1 \le i \le n, \\ 1 \le k \le K}} \quad 1 \le j \le d_\bullet.$$

A Bayesian Gaussian response generalized additive model is

$$\begin{gathered} \boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{d_\bullet}, \sigma_\varepsilon^2 \sim N\Big(\boldsymbol{X\beta} + \sum_{j=1}^{d_\bullet} \boldsymbol{Z}_j\, \boldsymbol{u}_j, \sigma_\varepsilon^2\, \boldsymbol{I}_n\Big), \quad \boldsymbol{\beta} \sim N\big(\boldsymbol{0}, \sigma_\beta^2 \boldsymbol{I}_{1+d_\circ+d_\bullet}\big), \\ \boldsymbol{u}_j|\sigma_{uj}^2 \overset{\text{ind.}}{\sim} N(\boldsymbol{0}, \sigma_{uj}^2 \boldsymbol{I}_{K_j}),\ \sigma_{uj}^{-2}|b_{uj} \overset{\text{ind.}}{\sim} \text{Gamma}\big(\tfrac{1}{2}, b_{uj}\big),\ b_{uj} \overset{\text{ind.}}{\sim} \text{Gamma}\big(\tfrac{1}{2}, s_u^{-2}\big), \\ 1 \le j \le d_\bullet, \quad \sigma_\varepsilon^{-2}|b_\varepsilon \sim \text{Gamma}\big(\tfrac{1}{2}, b_\varepsilon\big), \quad b_\varepsilon \sim \text{Gamma}\big(\tfrac{1}{2}, s_\varepsilon^{-2}\big). \end{gathered} \tag{11}$$

The binary response case involves replacement of the first distributional statement in (11) by

$$\begin{gathered} \boldsymbol{\alpha}\ (n \times 1) \text{ has } i\text{th entry } \alpha_i \text{ such that } y_i = I\big(\alpha_i \ge 0\big), \quad 1 \le i \le n, \\ \boldsymbol{\alpha}|\boldsymbol{\beta}, \boldsymbol{u}_1, \ldots, \boldsymbol{u}_{d_\bullet} \sim N\Big(\boldsymbol{X\beta} + \sum_{j=1}^{d_\bullet} \boldsymbol{Z}_j\, \boldsymbol{u}_j, \boldsymbol{I}_n\Big) \end{gathered} \tag{12}$$

and removal of the σ_ε^2 and b_ε variables. An appropriate orthogonalized design matrix reparametrization of the linear predictor vector is

$$\boldsymbol{X}\boldsymbol{\beta}+\sum_{j=1}^{d_\bullet}\boldsymbol{Z}_j\,\boldsymbol{u}_j=\breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}}+\sum_{j=1}^{d_\bullet}\breve{\boldsymbol{Z}}_j\,\breve{\boldsymbol{u}}_j$$

where $\breve{\boldsymbol{X}}$ and $\breve{\boldsymbol{\beta}}$ are defined as in Section 2. For each $1\le j\le d_\bullet$,

$$\breve{\boldsymbol{Z}}_j\equiv\boldsymbol{U}_{\boldsymbol{Z}_j}\mathrm{diag}(\boldsymbol{d}_{\boldsymbol{Z}_j})\quad\text{and}\quad\breve{\boldsymbol{u}}_j\equiv\boldsymbol{V}_{\boldsymbol{Z}_j}^T\boldsymbol{u}_j\quad\text{where}\quad\boldsymbol{Z}_j=\boldsymbol{U}_{\boldsymbol{Z}_j}\mathrm{diag}(\boldsymbol{d}_{\boldsymbol{Z}_j})\boldsymbol{V}_{\boldsymbol{Z}_j}^T$$

is the singular value decomposition of $\boldsymbol{Z}_j$ with $\boldsymbol{U}_{\boldsymbol{Z}_j}$ being $n\times K_j$, $\boldsymbol{d}_{\boldsymbol{Z}_j}$ being $K_j\times 1$ and $\boldsymbol{V}_{\boldsymbol{Z}_j}$ being $K_j\times K_j$ such that $\boldsymbol{U}_{\boldsymbol{Z}_j}^T\boldsymbol{U}_{\boldsymbol{Z}_j}=\boldsymbol{V}_{\boldsymbol{Z}_j}^T\boldsymbol{V}_{\boldsymbol{Z}_j}=\boldsymbol{V}_{\boldsymbol{Z}_j}\boldsymbol{V}_{\boldsymbol{Z}_j}^T=\boldsymbol{I}_{K_j}$.

We are now ready to list Algorithm 4, which describes Gibbs sampling for fitting the above Bayesian generalized additive model with orthogonalized design matrices speed-ups. It uses the following notation:

$$\begin{aligned}
&K_j\equiv\text{the number of columns in }\breve{\boldsymbol{Z}}_j,\ 1\le j\le d_\bullet,\\
&\mathfrak{c}\text{ is the }(d_\bullet+1)\times 1\text{vector with entries }\mathfrak{c}_1\equiv 0\text{and }\mathfrak{c}_{j+1}\equiv\textstyle\sum_{k=1}^{j}K_k,\ 1\le j\le d_\bullet,\\
&\breve{\mathrm{Z}}\mathrm{Ty}_{\mathrm{adj}}^{\langle j\rangle}\equiv\text{the sub-block of }\breve{\mathrm{Z}}\mathrm{Ty}_{\mathrm{adj}}\text{ corresponding to rows }(\mathfrak{c}_j+1)\text{ to }\mathfrak{c}_{j+1},\ 1\le j\le d_\bullet,\\
&\breve{\mathrm{Z}}\mathrm{T}\breve{\mathrm{X}}^{\langle j\rangle}\equiv\text{the sub-block of }\breve{\mathrm{Z}}\mathrm{T}\breve{\mathrm{X}}\text{ corresponding to rows }(\mathfrak{c}_j+1)\text{ to }\mathfrak{c}_{j+1},\ 1\le j\le d_\bullet,\\
&\breve{\mathrm{Z}}\mathrm{T}\breve{\mathrm{Z}}^{\langle j,j'\rangle}\equiv\text{the sub-block of }\breve{\mathrm{Z}}\mathrm{T}\breve{\mathrm{Z}}\text{ corresponding to rows }(\mathfrak{c}_j+1)\text{ to }\mathfrak{c}_{j+1},\\
&\qquad\qquad\text{and columns }(\mathfrak{c}_{j'}+1)\text{ to }\mathfrak{c}_{j'+1},\ 1\le j,j'\le d_\bullet.\\
&\boldsymbol{X}_{\boldsymbol{g}}\text{ and }\boldsymbol{Z}_{j,\boldsymbol{g}},1\le j\le d_\bullet,\text{ are grid-wise versions of the design for plotting grids of}\\
&\qquad\qquad\text{size }N_{\mathrm{grid}},\\
&\mathring{\boldsymbol{\beta}}\equiv\text{the }d_\circ\times 1\text{ vector corresponding to the coefficients of the }\mathring{\boldsymbol{x}}_i.
\end{aligned}\tag{13}$$

We ran a computer experiment to compare the practical performance of Algorithm 4 with the direct computation alternative in the case of Gaussian responses. The sample sizes ranged over $n\in\{100,200,400\}$ and the number of predictors entering the model non-linearly ranged over $d_\bullet\in\{2,4,8,16\}$. The number of predictors entering the model linearly was fixed at $d_\circ=0$ and number of basis functions for each predictor was fixed at $K=25$. The chains that were monitored are the vertical slice corresponding to the population median of all $d_\bullet$ predictors, which we denote by $f(Q_2)$, and the error standard deviation σ_ε. Figure 2 summarises the effective sample size per second ratios using side-by-side boxplots. It shows that effective sample sizes per second are about 5–20 times larger when the orthogonalized design matrices approach is used. The advantage tends to decrease for larger models but there are still around 5-fold improvements for generalized additive models with 16 predictors.

4 Group-Specific Curves Models

The group-specific curves models (e.g. Donnelly *et al.* 1995) that we consider here are based on grouped data of the form

$$(x_{ij},y_{ij}),\quad 1\le j\le n_i,\ 1\le i\le m,$$

where, for example, x_{ij} is the jth predictor measurement within the ith group. The number of groups is m. The models have generic forms

$$y_{ij}\sim\begin{cases}N\big(f(x_{ij})+g_i(x_{ij}),\sigma_\varepsilon^2\big), & y_{ij}\text{ continuous},\\ \text{Bernoulli}\big(\Phi\big(f(x_{ij})+g_i(x_{ij})\big)\big), & y_{ij}\text{ binary},\end{cases}\tag{14}$$

Inputs: $\boldsymbol{y}\ (n\times 1)$, $\boldsymbol{X}\ \big(n\times(1+d_\circ+d_\bullet)\big)$, $\boldsymbol{Z}_1\ (n\times K_1),\ldots,\boldsymbol{Z}_{d_\bullet}\ (n\times K_{d_\bullet})$,
$\boldsymbol{X_g}\ \big(N_{\text{grid}}\times(1+d_\circ+d_\bullet)\big)$, $\boldsymbol{Z}_{1,\boldsymbol{g}}\ (N_{\text{grid}}\times K_1),\ldots,\boldsymbol{Z}_{d_\bullet,\boldsymbol{g}}\ (N_{\text{grid}}\times K_{d_\bullet})$,
$\sigma_\beta, s_u, s_\varepsilon > 0$, $N_{\text{burn}}, N_{\text{kept}}\in\mathbb{N}$, **responseType** $\in\{\text{Gaussian}, \text{Bernoulli}\}$.

