Thailand Statistician
April 2026; 24(2): 371-401
http://statassoc.or.th
Contributed paper

# Bayesian Group and Sparse-Group LASSO with Spike-and-Slab Priors in Quantile Mixed Models: An Application to Child Growth Data

Taweesak Channgam*[a, b] Craig Anderson [b] and Tereza Neocleous [b]
[a] Department of Statistics and Information Management, Faculty of Science, Maejo University, Chiang Mai, Thailand
[b] School of Mathematics and Statistics, University of Glasgow, Glasgow, UK
*Corresponding author; e-mail: taweesak_c@mju.ac.th

Received: 11 August 2025
Revised: 6 January 2026
Accepted: 8 January 2026

**Abstract**

Identifying risk factors that exhibit significant associations with child growth and development is crucial for preventing unhealthy growth and supporting children's overall development. Given that children have a diverse range of growth patterns, it is particularly relevant to evaluate these associations across quantiles rather than simply focusing on the mean or median values. In this paper, we develop a Bayesian variable selection method within quantile regression (QR) and quantile mixed models (QMMs). In particular, these methods are designed to analyse longitudinal data, such as child growth data. This novel methodology combines several key components, including the Bayesian sparse group LASSO method, a likelihood function based on the scale mixture representation of the asymmetric Laplace (AL) distribution. It also incorporates spike-and-slab priors for regression coefficients and utilises linear mixed models based on a decomposition for the covariance matrix of random effects. By combining these elements, our approach offers a comprehensive solution for simultaneous selection and estimation of fixed and random effects in QMMs. We assess the performance of the proposed method through simulation studies, which demonstrate its strong variable selection and predictive capabilities. Furthermore, we illustrate its practical utility by applying it to the Growing Up in Scotland (GUS) dataset, providing practical insights into its real-world applicability.

**Keywords:** Bayesian methods, group LASSO, sparse group LASSO, MCMC algorithm, quantile regression, mixed models, longitudinal data, child growth.

## 1. Introduction

The growth and development of children is a vital area of health study in humans, involving the assessment and monitoring of physical growth, cognitive development, emotional development and social development during childhood (Berk, 2013). The aim of this study is to understand how individuals or groups progress through different stages of life and to support children in experiencing their growth in a normal process, from birth to adolescence (Berk and Meyers, 2016). Several factors, such as biological, nutritional and environmental influences, can impact children's growth, leading to either normal or abnormal development during this period (Berk, 2013; Black et al., 2008; Denham, 2006; Shonkoff et al., 2000; Starfield et al., 2005). In reality, children exhibit a diverse range of

growth patterns (underweight, normal, and overweight), and these groups may face different risk factors. Therefore, stakeholders at individual, community, and global levels need appropriate tools and knowledge to prevent risk factors affecting child growth and development. In this paper, we propose statistical models that enable us to identify such risk factors associated with child weight measurements among Scottish children.

Longitudinal studies are a key approach for understanding mechanisms of child growth and development, as repeated measurements on the same individuals allow trends and within-individual changes over time to be characterised (Bukatko and Daehler, 1995; Cole, 1994; Grammer et al., 2013; Weiss and Ware, 1996). The resulting repeated measurements form longitudinal child growth data (LCGD), where each child is followed over time (Fitzmaurice and Ravichandran, 2008). In literature, linear mixed-effects models (LMMs) (Laird and Ware, 1982) and their extensions (Cole et al., 2010; Muggeo et al., 2014; Aniley et al., 2019; Durbán et al., 2005; Ruppert et al., 2003) are a statistical method that has been widely used to analyse LCGD. These models use random effects to capture individual-specific variation and within-subject correlation alongside population-level (fixed) effects, but they are limited to modelling the mean of growth measurements and do not characterise other parts of the response distribution.

To address these limitations of mean-only models, quantile regression (QR) is a suitable approach (Koenker, 2005). However, applying QR to LCGD remains challenging due to its distribution-free framework, which complicates its incorporation with other frameworks, such as the mixed-effects framework. Fortunately, the discovery of an equivalence between maximising the likelihood function based on an asymmetric Laplace (AL) distribution and minimising the quantile loss function in the classical QR model has opened the door (Koenker and Machado, 1999). Recently, Geraci and Bottai (2014) applied this likelihood to propose linear quantile mixed models. Geraci (2019) also extended an idea of Geraci and Bottai (2014) to introduce more flexible model by allowing the use of smoothing methods with penalisation, formulated in a mixed-effects framework, named the additive quantile mixed models (AQMM). However, fixed-effect selection in these models via bootstrap-based inference remains contentious: penalised parametric bootstrap may introduce smoothing bias, nonparametric bootstrap can lead to under-smoothing due to resampling (Wood, 2006), and there is still little guidance on which random effects should be included. These limitations warrant cautious use of this approach.