Decompose $\boldsymbol{X}=\boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})\boldsymbol{V_X}^T$ where $\boldsymbol{U_X}^T\boldsymbol{U_X}=\boldsymbol{V_X}^T\boldsymbol{V_X}=\boldsymbol{V_X}\boldsymbol{V_X}^T=\boldsymbol{I}_{1+d_\circ+d_\bullet}$

$\breve{\boldsymbol{X}}\longleftarrow\boldsymbol{U_X}\text{diag}(\boldsymbol{d_X})$; $\boldsymbol{d}_{\boldsymbol{X}}^2\longleftarrow\boldsymbol{d_X}\odot\boldsymbol{d_X}$; $\breve{\text{X}}\text{Ty}_{\text{adj}}\longleftarrow\breve{\boldsymbol{X}}^T\boldsymbol{y}$

For $j=1,\ldots,d_\bullet$:

Decompose $\boldsymbol{Z}_j=\boldsymbol{U}_{\boldsymbol{Z}j}\text{diag}(\boldsymbol{d}_{\boldsymbol{Z}j})\boldsymbol{V}_{\boldsymbol{Z}j}^T$ where $\boldsymbol{U}_{\boldsymbol{Z}j}^T\boldsymbol{U}_{\boldsymbol{Z}j}=\boldsymbol{V}_{\boldsymbol{Z}j}^T\boldsymbol{V}_{\boldsymbol{Z}j}=\boldsymbol{V}_{\boldsymbol{Z}j}\boldsymbol{V}_{\boldsymbol{Z}j}^T=\boldsymbol{I}_{K_j}$

$\breve{\boldsymbol{Z}}_j\longleftarrow\boldsymbol{U}_{\boldsymbol{Z}j}\text{diag}(\boldsymbol{d}_{\boldsymbol{Z}j})$; $\boldsymbol{d}_{\boldsymbol{Z}j}^2\longleftarrow\boldsymbol{d}_{\boldsymbol{Z}j}\odot\boldsymbol{d}_{\boldsymbol{Z}j}$

$\breve{\boldsymbol{Z}}\longleftarrow[\breve{\boldsymbol{Z}}_1\cdots\breve{\boldsymbol{Z}}_{d_\bullet}]$; $\breve{\text{Z}}\text{Ty}_{\text{adj}}\longleftarrow\breve{\boldsymbol{Z}}^T\boldsymbol{y}$; $\breve{\text{Z}}\text{T}\breve{\text{X}}\longleftarrow\breve{\boldsymbol{Z}}^T\breve{\boldsymbol{X}}$; $\breve{\text{Z}}\text{T}\breve{\text{Z}}\longleftarrow\breve{\boldsymbol{Z}}^T\breve{\boldsymbol{Z}}$

Initialize: $(\sigma_\varepsilon^{-2})^{[0]}, b_\varepsilon>0$; For each $1\le j\le d_\bullet$: $\breve{\boldsymbol{u}}^{[j]}\ (K_j\times 1)$, $(\sigma_{uj}^{-2})^{[0]}, b_{uj}>0$

For $s=1,\ldots,N_{\text{burn}}+N_{\text{kept}}$:

$$\boldsymbol{r}_{\breve{\beta}}\longleftarrow\breve{\text{X}}\text{Ty}_{\text{adj}}-\sum_{j=1}^{d_\bullet}\breve{\text{Z}}\text{T}\breve{\text{X}}^{\langle j\rangle T}\,\breve{\boldsymbol{u}}_j^{[s-1]}\ ;\ \boldsymbol{\psi}_{\breve{\beta}}\longleftarrow(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{d}_{\boldsymbol{X}}^2+\sigma_\beta^{-2}\boldsymbol{1}_{1+d_\circ+d_\bullet}$$

$$\boldsymbol{z}_{\breve{\beta}}\sim N(\boldsymbol{0},\boldsymbol{I}_{1+d_\circ+d_\bullet})\ ;\ \breve{\boldsymbol{\beta}}^{[s]}\longleftarrow\frac{\boldsymbol{z}_{\breve{\beta}}}{\sqrt{\boldsymbol{\psi}_{\breve{\beta}}}}+\frac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{r}_{\breve{\beta}}}{\boldsymbol{\psi}_{\breve{\beta}}}$$

For $j=1,\ldots,d_\bullet$: $\breve{\boldsymbol{u}}_j^{\text{curr}}\longleftarrow\breve{\boldsymbol{u}}_j^{[s-1]}$

For $j=1,\ldots,d_\bullet$:

$$\boldsymbol{r}_{\breve{u}j}\longleftarrow\breve{\text{Z}}\text{Ty}_{\text{adj}}^{\langle j\rangle}-\breve{\text{Z}}\text{T}\breve{\text{X}}^{\langle j\rangle}\,\breve{\boldsymbol{\beta}}^{[s]}-\sum_{j'\neq j}^{d_\bullet}\breve{\text{Z}}\text{T}\breve{\text{Z}}^{\langle j,j'\rangle}\breve{\boldsymbol{u}}_{j'}^{\text{curr}}$$

$$\boldsymbol{\psi}_{\breve{u}j}\longleftarrow(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{d}_{\boldsymbol{Z}j}^2+(\sigma_{uj}^{-2})^{[s-1]}\boldsymbol{1}_{K_j}$$

$$\boldsymbol{z}_{\breve{u}j}\sim N(\boldsymbol{0},\boldsymbol{I}_{K_j})\ ;\ \breve{\boldsymbol{u}}_j^{\text{curr}}\longleftarrow\frac{\boldsymbol{z}_{\breve{u}j}}{\sqrt{\boldsymbol{\psi}_{\breve{u}j}}}+\frac{(\sigma_\varepsilon^{-2})^{[s-1]}\boldsymbol{r}_{\breve{u}j}}{\boldsymbol{\psi}_{\breve{u}j}}$$

For $j=1,\ldots,d_\bullet$: $\breve{\boldsymbol{u}}_j^{[s]}\longleftarrow\breve{\boldsymbol{u}}_j^{\text{curr}}$

For $j=1,\ldots,d_\bullet$:

$$(\sigma_{uj}^{-2})^{[s]}\longleftarrow\text{Gamma}\Big(\tfrac{1}{2}(K_j+1), b_{uj}+\tfrac{1}{2}\big\|\breve{\boldsymbol{u}}_j^{[s]}\big\|^2\Big)\ ;\ b_{uj}\longleftarrow\text{Gamma}\Big(1,(\sigma_{uj}^{-2})^{[s]}+s_u^{-2}\Big)$$

$$\boldsymbol{\eta}\longleftarrow\breve{\boldsymbol{X}}\breve{\boldsymbol{\beta}}^{[s]}+\sum_{j=1}^{d_\bullet}\breve{\boldsymbol{Z}}_j\breve{\boldsymbol{u}}_j^{[s]}$$

If **responseType** is Gaussian then

$$(\sigma_\varepsilon^{-2})^{[s]}\longleftarrow\text{Gamma}\Big(\tfrac{1}{2}(n+1), b_\varepsilon+\tfrac{1}{2}\|\boldsymbol{y}-\boldsymbol{\eta}\|^2\Big)\ ;\ b_\varepsilon\longleftarrow\text{Gamma}\Big(1,(\sigma_\varepsilon^{-2})^{[s]}+s_\varepsilon^{-2}\Big)$$

If **responseType** is Bernoulli then

$(\sigma_\varepsilon^{-2})^{[s]}\longleftarrow 1$; $\zeta\sim\text{Truncated-Normal}_+\big((2y_i-1)\eta_i,1\big)$; $\alpha_i\longleftarrow(2y_i-1)\zeta$

$\breve{\text{X}}\text{Ty}_{\text{adj}}\longleftarrow\breve{\boldsymbol{X}}^T\boldsymbol{\alpha}$; $\breve{\text{Z}}\text{Ty}_{\text{adj}}\longleftarrow\breve{\boldsymbol{Z}}^T\boldsymbol{\alpha}$

continued on a subsequent page . . .

Algorithm 4: *An orthogonalized design matrices Gibbs sampling scheme for the Bayesian generalized additive models (11) and (12).*

for smooth functions f and g_i, $1\le i\le m$. The function f models the global mean response, whereas g_i models the deviation from the global mean response for group i.

As in this article's previous sections, mixed model-based penalized splines can be used to formulate a hierarchical Bayesian model for (14). For $1\le i\le m$ define

$$\boldsymbol{x}_i\equiv[x_{ij}]_{1\le j\le n_i},\quad \boldsymbol{y}_i\equiv[y_{ij}]_{1\le j\le n_i},\quad \boldsymbol{X}_i\equiv[\boldsymbol{1}_{n_i}\ \boldsymbol{x}_i],$$

$\breve{\boldsymbol{X}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{V}_{\boldsymbol{X}}$; For $j = 1, \ldots, d_\bullet$: $\breve{\boldsymbol{Z}}_{j,\boldsymbol{g}} \longleftarrow \boldsymbol{Z}_{j,\boldsymbol{g}} \boldsymbol{V}_{\boldsymbol{Z}j}$

For $s = 1, \ldots, N_{\text{kept}}$:

$\breve{\boldsymbol{\beta}}_{\text{temp}} \longleftarrow$ entries 2 to $(1 + d_\circ)$ of $\breve{\boldsymbol{\beta}}^{[s+N_{\text{burn}}]}$; $\mathring{\boldsymbol{\beta}}^{[s]} \longleftarrow \boldsymbol{V}_{\boldsymbol{X}} \breve{\boldsymbol{\beta}}_{\text{temp}}$

For $j = 1, \ldots, d_\bullet$:

$\boldsymbol{x}_{\boldsymbol{g}}^{\text{curr}} \longleftarrow (1 + d_\circ + j)$th column of $\breve{\boldsymbol{X}}_{\boldsymbol{g}}$; $\breve{\beta}^{\text{curr}} \longleftarrow (1 + d_\circ + j)$th entry of $\breve{\boldsymbol{\beta}}^{[s+N_{\text{burn}}]}$

$\widehat{\boldsymbol{f}}_{j\boldsymbol{g}}^{[s]} \longleftarrow \boldsymbol{x}_{\boldsymbol{g}}^{\text{curr}} \breve{\beta}^{\text{curr}} + \breve{\boldsymbol{Z}}_{j,\boldsymbol{g}} \breve{\boldsymbol{u}}_j^{[s+N_{\text{burn}}]}$

Outputs: $\big\{\mathring{\boldsymbol{\beta}}^{[s]}, \widehat{\boldsymbol{f}}_{j\boldsymbol{g}}^{[s]}, (\sigma_{uj}^{-2})^{[s]} : 1 \leq s \leq N_{\text{kept}},\ 1 \leq j \leq d_\bullet\big\}$

If **responseType** is Gaussian then also output $\big\{(\sigma_\varepsilon^{-2})^{[s]} : 1 \leq s \leq N_{\text{kept}}\big\}$

Algorithm 4: **continued.** *This is a continuation of the description of this algorithm that commences on a preceding page.*

$$\boldsymbol{Z}_{\text{gbl},i} \equiv \big[z_{\text{gbl},1}(\boldsymbol{x}_i) \cdots z_{\text{gbl},K_{\text{gbl}}}(\boldsymbol{x}_i)\big] \quad \text{and} \quad \boldsymbol{Z}_{\text{grp},i} \equiv \big[z_{\text{grp},1}(\boldsymbol{x}_i) \cdots z_{\text{grp},K_{\text{grp}}}(\boldsymbol{x}_i)\big]$$

where $\{z_{\text{gbl},k}(\cdot) : 1 \leq k \leq K_{\text{gbl}}\}$ and $\{z_{\text{grp},k}(\cdot) : 1 \leq k \leq K_{\text{grp}}\}$ are spline basis functions for the global f and group-specific g_i functions, respectively. The full Gaussian response Bayesian group-specific curves model that we consider here is

$$\begin{gathered}
\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{u}_{\text{gbl}}, \boldsymbol{u}_{\text{lin},i}, \boldsymbol{u}_{\text{grp},i}, \sigma_\varepsilon^2 \stackrel{\text{ind.}}{\sim} N\big(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_{\text{gbl},i}\boldsymbol{u}_{\text{gbl}} + \boldsymbol{X}_i\boldsymbol{u}_{\text{lin},i} + \boldsymbol{Z}_{\text{grp},i}\boldsymbol{u}_{\text{grp},i}, \sigma_\varepsilon^2\boldsymbol{I}_{n_i}\big),\ 1 \leq i \leq m,\\
\boldsymbol{u}_{\text{gbl}}|\sigma_{\text{gbl}}^2 \sim N(\boldsymbol{0}, \sigma_{\text{gbl}}^2\boldsymbol{I}_{K_{\text{gbl}}}), \quad \boldsymbol{u}_{\text{lin},i}|\boldsymbol{\Sigma} \stackrel{\text{ind.}}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{u}_{\text{grp},i}|\sigma_{\text{grp}}^2 \stackrel{\text{ind.}}{\sim} N(\boldsymbol{0}, \sigma_{\text{grp}}^2\boldsymbol{I}_{K_{\text{grp}}}),\ 1 \leq i \leq m,\\
\boldsymbol{\beta} \sim N(\boldsymbol{0}, \sigma_\beta^2\boldsymbol{I}_2), \quad \sigma_\varepsilon^2|b_\varepsilon \sim \text{Gamma}(\tfrac{1}{2}, b_\varepsilon), \quad b_\varepsilon \sim \text{Gamma}(\tfrac{1}{2}, s_\varepsilon^{-2}),\\
\sigma_{\text{gbl}}^{-2}|b_{\text{gbl}} \sim \text{Gamma}(\tfrac{1}{2}, b_{\text{gbl}}), \quad b_{\text{gbl}} \sim \text{Gamma}(\tfrac{1}{2}, s_{\text{gbl}}^{-2}),\\
\sigma_{\text{grp}}^{-2}|b_{\text{grp}} \sim \text{Gamma}(\tfrac{1}{2}, b_{\text{grp}}), \quad b_{\text{grp}} \sim \text{Gamma}(\tfrac{1}{2}, s_{\text{grp}}^{-2}),\\
\boldsymbol{\Sigma}^{-1}|b_{\text{lin},1}, b_{\text{lin},2} \sim \text{Wishart}\left(3, 4\begin{bmatrix} b_{\text{lin},1} & 0\\ 0 & b_{\text{lin},2}\end{bmatrix}\right), \quad b_{\text{lin},1}, b_{\text{lin},2} \stackrel{\text{ind.}}{\sim} \text{Gamma}(\tfrac{1}{2}, s_{\text{lin}}^{-2})
\end{gathered} \tag{15}$$

for hyperparameters $\sigma_\beta, s_\varepsilon, s_{\text{gbl}}, s_{\text{grp}}, s_{\text{lin}} > 0$. The notation $\boldsymbol{X} \sim \text{Wishart}(\kappa, \boldsymbol{\Lambda})$ signifies that $\boldsymbol{X}$ has a Wishart distribution with shape parameter κ and rate matrix $\boldsymbol{\Lambda}$ and is such that the corresponding density function satisfies $\mathfrak{p}(\boldsymbol{X}) \propto |\boldsymbol{X}|^{(\kappa-3)/2}\exp\{-\tfrac{1}{2}\text{tr}(\boldsymbol{\Lambda}\boldsymbol{X})\}$ for $\boldsymbol{X}$ symmetric and positive definite. The last line of (15) corresponds to $\boldsymbol{\Sigma}$ having a marginally non-informative prior distribution, for sufficiently large s_{lin}, as described in Huang & Wand (2013).

For the second case of (14), for which the y_{ij}s are binary, the Albert & Chib (1993) approach involves the introduction of the auxiliary variables, α_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n_i$, such that

$$y_{ij} \equiv I(\alpha_{ij} > 0) \quad \text{and the vectors} \quad \boldsymbol{\alpha}_i \equiv [\alpha_{ij}]_{1 \leq j \leq n_i}, \quad 1 \leq i \leq m. \tag{16}$$

With (16) in place, the full binary response Bayesian group-specific curves model is the same as (15) but with the first line replaced by

$$\boldsymbol{\alpha}_i|\boldsymbol{\beta}, \boldsymbol{u}_{\text{gbl}}, \boldsymbol{u}_{\text{lin},i}, \boldsymbol{u}_{\text{grp},i} \stackrel{\text{ind.}}{\sim} N\big(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_{\text{gbl},i}\boldsymbol{u}_{\text{gbl}} + \boldsymbol{X}_i\boldsymbol{u}_{\text{lin},i} + \boldsymbol{Z}_{\text{grp},i}\boldsymbol{u}_{\text{grp},i}, \boldsymbol{I}_{n_i}\big), \quad 1 \leq i \leq m,$$

We now present Algorithm 5 for Gibbs sampling with orthogonalized design matrices speed-ups, and let the matrices $\breve{\boldsymbol{X}}_i$ and $\breve{\boldsymbol{Z}}_{\text{gbl},i}$, $1 \leq i \leq m$, be defined according to

$$\breve{\boldsymbol{X}} = \begin{bmatrix} \breve{\boldsymbol{X}}_1\\ \vdots\\ \breve{\boldsymbol{X}}_m \end{bmatrix} \quad \text{and} \quad \breve{\boldsymbol{Z}}_{\text{gbl}} = \begin{bmatrix} \breve{\boldsymbol{Z}}_{\text{gbl},1}\\ \vdots\\ \breve{\boldsymbol{Z}}_{\text{gbl},m} \end{bmatrix}$$

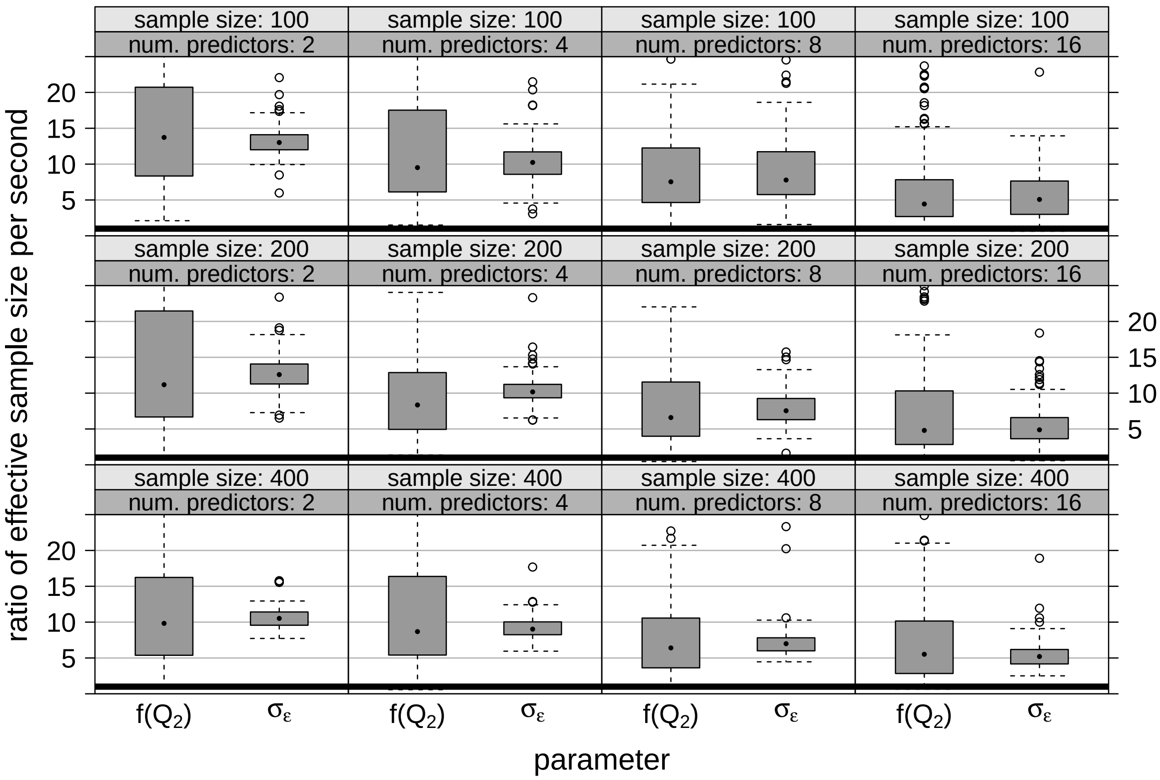

Figure 2: *Side-by-side boxplots of the effective sample size per second ratios for the computer experiment involving Bayesian generalized additive models described in the text. Each ratio corresponds to the effective sample size per second for the orthogonalized design matrices approach divided by the same quantity for the direct approach. The thick horizontal lines correspond to a ratio of* 1.

with the number of rows in $\breve{\boldsymbol{X}}_i$ and $\breve{\boldsymbol{Z}}_{\text{gbl},i}$ being the same as the number of rows in $\boldsymbol{y}_i$. The grid-wise design matrices $\boldsymbol{X}_{\boldsymbol{g}}$, $\boldsymbol{Z}_{\text{gbl},\boldsymbol{g}}$ and $\boldsymbol{Z}_{\text{grp},\boldsymbol{g}}$ are also inputted to Algorithm 5.

Figure 3 shows the results of a computer experiment in which synthetic data were generated according to the binary response version of (14) with the number of groups ranging over $m \in \{250, 500, 1000\}$ and the number of observations within the ith group fixed to be $n_i = n$ with $n \in \{30, 60\}$. The spline basis sizes were fixed at $K_{\text{gbl}} = 25$ and $K_{\text{grp}} = 15$ and the Gibbs sample sizes were $N_{\text{burn}} = N_{\text{kept}} = 5,000$. The quantities of interest are f evaluated at each of the population quantiles, $f(Q_k)$, $k = 1, 2, 3$, and the diagonal entries of $\boldsymbol{\Sigma}$, Σ_{kk}, $k = 1, 2$. The side-by-side boxplots in Figure 3 show that the orthogonalized design matrices approach improves upon the direct approach by factors of around 5–15.