From a Bayesian perspective, several approaches have been utilised for parameter estimation and variable selection in QR models, especially when analysing independent data. In the context of estimation, Yu and Moyeed (2001) made a pioneering contribution by introducing the Bayesian framework for QR. They employed the AL distribution as an augmented likelihood and improper uniform prior distributions for each regression coefficient. Additionally, Kozumi and Kobayashi (2011) introduced a location-scale mixture representation of the AL distribution, incorporating a normal prior and a double exponential prior, to develop a Gibbs sampling algorithm. Regarding variable selection, Chen, Dunson, Reed and Yu (2013) introduced a stochastic search variable selection (SSVS) (George and McCulloch, 1997) to identify important variables in QR models. Furthermore, Alhamzawi (2013) proposed a prior for variable selection based on percentage bend correlation within a Gibbs sampler. These approaches involve specifying prior distributions (e.g. Bernoulli priors) for the model indicator, which is used to represent the inclusion or exclusion of each variable in the model. This specification can be computationally demanding, particularly for large datasets or complex models.

To address this issue, it is recommended to use the Bayesian LASSO (BLASSO) method (Park and Casella, 2008). This approach assumes the regression coefficients to follow independent Laplace priors and performs variable selection using Markov chain Monte Carlo (MCMC)-based computational techniques, such as Gibbs sampling. The BLASSO method offers several notable advantages, including the ability to incorporate automated predictor selection and deselection mechanisms within the model, thereby enhancing both its efficiency and interpretability. In the QR models, Li et al. (2010) made a significant contribution by introducing BLASSO to QR. Furthermore, Alhamzawi and Yu (2014) proposed a BLASSO method for variable selection and estimation in linear quantile mixed

models (LQMMs). Ji and Shi (2022) also considered several Bayesian regularised approaches to estimate fixed and random effects in LQMMs. However, such models do not support grouped variables, such as a group of dummy variables forming a categorical variable, because they treat each variable independently and apply shrinkage individually to each predictor. Moreover, the estimates obtained from these BLASSO methods may not precisely converge to zero (Park and Casella, 2008; Xu and Ghosh, 2015; Alhamzawi and Ali, 2018), potentially impacting the accuracy of variable selection.

To accommodate grouped variables and ensure exact zero convergence, Xu and Ghosh (2015) proposed two variants of the BLASSO method with spike-and-slab priors. These two variants consist of Bayesian group LASSO (BGLASSO) and Bayesian sparse group LASSO (BSGLASSO). The former handles the selection process for entire groups of variables, while the latter goes a step further by also encouraging selection within the groups. Meanwhile, the spike-and-slab priors are employed to construct the point mass at zero for regression parameters. However, these methods were only applied to independent data and conditional mean model. In this paper, we firstly focus on an extension of BGLASSO and BSGLASSO with spike-and-slab priors in the context of the QR model. Further, we consider BSGLASSO to be incorporated with a modified Cholesky decomposition of covariance matrix of random effects (Chen and Dunson, 2003; Kinney and Dunson, 2007), enabling the simultaneous selection of both fixed effects and random effects for quantile mixed model. In summary, our contributions are: (1) extending Bayesian group and sparse-group LASSO with spike-and-slab priors to quantile regression, and (2) developing a novel quantile mixed-effects model that performs simultaneous selection of fixed and random effects. Furthermore, this method addresses key questions in the analysis of LCGD in Scotland (i.e. the Growing Up in Scotland (GUS) data), particularly in identifying risk factors influencing physical growth. This is exemplified by the rising rates of childhood obesity in Scotland (Royal College of Physicians and Surgeons of Glasgow, 2023), which we investigate latter in the paper.

In brief, our proposed approach enables automatic selection of both significant fixed effects and necessary random effects in quantile mixed models, which we demonstrate through simulations and an application. The rest of this paper is organised as follows: First, we provide an overview of the QR model and the quantile mixed model in Section 2. Section 3 presents the proposed methodologies, including Bayesian group and sparse group LASSO methods, Bayesian group and sparse group LASSO with spike-and-slab priors in the QR context, the methods for estimating the tuning parameters in our proposed models, Bayesian sparse group LASSO mixed QR model, and the variable selection process for the proposed models. We delve into the simulation studies in Section 4. Moving on to Section 5, we present our results, encompassing the simulation study findings, the illustrative analysis for simultaneous selection. The application to real data is presented in Section 6. Furthermore, Section 7 presents our discussion. Finally, we conclude the paper in Section 8.