4.1 Application to Adolescent Somatic Growth Data

We applied the Gaussian response version of Algorithm 5 and its direct counterpart to data on adolescent somatic growth from the study described in Pratt *et al.* (1989). The data are part of the R package HRW (Harezlak *et al.* 2021) and stored in the data frame `growthIndiana`. It consists of 9 or more longitudinal height measurements taken approximately every six months for each of 216 adolescents from Indiana, U.S.A.,

Figure 4 shows the Bayesian group-specific curve model fits. The basis sizes are $K_{\text{gbl}} = 25$ and $K_{\text{grp}} = 9$ and the hyperparameters values were set to $\sigma_\beta = s_\varepsilon = s_{\text{gbl}} = s_{\text{lin}} = s_{\text{grp}} = 10^5$ after global standardisation of the age and height data before input into Algorithm 5. The Gibbs sample size values are $N_{\text{burn}} = N_{\text{kept}} = 10,000$. In each panel of Figure 4 the curve corresponds to the posterior mean of $f(\texttt{age}) + g_i(\texttt{age})$, $1 \le i \le 216$, (in the notation of (14)) and the

Inputs: $\boldsymbol{y}_i\ (n_i \times 1)$, $\boldsymbol{X}_i\ (n_i \times 2)$, $\boldsymbol{Z}_{\text{gbl},i}\ (n_i \times K_{\text{gbl}})$, $\boldsymbol{Z}_{\text{grp},i}\ (n_i \times K_{\text{grp}})$, $1 \le i \le m$,
$\boldsymbol{X}_{\boldsymbol{g}}\ (N_{\text{grid}} \times 2)$, $\boldsymbol{Z}_{\text{gbl},\boldsymbol{g}}\ (N_{\text{grid}} \times K_{\text{gbl}})$, $\boldsymbol{Z}_{\text{grp},\boldsymbol{g}}\ (N_{\text{grid}} \times K_{\text{grp}})$,
$\sigma_\beta, s_\varepsilon, s_{\text{gbl}}, s_{\text{lin}}, s_{\text{grp}} > 0$, $N_{\text{burn}}, N_{\text{kept}} \in \mathbb{N}$, **responseType** $\in$ {Gaussian, Bernoulli}.

Decompose $\boldsymbol{X} = \boldsymbol{U}_{\boldsymbol{X}} \text{diag}(\boldsymbol{d}_{\boldsymbol{X}}) \boldsymbol{V}_{\boldsymbol{X}}^T$ where $\boldsymbol{U}_{\boldsymbol{X}}^T \boldsymbol{U}_{\boldsymbol{X}} = \boldsymbol{V}_{\boldsymbol{X}}^T \boldsymbol{V}_{\boldsymbol{X}} = \boldsymbol{V}_{\boldsymbol{X}} \boldsymbol{V}_{\boldsymbol{X}}^T = \boldsymbol{I}_2$.

Decompose $\boldsymbol{Z}_{\text{gbl}} = \boldsymbol{U}_{\boldsymbol{Z}_{\text{gbl}}} \text{diag}(\boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}}) \boldsymbol{V}_{\boldsymbol{Z}_{\text{gbl}}}^T$ where

$$\boldsymbol{U}_{\boldsymbol{Z}_{\text{gbl}}}^T \boldsymbol{U}_{\boldsymbol{Z}_{\text{gbl}}} = \boldsymbol{V}_{\boldsymbol{Z}_{\text{gbl}}}^T \boldsymbol{V}_{\boldsymbol{Z}_{\text{gbl}}} = \boldsymbol{V}_{\boldsymbol{Z}_{\text{gbl}}} \boldsymbol{V}_{\boldsymbol{Z}_{\text{gbl}}}^T = \boldsymbol{I}_{K_{\text{gbl}}}.$$

$\breve{\boldsymbol{X}} \longleftarrow \boldsymbol{U}_{\boldsymbol{X}} \text{diag}(\boldsymbol{d}_{\boldsymbol{X}})$; $\breve{\boldsymbol{Z}}_{\text{gbl}} \longleftarrow \boldsymbol{U}_{\boldsymbol{Z}_{\text{gbl}}} \text{diag}(\boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}})$; $\boldsymbol{d}_{\boldsymbol{X}}^2 \longleftarrow \boldsymbol{d}_{\boldsymbol{X}} \odot \boldsymbol{d}_{\boldsymbol{X}}$; $\boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}}^2 \longleftarrow \boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}} \odot \boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}}$

$(\text{XTy}_i)_{\text{adj}} \longleftarrow \boldsymbol{X}_i^T \boldsymbol{y}_i$; $\text{XTX}_i \longleftarrow \boldsymbol{X}_i^T \boldsymbol{X}_i$; $\text{XT}\breve{\text{X}}_i \longleftarrow \boldsymbol{X}_i^T \breve{\boldsymbol{X}}_i$; $\text{XT}\breve{\text{Z}}_{\text{gbl},i} \longleftarrow \boldsymbol{X}_i^T \breve{\boldsymbol{Z}}_{\text{gbl},i}$

$\text{SUM}\breve{\text{X}}\text{Ty}_{\text{adj}} \longleftarrow \sum_{i=1}^m \breve{\boldsymbol{X}}_i^T \boldsymbol{y}_i$; $\text{SUM}\breve{\text{Z}}_{\text{gbl}}\text{Ty}_{\text{adj}} \longleftarrow \sum_{i=1}^m \breve{\boldsymbol{Z}}_{\text{gbl},i}^T \boldsymbol{y}_i$; $\text{SUM}\breve{\text{X}}\text{T}\,\breve{\text{Z}}_{\text{gbl}} \longleftarrow \sum_{i=1}^m \breve{\boldsymbol{X}}_i^T \breve{\boldsymbol{Z}}_{\text{gbl},i}$

For $i = 1, \ldots, m$:

Decompose $\boldsymbol{Z}_{\text{grp},i} = \boldsymbol{U}_{\boldsymbol{Z}_{\text{grp},i}} \text{diag}(\boldsymbol{d}_{\boldsymbol{Z}_{\text{grp},i}}) \boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}}^T$ where $\boldsymbol{U}_{\boldsymbol{Z}_{\text{grp},i}}^T \boldsymbol{U}_{\boldsymbol{Z}_{\text{grp},i}} = \boldsymbol{I}_{K_{\text{grp}}}$
and $\boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}}^T \boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}} = \boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}} \boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}}^T = \boldsymbol{I}_{K_{\text{grp}}}$

$\breve{\boldsymbol{Z}}_{\text{grp},i} \longleftarrow \boldsymbol{U}_{\boldsymbol{Z}_{\text{grp},i}} \text{diag}(\boldsymbol{d}_{\boldsymbol{Z}_{\text{grp},i}})$; $\boldsymbol{d}_{\boldsymbol{Z}_{\text{grp},i}}^2 \longleftarrow \boldsymbol{d}_{\boldsymbol{Z}_{\text{grp},i}} \odot \boldsymbol{d}_{\boldsymbol{Z}_{\text{grp},i}}$

$(\breve{\text{Z}}_{\text{grp}}\text{Ty}_i)_{\text{adj}} \longleftarrow \breve{\boldsymbol{Z}}_{\text{grp},i}^T \boldsymbol{y}_i$; $\breve{\text{X}}\text{T}\breve{\text{Z}}_{\text{grp},i} \longleftarrow \breve{\boldsymbol{X}}_i^T \breve{\boldsymbol{Z}}_{\text{grp},i}$; $\breve{\text{Z}}_{\text{gbl}}\text{T}\breve{\text{Z}}_{\text{grp},i} \longleftarrow \breve{\boldsymbol{Z}}_{\text{gbl},i}^T \breve{\boldsymbol{Z}}_{\text{grp},i}$

Initialize: $\breve{\boldsymbol{u}}_{\text{gbl}}^{[0]}\ (K_{\text{gbl}} \times 1)$, $\boldsymbol{u}_{\text{lin},i}^{[0]}\ (2 \times m)$, $\breve{\boldsymbol{u}}_{\text{grp},i}^{[0]}\ (K_{\text{grp}} \times 1)$,
$(\sigma_\varepsilon^{-2})^{[0]}, (\sigma_{\text{gbl}}^{-2})^{[0]}, (\boldsymbol{\Sigma}^{-1})_{11}^{[0]}, (\boldsymbol{\Sigma}^{-1})_{22}^{[0]}, (\sigma_{\text{grp}}^{-2})^{[0]}, b_\varepsilon, b_{\text{gbl}}, b_{\text{lin},1}, b_{\text{lin},2}, b_{\text{grp}} > 0$,
$(\text{XTy}_i)_{\text{adj}}\ (2\times1)$, $\text{SUM}\breve{\text{X}}\text{Ty}_{\text{adj}}\ (2\times1)$, $\text{SUM}\breve{\text{Z}}_{\text{gbl}}\text{Ty}_{\text{adj}}\ (K_{\text{gbl}}\times1)$, $(\breve{\text{Z}}_{\text{grp}}\text{Ty}_i)_{\text{adj}}\ (K_{\text{grp}}\times1)$.

For $s = 1, \ldots, N_{\text{burn}} + N_{\text{kept}}$:

$$\boldsymbol{r}_{\breve{\beta}} \longleftarrow \text{SUM}\breve{\text{X}}\text{Ty}_{\text{adj}} - \text{SUM}\breve{\text{X}}\text{T}\,\breve{\text{Z}}_{\text{gbl}}\,\breve{\boldsymbol{u}}_{\text{gbl}}^{[s-1]} - \sum_{i=1}^m (\text{XT}\breve{\text{X}}_i)^T \boldsymbol{u}_{\text{lin},i}^{[s-1]} - \sum_{i=1}^m \breve{\text{X}}\text{T}\breve{\text{Z}}_{\text{grp},i} \breve{\boldsymbol{u}}_{\text{grp},i}^{[s-1]}$$

$$\boldsymbol{\psi}_{\breve{\beta}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{d}_{\boldsymbol{X}}^2 + \sigma_\beta^{-2} \boldsymbol{1}_2$$

$$\boldsymbol{z}_{\breve{\beta}} \sim N(\boldsymbol{0}, \boldsymbol{I}_2) \quad ; \quad \breve{\boldsymbol{\beta}}^{[s]} \longleftarrow \frac{\boldsymbol{z}_{\breve{\beta}}}{\sqrt{\boldsymbol{\psi}_{\breve{\beta}}}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{r}_{\breve{\beta}}}{\boldsymbol{\psi}_{\breve{\beta}}}$$

$$\boldsymbol{r}_{\breve{u}_{\text{gbl}}} \longleftarrow \text{SUM}\breve{\text{Z}}_{\text{gbl}}\text{Ty}_{\text{adj}} - (\text{SUM}\breve{\text{X}}\text{T}\,\breve{\text{Z}}_{\text{gbl}})^T \breve{\boldsymbol{\beta}}^{[s]} - \sum_{i=1}^m (\text{XT}\breve{\text{Z}}_{\text{gbl},i})^T \boldsymbol{u}_{\text{lin},i}^{[s-1]} - \sum_{i=1}^m \breve{\text{Z}}_{\text{gbl}}\text{T}\breve{\text{Z}}_{\text{grp},i} \breve{\boldsymbol{u}}_{\text{grp},i}^{[s-1]}$$

$$\boldsymbol{\psi}_{\breve{u}_{\text{gbl}}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{d}_{\boldsymbol{Z}_{\text{gbl}}}^2 + (\sigma_{\text{gbl}}^{-2})^{[s-1]} \boldsymbol{1}_{K_{\text{gbl}}}$$

$$\boldsymbol{z}_{\breve{u}_{\text{gbl}}} \sim N(\boldsymbol{0}, \boldsymbol{I}_{K_{\text{gbl}}}) \quad ; \quad \breve{\boldsymbol{u}}_{\text{gbl}}^{[s]} \longleftarrow \frac{\boldsymbol{z}_{\breve{u}_{\text{gbl}}}}{\sqrt{\boldsymbol{\psi}_{\breve{u}_{\text{gbl}}}}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{r}_{\breve{u}_{\text{gbl}}}}{\boldsymbol{\psi}_{\breve{u}_{\text{gbl}}}}$$

For $i = 1, \ldots, m$:

$\boldsymbol{r}_{u_{\text{lin},i}} \longleftarrow (\text{XTy}_i)_{\text{adj}} - \text{XT}\breve{\text{X}}_i\,\breve{\boldsymbol{\beta}}^{[s]} - \text{XT}\breve{\text{Z}}_{\text{gbl},i} \breve{\boldsymbol{u}}_{\text{gbl}}^{[s]} - \text{XT}\breve{\text{Z}}_{\text{grp},i} \breve{\boldsymbol{u}}_{\text{grp},i}^{[s]}$
$\boldsymbol{\Psi}_{u_{\text{lin},i}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]} \text{XTX}_i + (\boldsymbol{\Sigma}^{-1})^{[s-1]}$
Decompose $\boldsymbol{\Psi}_{u_{\text{lin},i}} = \boldsymbol{U}_{u_{\text{lin},i}} \text{diag}(\boldsymbol{d}_{u_{\text{lin},i}}) \boldsymbol{U}_{u_{\text{lin},i}}^T$ where $\boldsymbol{U}_{u_{\text{lin},i}}^T \boldsymbol{U}_{u_{\text{lin},i}} = \boldsymbol{U}_{u_{\text{lin},i}} \boldsymbol{U}_{u_{\text{lin},i}}^T = \boldsymbol{I}_2$

$$\boldsymbol{z}_{u_{\text{lin},i}} \sim N(\boldsymbol{0}, \boldsymbol{I}_2) \quad ; \quad \boldsymbol{u}_{\text{lin},i}^{[s]} \longleftarrow \boldsymbol{U}_{u_{\text{lin},i}} \left(\frac{\boldsymbol{U}_{u_{\text{lin},i}}^T \boldsymbol{z}_{u_{\text{lin},i}}}{\sqrt{\boldsymbol{d}_{u_{\text{lin},i}}}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{U}_{u_{\text{lin},i}}^T \boldsymbol{r}_{u_{\text{lin},i}}}{\boldsymbol{d}_{u_{\text{lin},i}}} \right)$$

continued on a subsequent page ...

Algorithm 5: *An orthogonalized design matrices Gibbs sampling scheme for the group-specific curves models (15) and (16).*

For $i = 1, \ldots, m$:

$$\boldsymbol{r}_{\breve{u}_{\text{grp},i}} \longleftarrow (\breve{\mathrm{Z}}_{\text{grp}}\mathrm{Ty}_i)_{\text{adj}} - (\breve{\mathrm{X}}\mathrm{T}\breve{\mathrm{Z}}_{\text{grp},i})^T \breve{\boldsymbol{\beta}}^{[s]} - (\breve{\mathrm{Z}}_{\text{gbl}}\mathrm{T}\breve{\mathrm{Z}}_{\text{grp},i})^T \breve{\boldsymbol{u}}_{\text{gbl}}^{[s]} - (\mathrm{XT}\breve{\mathrm{Z}}_{\text{grp},i})^T \boldsymbol{u}_{\text{lin},i}^{[s]}$$

$$\boldsymbol{\psi}_{\breve{u}_{\text{grp},i}} \longleftarrow (\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{d}^2_{\boldsymbol{Z}_{\text{grp},i}} + (\sigma_{\text{grp}}^{-2})^{[s-1]} \boldsymbol{1}_{K_{\text{grp}}}$$

$$\boldsymbol{z}_{\breve{u}_{\text{grp},i}} \sim N(\boldsymbol{0}, \boldsymbol{I}_{K_{\text{grp}}}) \quad ; \quad \breve{\boldsymbol{u}}_{\text{grp},i}^{[s]} \longleftarrow \frac{\boldsymbol{z}_{\breve{u}_{\text{grp},i}}}{\sqrt{\boldsymbol{\psi}_{\breve{u}_{\text{grp},i}}}} + \frac{(\sigma_\varepsilon^{-2})^{[s-1]} \boldsymbol{r}_{\breve{u}_{\text{grp},i}}}{\boldsymbol{\psi}_{\breve{u}_{\text{grp},i}}}$$

$$\boldsymbol{\eta}_i \longleftarrow \breve{\boldsymbol{X}}_i \breve{\boldsymbol{\beta}}^{[s]} + \breve{\boldsymbol{Z}}_{\text{gbl},i} \breve{\boldsymbol{u}}_{\text{gbl}}^{[s]} + \boldsymbol{X}_i \boldsymbol{u}_{\text{lin},i}^{[s]} + \breve{\boldsymbol{Z}}_{\text{grp},i} \breve{\boldsymbol{u}}_{\text{grp},i}^{[s]}, \; i = 1, \ldots, m$$

If **responseType** is Gaussian then

$$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}\Big(\sum_{i=1}^m n_i + 1\Big), b_\varepsilon + \tfrac{1}{2}\sum_{i=1}^m \|\boldsymbol{y}_i - \boldsymbol{\eta}_i\|^2\Big)$$

$$b_\varepsilon \longleftarrow \text{Gamma}\Big(1, (\sigma_\varepsilon^{-2})^{[s]} + s_\varepsilon^{-2}\Big)$$

If **responseType** is Bernoulli then

$(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow 1$

For $1 \le i \le m$:

For $1 \le j \le n_i$:

$\zeta \sim \text{Truncated-Normal}_+\big((2y_{ij} - 1)(\boldsymbol{\eta}_i)_j, 1\big) \;\; ; \alpha_{ij} \longleftarrow (2y_{ij} - 1)\zeta$

$$(\mathrm{XTy}_i)_{\text{adj}} \longleftarrow \boldsymbol{X}_i^T \boldsymbol{\alpha}_i \;\; ; \; \mathrm{SUM}\breve{\mathrm{X}}\mathrm{Ty}_{\text{adj}} \longleftarrow \sum_{i=1}^m \breve{\boldsymbol{X}}_i^T \boldsymbol{\alpha}_i$$

$$\mathrm{SUM}\breve{\mathrm{Z}}_{\text{gbl}}\mathrm{Ty}_{\text{adj}} \longleftarrow \sum_{i=1}^m \breve{\boldsymbol{Z}}_{\text{gbl},i}^T \boldsymbol{\alpha}_i \;\; ; (\breve{\mathrm{Z}}_{\text{grp}}\mathrm{Ty}_i)_{\text{adj}} \longleftarrow \breve{\boldsymbol{Z}}_{\text{grp},i}^T \boldsymbol{\alpha}_i$$

$$(\sigma_{\text{gbl}}^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(K_{\text{gbl}} + 1), b_{\text{gbl}} + \tfrac{1}{2}\big\|\breve{\boldsymbol{u}}_{\text{gbl}}^{[s]}\big\|^2\Big) \;\; ; \; b_{\text{gbl}} \longleftarrow \text{Gamma}\Big(1, (\sigma_{\text{gbl}}^{-2})^{[s]} + s_{\text{gbl}}^{-2}\Big)$$

$$(\boldsymbol{\Sigma}^{-1})^{[s]} \longleftarrow \text{Wishart}\Big(m + 3, 4\text{diag}(b_{\text{lin},1}, b_{\text{lin},2}) + \sum_{i=1}^m \boldsymbol{u}_{\text{lin},i}^{[s]} \boldsymbol{u}_{\text{lin},i}^{[s]T}\Big)$$

$$b_{\text{lin},1} \longleftarrow \text{Gamma}\Big(2, 2(\boldsymbol{\Sigma}^{-1})_{11}^{[s]} + s_{\text{lin}}^{-2}\Big) \;\; ; \; b_{\text{lin},2} \longleftarrow \text{Gamma}\Big(2, 2(\boldsymbol{\Sigma}^{-1})_{22}^{[s]} + s_{\text{lin}}^{-2}\Big)$$

$$(\sigma_{\text{grp}}^{-2})^{[s]} \longleftarrow \text{Gamma}\Big(\tfrac{1}{2}(mK_{\text{grp}} + 1), b_{\text{grp}} + \tfrac{1}{2}\sum_{i=1}^m \big\|\breve{\boldsymbol{u}}_{\text{grp},i}^{[s]}\big\|^2\Big)$$

$$b_{\text{grp}} \longleftarrow \text{Gamma}\Big(1, (\sigma_{\text{grp}}^{-2})^{[s]} + s_{\text{grp}}^{-2}\Big)$$

$\breve{\boldsymbol{X}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{V}_{\boldsymbol{X}} \quad ; \quad \breve{\boldsymbol{Z}}_{\text{gbl},\boldsymbol{g}} \longleftarrow \boldsymbol{Z}_{\text{gbl},\boldsymbol{g}} \boldsymbol{V}_{\boldsymbol{Z}} \quad ;$ For $i = 1, \ldots, m$: $\breve{\boldsymbol{Z}}_{\text{grp},i,\boldsymbol{g}} \longleftarrow \boldsymbol{Z}_{\text{grp},\boldsymbol{g}} \boldsymbol{V}_{\boldsymbol{Z}_{\text{grp},i}}$

For $s = 1, \ldots, N_{\text{kept}}$:

$$\widehat{\boldsymbol{\eta}}_{\boldsymbol{g}}^{[s]} \longleftarrow \breve{\boldsymbol{X}}_{\boldsymbol{g}} \breve{\boldsymbol{\beta}}^{[s+N_{\text{burn}}]} + \breve{\boldsymbol{Z}}_{\text{gbl},\boldsymbol{g}} \breve{\boldsymbol{u}}_{\text{gbl}}^{[s+N_{\text{burn}}]}$$

For $i = 1, \ldots, m$:

$$\widehat{\boldsymbol{\eta}}_{i\boldsymbol{g}}^{[s]} \longleftarrow \widehat{\boldsymbol{\eta}}_{\boldsymbol{g}}^{[s]} + \boldsymbol{X}_{\boldsymbol{g}} \boldsymbol{u}_{\text{lin},i}^{[s+N_{\text{burn}}]} + \breve{\boldsymbol{Z}}_{\text{grp},i,\boldsymbol{g}} \breve{\boldsymbol{u}}_{\text{grp},i}^{[s+N_{\text{burn}}]}$$

$$(\sigma_{\text{gbl}}^{-2})^{[s]} \longleftarrow (\sigma_{\text{gbl}}^{-2})^{[s+N_{\text{burn}}]} \;\; ; \;\; (\boldsymbol{\Sigma}^{-1})^{[s]} \longleftarrow (\boldsymbol{\Sigma}^{-1})^{[s+N_{\text{burn}}]} \;\; ; \;\; (\sigma_{\text{grp}}^{-2})^{[s]} \longleftarrow (\sigma_{\text{grp}}^{-2})^{[s+N_{\text{burn}}]}$$

If **responseType** is Gaussian then $(\sigma_\varepsilon^{-2})^{[s]} \longleftarrow (\sigma_\varepsilon^{-2})^{[s+N_{\text{burn}}]}$

Outputs: $\big\{\widehat{\boldsymbol{\eta}}_{\boldsymbol{g}}^{[s]}, \; \widehat{\boldsymbol{\eta}}_{i\boldsymbol{g}}^{[s]}, \; (\sigma_{\text{gbl}}^{-2})^{[s]}, \; (\boldsymbol{\Sigma}^{-1})^{[s]}, \; (\sigma_{\text{grp}}^{-2})^{[s]} : 1 \le s \le N_{\text{kept}}, \; 1 \le i \le m\big\}$.

If **responseType** is Gaussian then also output $\big\{(\sigma_\varepsilon^{-2})^{[s]} : 1 \le s \le N_{\text{kept}}\big\}$.

Algorithm 5: **continued.** *This is a continuation of the description of this algorithm that commences on a preceding page.*

shaded region corresponds to pointwise 95% credible intervals after back-transformation to the original units. Simple enhancements of model (15) and Algorithm 5 could be used to

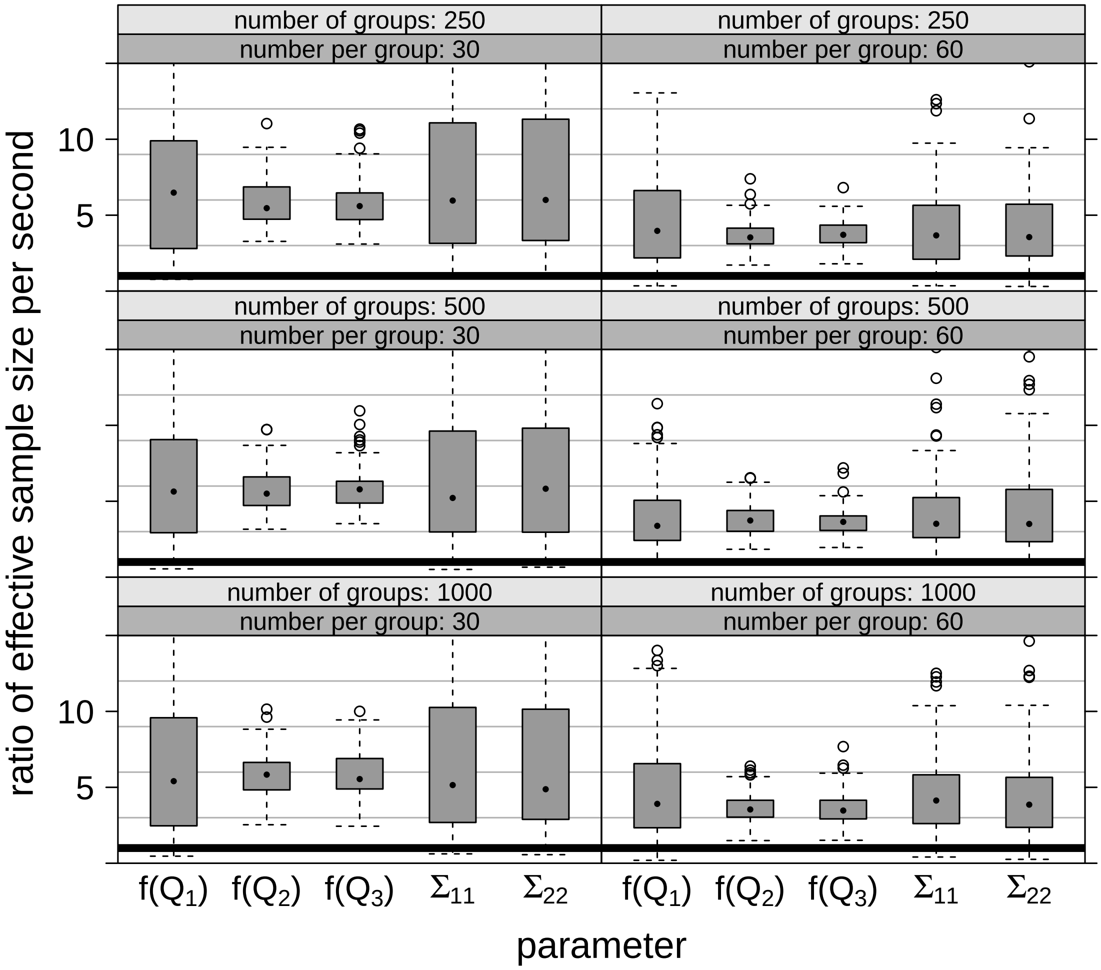

Figure 3: *Side-by-side boxplots of the effective sample size per second ratios for the computer experiment involving Bayesian group-specific curves models described in the text. Each ratio corresponds to the effective sample size per second for the orthogonalized design matrices approach divided by the same quantity for the direct approach. The thick horizontal lines correspond to a ratio of* 1.

make inferences concerning ethnicity and gender contrasts. Of interest here is the speed-ups afforded by use of orthogonalized design matrices.

When run on the first author's MacBook Air computer, with specifications given in Section 2.5, the direct approach took 51.8 seconds whereas Algorithm 5 only took 10.8 seconds. This approximately 5-fold speed-up exemplifies the advantages of orthogonalized design matrices in applications.

5 Variational Inference

Orthogonalized design matrices speed-ups also apply to other iterative Bayesian fitting and inference procedures. Here we provide a flavor of how the principle used in the previous few sections for Gibbs sampling also applies to variational inference.

Consider again the Bayesian nonparametric regression model (2). Suppose that we approximate the joint posterior density function of the model parameters:

$$\mathfrak{p}(\boldsymbol{\beta}, \boldsymbol{u}, b_u, b_\varepsilon, \sigma_u^2, \sigma_\varepsilon^2|\boldsymbol{y}) \quad \text{by the product density form} \quad \mathfrak{q}(\boldsymbol{\beta}, \boldsymbol{u}, b_u, b_\varepsilon)\mathfrak{q}(\sigma_u^2, \sigma_\varepsilon^2). \tag{17}$$

Using results laid out in, for example, Chapter 10 of Bishop (2006) the optimal $\mathfrak{q}$-densities in

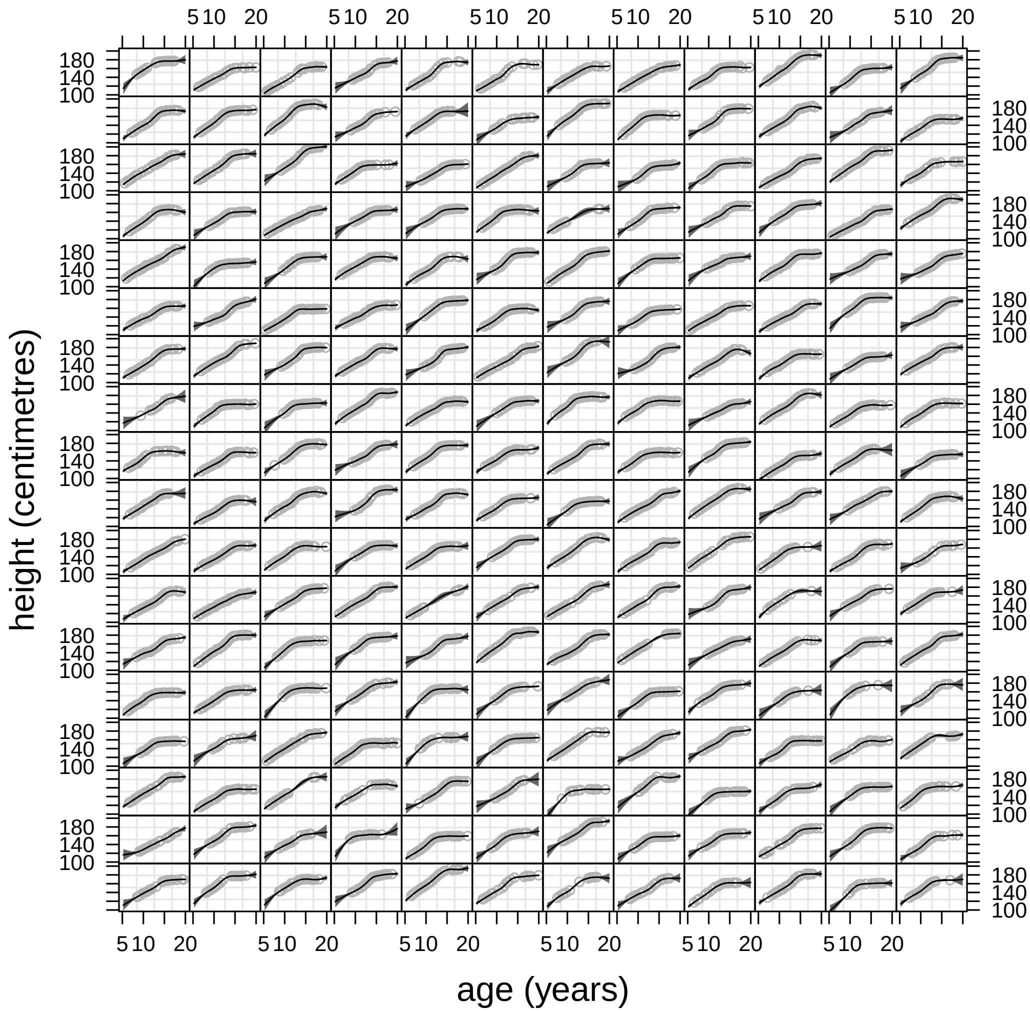

Figure 4: *Somatic growth data for each of 216 adolescents from Indiana, U.S.A., from the study described in Pratt et al. (1989). The curves are posterior means and the shading indicates pointwise 95% credible intervals for the Bayesian group-specific curves model described in Section 4 and obtained using Algorithm 5.*

terms of minimizing the Kullback-Leibler divergence of the first density function in (17) from the second are

$$\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u},b_u,b_\varepsilon)=\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u})\mathfrak{q}^*(b_u)\mathfrak{q}^*(b_\varepsilon)\quad\text{and}\qquad \mathfrak{q}^*(\sigma_u^2)\mathfrak{q}^*(\sigma_\varepsilon^2)$$

where the $\mathfrak{q}^*$-densities have forms

$$\begin{aligned}&\mathfrak{q}^*(\boldsymbol{\beta},\boldsymbol{u}):N(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})},\Sigma_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}),\ \mathfrak{q}^*(b_u):\text{Gamma}(\kappa_{\mathfrak{q}(b_u)},\lambda_{\mathfrak{q}(b_u)}),\ \mathfrak{q}^*(b_\varepsilon):\text{Gamma}(\kappa_{\mathfrak{q}(b_\varepsilon)},\lambda_{\mathfrak{q}(b_\varepsilon)}),\\ &\mathfrak{q}^*(\sigma_u^2):\ \text{Inverse-Gamma}(\kappa_{\mathfrak{q}(\sigma_u^2)},\lambda_{\mathfrak{q}(\sigma_u^2)})\quad\text{and}\quad \mathfrak{q}^*(\sigma_\varepsilon^2):\text{Inverse-Gamma}(\kappa_{\mathfrak{q}(\sigma_\varepsilon^2)},\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}).\end{aligned}$$

The optimal parameters of the $\mathfrak{q}^*$-densities can be obtained using the coordinate ascent iterative algorithm listed in Algorithm 6. The stopping criterion depends on the approximate marginal log-likelihood under product restriction (17), which we denote by $\log\{\underline{p}(\boldsymbol{y};\mathfrak{q})\}$. Its explicit form is given later in this section. The approximate posterior density function, of the grid-wise fit values vector $\widehat{\boldsymbol{f}}_{\boldsymbol{g}}\equiv\boldsymbol{X}_{\boldsymbol{g}}\boldsymbol{\beta}+\boldsymbol{Z}_{\boldsymbol{g}}\boldsymbol{u}$ is Multivariate Normal with mean vector $E_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}})=\boldsymbol{C}_{\boldsymbol{g}}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$ and covariance matrix $\text{Cov}_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}})=\boldsymbol{C}_{\boldsymbol{g}}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\boldsymbol{C}_{\boldsymbol{g}}^T$. Each of these matrices is outputted by Algorithm 6 to facilitate approximate Bayesian inference for f.

Inputs: $\boldsymbol{y}$ $(n\times 1)$, $\boldsymbol{X}$ $(n\times 2)$, $\boldsymbol{Z}$ $(n\times K)$, $\boldsymbol{X_g}$ $(N_{\text{grid}}\times 2)$, $\boldsymbol{Z_g}$ $(N_{\text{grid}}\times K)$ $\sigma_\beta, s_u, s_\varepsilon, \varepsilon_{\text{toler.}} > 0$

Initialize: $\mu_{\mathfrak{q}(\sigma_u^{-2})}, \mu_{\mathfrak{q}(b_u)}, \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}, \mu_{\mathfrak{q}(b_\varepsilon)} > 0$; $\kappa_{\mathfrak{q}(\sigma_u^2)} \longleftarrow \frac{1}{2}(K+1)$, ; $\kappa_{\mathfrak{q}(\sigma_\varepsilon^2)} \longleftarrow \frac{1}{2}(n+1)$

$\boldsymbol{C} \longleftarrow [\boldsymbol{X}\ \boldsymbol{Z}]$; $\boldsymbol{C_g} \longleftarrow [\boldsymbol{X_g}\ \boldsymbol{Z_g}]$; CTy $\longleftarrow \boldsymbol{C}^T\boldsymbol{y}$; CTC $\longleftarrow \boldsymbol{C}^T\boldsymbol{C}$

Cycle:

> $\boldsymbol{\Omega}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}\text{CTC} + \text{diag}\big(\big[\sigma_\beta^{-2}\boldsymbol{1}_2^T\ \mu_{\mathfrak{q}(\sigma_u^{-2})}\boldsymbol{1}_K^T\big]^T\big)$
>
> Decompose $\boldsymbol{\Omega}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} = \boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\text{diag}(\boldsymbol{d}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})})\boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}^T$ where
> $\boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}^T\boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} = \boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}^T = \boldsymbol{I}_{K+2}$
>
> $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow \boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\text{diag}(\boldsymbol{1}/\boldsymbol{d}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})})\boldsymbol{U}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}^T$; $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})} \longleftarrow \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\text{CTy}$
>
> $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u})} \longleftarrow$ lower $K\times 1$ block of $\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$; $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u})} \longleftarrow$ lower right $K\times K$ block of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$
>
> $\lambda_{\mathfrak{q}(\sigma_u^2)} \longleftarrow \mu_{\mathfrak{q}(b_u)} + \frac{1}{2}\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u})}\|^2 + \frac{1}{2}\text{tr}\big(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u})}\big)$; $\mu_{\mathfrak{q}(\sigma_u^{-2})} \longleftarrow \kappa_{\mathfrak{q}(\sigma_u^2)}\big/\lambda_{\mathfrak{q}(\sigma_u^2)}$
>
> $\lambda_{\mathfrak{q}(b_u)} \longleftarrow \mu_{\mathfrak{q}(\sigma_u^{-2})} + s_u^{-2}$; $\mu_{\mathfrak{q}(b_u)} \longleftarrow 1\big/\lambda_{\mathfrak{q}(b_u)}$
>
> $\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)} \longleftarrow \mu_{\mathfrak{q}(b_\varepsilon)} + \frac{1}{2}\big\|\boldsymbol{y} - \boldsymbol{C}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\big\|^2 + \frac{1}{2}\text{tr}\big(\text{CTC}\,\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\big)$; $\mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})} \longleftarrow \kappa_{\mathfrak{q}(\sigma_\varepsilon^2)}\big/\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}$
>
> $\lambda_{\mathfrak{q}(b_\varepsilon)} \longleftarrow \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})} + s_\varepsilon^{-2}$; $\mu_{\mathfrak{q}(b_\varepsilon)} \longleftarrow 1\big/\lambda_{\mathfrak{q}(b_\varepsilon)}$

until the relative change in $\log\{\underline{p}(\boldsymbol{y};\mathfrak{q})\}$ is less than $\varepsilon_{\text{toler.}}$.

$E_{\mathfrak{q}}\big(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}\big) \longleftarrow \boldsymbol{C_g}\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$; $\text{Cov}_{\mathfrak{q}}\big(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}\big) \longleftarrow \boldsymbol{C_g}\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}\boldsymbol{C_g}^T$

Outputs: $\big\{E_{\mathfrak{q}}\big(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}\big), \text{Cov}_{\mathfrak{q}}\big(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}\big), \kappa_{\mathfrak{q}(\sigma_u^2)}, \lambda_{\mathfrak{q}(\sigma_u^2)}\ \kappa_{\mathfrak{q}(\sigma_\varepsilon^2)}, \lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}\big\}$

Algorithm 6: *A direct mean field variational Bayes algorithm for fitting and inference for the Bayesian nonparametric regression model (2).*

In Algorithm 6 note that an $O(nK^2)$ singular value decomposition, for inversion of $\boldsymbol{\Omega}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$, is carried out within each iteration. Algorithm 7 avoids this computational cost by working with an orthogonalized version of $\boldsymbol{C} \equiv [\boldsymbol{X}\ \boldsymbol{Z}]$. The approximate marginal log-likelihood with respect to the $(\breve{\boldsymbol{\beta}}, \breve{\boldsymbol{u}})$ parameterization is denoted by $\log\{\underline{\breve{p}}(\boldsymbol{y};\mathfrak{q})\}$. An explicit expression is given later in this section.

Inputs: $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times 2)$, $\boldsymbol{Z}$ $(n \times K)$, $\boldsymbol{X_g}$ $(N_{\text{grid}} \times 2)$, $\boldsymbol{Z_g}$ $(N_{\text{grid}} \times K)$, $\sigma_\beta, s_u, s_\varepsilon, \varepsilon_{\text{toler.}} > 0$

Initialize: $\mu_{\mathfrak{q}(\sigma_u^{-2})}, \mu_{\mathfrak{q}(b_u)}, \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}, \mu_{\mathfrak{q}(b_\varepsilon)} > 0$; $\kappa_{\mathfrak{q}(\sigma_u^2)} \longleftarrow \frac{1}{2}(K+1)$, ; $\kappa_{\mathfrak{q}(\sigma_\varepsilon^2)} \longleftarrow \frac{1}{2}(n+1)$

$\boldsymbol{C} \longleftarrow [\boldsymbol{X}\ \boldsymbol{Z}]$; $\boldsymbol{C_g} \longleftarrow [\boldsymbol{X_g}\ \boldsymbol{Z_g}]$

Decompose $\boldsymbol{C} = \boldsymbol{U}_{\text{C}}\text{diag}(\boldsymbol{d}_{\text{C}})\boldsymbol{V}_{\text{C}}^T$ where $\boldsymbol{U}_{\text{C}}^T\boldsymbol{U}_{\text{C}} = \boldsymbol{V}_{\text{C}}^T\boldsymbol{V}_{\text{C}} = \boldsymbol{V}_{\text{C}}\boldsymbol{V}_{\text{C}}^T = \boldsymbol{I}_{K+2}$
$\breve{\boldsymbol{C}} \longleftarrow \boldsymbol{U}_{\text{C}}\text{diag}(\boldsymbol{d}_{\text{C}})$; $\breve{\text{C}}\text{Ty} \longleftarrow \breve{\boldsymbol{C}}^T\boldsymbol{y}$; $\boldsymbol{d}_{\text{C}}^2 \longleftarrow \boldsymbol{d}_{\text{C}} \odot \boldsymbol{d}_{\text{C}}$

Cycle:

$\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})} \longleftarrow \boldsymbol{1}_{K+2}\Big/\Big(\mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}\boldsymbol{d}_{\text{C}}^2 + \big[\sigma_\beta^{-2}\boldsymbol{1}_2^T\ \mu_{\mathfrak{q}(\sigma_u^{-2})}\boldsymbol{1}_K^T\big]^T\Big)$; $\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})} \longleftarrow \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})}\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}\breve{\text{C}}\text{Ty}$

$\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{u}})} \longleftarrow$ lower $K \times 1$ block of $\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}$; $\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{u}})} \longleftarrow$ lower $K \times 1$ block of $\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}$

$\lambda_{\mathfrak{q}(\sigma_u^2)} \longleftarrow \mu_{\mathfrak{q}(b_u)} + \frac{1}{2}\|\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{u}})}\|^2 + \frac{1}{2}\boldsymbol{1}_K^T\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{u}})}$; $\mu_{\mathfrak{q}(\sigma_u^{-2})} \longleftarrow \kappa_{\mathfrak{q}(\sigma_u^2)}/\lambda_{\mathfrak{q}(\sigma_u^2)}$

$\lambda_{\mathfrak{q}(b_u)} \longleftarrow \mu_{\mathfrak{q}(\sigma_u^{-2})} + s_u^{-2}$; $\mu_{\mathfrak{q}(b_u)} \longleftarrow 1/\lambda_{\mathfrak{q}(b_u)}$

$\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)} \longleftarrow \mu_{\mathfrak{q}(b_\varepsilon)} + \frac{1}{2}\Big\|\boldsymbol{y} - \breve{\boldsymbol{C}}\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}\Big\|^2 + \frac{1}{2}(\boldsymbol{d}_{\text{C}}^2)^T\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}$; $\mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})} \longleftarrow \kappa_{\mathfrak{q}(\sigma_\varepsilon^2)}/\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}$

$\lambda_{\mathfrak{q}(b_\varepsilon)} \longleftarrow \mu_{\mathfrak{q}(\sigma_\varepsilon^{-2})} + s_\varepsilon^{-2}$; $\mu_{\mathfrak{q}(b_\varepsilon)} \longleftarrow 1/\lambda_{\mathfrak{q}(b_\varepsilon)}$

until the relative change in $\log\{\underline{\breve{p}}(\boldsymbol{y};\mathfrak{q})\}$ is less than $\varepsilon_{\text{toler.}}$.

$\breve{\boldsymbol{C}}_{\boldsymbol{g}} \longleftarrow \boldsymbol{C_g}\boldsymbol{V}_{\text{C}}$; $E_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}) \longleftarrow \breve{\boldsymbol{C}}_{\boldsymbol{g}}\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}$; $\text{Cov}_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}) \longleftarrow \breve{\boldsymbol{C}}_{\boldsymbol{g}}\text{diag}(\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})})\breve{\boldsymbol{C}}_{\boldsymbol{g}}^T$

Outputs: $\{E_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}), \text{Cov}_{\mathfrak{q}}(\widehat{\boldsymbol{f}}_{\boldsymbol{g}}), \kappa_{\mathfrak{q}(\sigma_u^2)}, \lambda_{\mathfrak{q}(\sigma_u^2)}\ \kappa_{\mathfrak{q}(\sigma_\varepsilon^2)}, \lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}\}$

Algorithm 7: *An orthogonalized design matrices speed-up of Algorithm 6 for the Bayesian nonparametric regression model (2).*

For the stopping criteria in each of Algorithm 6 and Algorithm 7 first define

$$\begin{aligned}\mathcal{T}_{\mathfrak{q}(\sigma_u^2,b_u,\sigma_\varepsilon^2,b_\varepsilon)} \equiv\ &\mu_{\mathfrak{q}(b_u)}\big(\mu_{\mathfrak{q}(1/\sigma_u^2)} + s_u^{-2}\big) + \tfrac{1}{2}(K+1)\log\big(\lambda_{\mathfrak{q}(\sigma_u^2)}\big) + \log\big(\lambda_{\mathfrak{q}(b_u)}\big)\\ &+\mu_{\mathfrak{q}(b_\varepsilon)}\big(\mu_{\mathfrak{q}(1/\sigma_\varepsilon^2)} + s_\varepsilon^{-2}\big) + \tfrac{1}{2}(n+1)\log\big(\lambda_{\mathfrak{q}(\sigma_\varepsilon^2)}\big) + \log\big(\lambda_{\mathfrak{q}(b_\varepsilon)}\big)\end{aligned}$$

and let 'const' denote constant terms such as $-\frac{1}{2}\,n\log(2\pi)$. Then expressions for the $\mathfrak{q}$-density dependent components of $\log\{\underline{p}(\boldsymbol{y};\mathfrak{q})\}$ and $\log\{\underline{\breve{p}}(\boldsymbol{y};\mathfrak{q})\}$ are:

$$\begin{aligned}\log\{\underline{p}(\boldsymbol{y};\mathfrak{q})\} = \tfrac{1}{2}\log|\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}| - \tfrac{1}{2}\sigma_\beta^{-2}\big\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\beta})}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})})\big\}\\ -\tfrac{1}{2}\mu_{\mathfrak{q}(\sigma_u^{-2})}\big\{\|\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{u})}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{u})})\big\} - \mathcal{T}_{\mathfrak{q}(\sigma_u^2,b_u,\sigma_\varepsilon^2,b_\varepsilon)} + \text{const}\end{aligned}$$

and

$$\begin{aligned}\log\{\underline{\breve{p}}(\boldsymbol{y};\mathfrak{q})\} = \tfrac{1}{2}\boldsymbol{1}_{K+2}^T\log\big(\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}},\breve{\boldsymbol{u}})}\big) - \tfrac{1}{2}\sigma_\beta^{-2}\big(\big\|\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{\beta}})}\big\|^2 + \boldsymbol{1}_2^T\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{\beta}})}\big)\\ -\tfrac{1}{2}\mu_{\mathfrak{q}(\sigma_u^{-2})}\big(\|\boldsymbol{\mu}_{\mathfrak{q}(\breve{\boldsymbol{u}})}\|^2 + \boldsymbol{1}_K^T\boldsymbol{\sigma}^2_{\mathfrak{q}(\breve{\boldsymbol{u}})}\big) - \mathcal{T}_{\mathfrak{q}(\sigma_u^2,b_u,\sigma_\varepsilon^2,b_\varepsilon)} + \text{const}\end{aligned}$$

where, for example, $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta})}$ is the upper left 2×2 block of $\boldsymbol{\Sigma}_{\mathfrak{q}(\boldsymbol{\beta},\boldsymbol{u})}$.

There are numerous extensions of Gaussian response nonparametric regression for which the principles illustrated in this section apply. Some of these are the extensions described in Sections 3 and 4. Others include streamlined variational inference for higher level random effects as described in Nolan *et al.* (2020) and multiply nested group-specific curves as described in Menictas *et al.* (2021).

6 Conclusions

The orthogonalized design matrices approach to Bayesian semiparametric regression is a small-cost adjustment that yields significant speed-ups. Bayesian computing algorithms such as

Gibbs sampling and coordinate ascent variational inference require two orders of magnitude fewer operations. Our computer experiments show practical speed-ups as high as factors exceeding 60 and almost always by factors of 5 to 10. In conclusion, of orthogonalized design matrices has clear benefits in semiparametric regression applications where speed is important.

A Result 1 and Derivation

Throughout this article we use a singular value decomposition approach to obtain draws from Multivariate Normal distributions. These are underpinned by Result 1, which we now state and prove.

Result 1: *Suppose that* $\boldsymbol{\psi}$ $(d \times 1)$ *and* $\boldsymbol{\Psi}$ $(d \times d)$ *are two matrices such* $\boldsymbol{\Psi}$ *is symmetric and positive definite. Next, suppose* $\boldsymbol{\Psi} = \boldsymbol{U}\text{diag}(\boldsymbol{d})\boldsymbol{U}^T$ *is a decomposition such that* $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}_d$*. If*

$$\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I}_d) \quad \text{and} \quad \boldsymbol{x} \equiv \boldsymbol{U}\left(\frac{\boldsymbol{U}^T\boldsymbol{z}}{\sqrt{\boldsymbol{d}}} + \frac{\boldsymbol{U}^T\boldsymbol{\psi}}{\boldsymbol{d}}\right) \quad \text{then} \quad \boldsymbol{x} \sim N(\boldsymbol{\Psi}^{-1}\boldsymbol{\psi}, \boldsymbol{\Psi}^{-1}).$$

Proof of Result 1.

First note that the mean of $\boldsymbol{x}$ is

$$E(\boldsymbol{x}) = \boldsymbol{U}\left(\frac{\boldsymbol{U}^T\boldsymbol{\psi}}{\boldsymbol{d}}\right) = \boldsymbol{U}\text{diag}(\boldsymbol{1}/\boldsymbol{d})\boldsymbol{U}^T\boldsymbol{\psi} = \{\boldsymbol{U}\text{diag}(\boldsymbol{d})\boldsymbol{U}^T\}^{-1}\boldsymbol{\psi} = \boldsymbol{\Psi}^{-1}\boldsymbol{\psi}.$$

The covariance matrix of $\boldsymbol{x}$ is

$$\begin{aligned}\text{Cov}(\boldsymbol{x}) = \text{Cov}\Big(\boldsymbol{U}\left(\frac{\boldsymbol{U}^T\boldsymbol{z}}{\sqrt{\boldsymbol{d}}}\right)\Big) &= \text{Cov}\Big(\boldsymbol{U}\text{diag}(\boldsymbol{1}/\sqrt{\boldsymbol{d}})\boldsymbol{U}^T\boldsymbol{z}\Big)\\ &= \boldsymbol{U}\text{diag}(\boldsymbol{1}/\sqrt{\boldsymbol{d}})\boldsymbol{U}^T\text{Cov}(\boldsymbol{z})\boldsymbol{U}\text{diag}(\boldsymbol{1}/\sqrt{\boldsymbol{d}})\boldsymbol{U}^T\\ &= \boldsymbol{U}\text{diag}(\boldsymbol{1}/\boldsymbol{d})\boldsymbol{U}^T = \big\{\boldsymbol{U}\text{diag}(\boldsymbol{d})\boldsymbol{U}^T\big\}^{-1} = \boldsymbol{\Psi}^{-1}.\end{aligned}$$

Since $\boldsymbol{x}$ is a linear transformation of $\boldsymbol{z}$, it also has a Multivariate Normal distribution and Result 1 holds.

Acknowledgements

This research was partially supported by the Australian Research Council Discovery Project DP230101179, the Indonesian Education Scholarship, Center for Higher Education Funding and Assessment and the Indonesian Endowment Fund for Education.

References

Albert, J.H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer.

Demmler, A & Reinsch, C. (1975). Oscillation matrices with spline smoothing. *Numerische Mathematik*, **24**, 375–382.

Donnelly, C.A., Laird, N.M. and Ware, J.H. (1995). Prediction and creation of smooth curves for temporally correlated longitudinal data. *Journal of the American Statistical Association*, **90**, 984–989.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**, 515–533.

Harezlak, J., Ruppert, D. and Wand, M.P. (2018). *Semiparametric Regression with R*. New York: Springer.

Harezlak, J., Ruppert, D. and Wand, M.P. (2021). HRW 1.0. Datasets, functions and scripts for semiparametric regression supporting Harezlak, Ruppert & Wand (2018). R package. `https://CRAN.R-project.org/package=HRW`

Huang, A. and Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, **8**, 439–452.

Menictas, M., Nolan, T.H., Simpson, D.G. and Wand, M.P. (2021). Streamlined variational inference for higher level group-specific curve models. *Statistical Modelling*, **21**, 479–519.

Nolan, T.H., Menictas, M. and Wand, M.P. (2020). Streamlined computing for variational inference with higher level random effects. *Journal of Machine Learning Research*, **21(157)**, 1–62.

Polson, N.G., Scott, J.G. & Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, **108**, 1339–1349.

Pratt, J.H., Jones, J.J., Miller, J.Z., Wagner, M.A. & Fineberg, N.S. (1989). Racial differences in aldosterone excretion and plasma aldosterone concentrations in children. *New England Journal of Medicine*, **321**, 1152–1157.

Robert, C.P. (1995). Simulation of truncated normal variates. *Statistics and Computing*, **5**, 121–125.

Stan Development Team (2025). rstan: the R interface to Stan. R package version 2.32.7. `https://mc-stan.org/`

Wand, M.P. and Ormerod, J.T. (2008). On semiparametric regression with O'Sullivan penalized splines. *Australian and New Zealand Journal of Statistics*, **50**, 179–198.