## 2. Background

### 2.1. Quantile regression model

Consider the following conditional linear quantile regression (QR) model for independent data (Koenker, 2005) in the form $(y_i, \mathbf{x}_i)$ for $i = 1, \ldots, n$:

$$Q_\tau(y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}_\tau, \tag{1}$$

where $\mathbf{x}_i$ is a vector of known predictors and $\boldsymbol{\beta}_\tau \in \mathbb{R}^{(p+1)}$ is a vector of unknown regression coefficients dependent on $\tau \in (0, 1)$, which can be interpreted as the marginal change in the $\tau$th quantile resulting from a marginal change in $\mathbf{x}$. To estimate the coefficient vector $\boldsymbol{\beta}$ of model (1) at a given quantile $\tau$ from a sample of $n$ pairs $\mathbf{x}_i$, solve

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\boldsymbol{\beta}), \tag{2}$$

where $\rho_\tau(u) = u(\tau - \mathbb{I}\{u < 0\})$ is the asymmetric absolute (check) loss and $\mathbb{I}(\cdot)$ denotes the indicator function. The estimator can be computed efficiently using linear-programming algorithms (Koenker, 2005).

Alternatively, the model in (1) can be conveyed through the linear model as

$$y_i = \mathbf{x}_i'\boldsymbol{\beta}_\tau + \epsilon_i, \tag{3}$$

where the random error term is defined as $\epsilon_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau$ and is specified to satisfy $Q_\tau(\epsilon_i|\mathbf{x}_i) = 0$. We assume that the errors in model (3) $\epsilon_i$ are independent and identically distributed according to the asymmetric Laplace distribution (AL), $\epsilon_i \sim \text{AL}(\mu, \sigma, \tau)$, where $\mu = 0$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\tau \in (0, 1)$ is the skew (quantile) parameter. Under this working likelihood, with conditional location $\mu = \mathbf{x}_i'\boldsymbol{\beta}$, the maximum likelihood estimator of $\mu$ is obtained by the minimisation in (2), in line with the standard quantile regression formulation (Yu and Moyeed, 2001). Therefore, QR can be represented by assuming that each $y_i$, conditional on $\mathbf{x}_i$, follows the AL distribution. Consequently, the likelihood function based on this distribution can be expressed as

$$\mathcal{L}(\boldsymbol{\epsilon}|\sigma) \propto \sigma^{-n} \exp\left\{ -\sum_{i=1}^{n} \frac{|\epsilon_i| + (2\tau - 1)\epsilon_i}{2\sigma} \right\}. \tag{4}$$

As demonstrated by Tsionas (2003), Kozumi and Kobayashi (2011) and Alhamzawi and Yu (2013), the AL distribution can be formulated as a mixture of normals representation. This discovery can lead to the use of the Gibbs sampling algorithm for effective estimation in the Bayesian inference for QR model. To establish this representation, we employ the method of Andrews and Mallows (1974), in which for any $a, b > 0$,

$$\exp\{-|ab|\} = \int_0^\infty \frac{a}{\sqrt{2\pi e}} \exp\left\{ -\frac{1}{2}(a^2 e + b^2 e^{-1}) \right\}. \tag{5}$$

We apply Equation (5) to the right-hand side of the function (4) by assuming $a = 1/\sqrt{2\sigma}$, $b = \epsilon/\sqrt{2\sigma}$ and multiplying by $\exp\{-(2\tau - 1)\epsilon/2\sigma\}$, it results in

$$\sigma^{-n} \exp\left\{ -\sum_{i=1}^{n} \frac{|\epsilon_i| + (2\tau - 1)\epsilon_i}{2\sigma} \right\} = \prod_{i=1}^{n} \int_0^\infty \frac{1}{\sigma\sqrt{4\pi\sigma v_i}} \exp - \left\{ \frac{(\epsilon_i - \xi v_i)^2}{4\sigma v_i} - \zeta v_i \right\} dv_i,$$

where $\xi = (1 - 2\tau)$, $\zeta = \tau(1 - \tau)/\sigma$ and $v_i \sim \text{Exp}(\zeta)$. Further, let $\epsilon_i = y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau$, this implies that

$$\mathcal{L}(\mathbf{y}|\boldsymbol{v}, \boldsymbol{\beta}, \sigma) \propto \prod_{i=1}^{n} \int_0^\infty \frac{1}{\sigma\sqrt{4\pi\sigma v_i}} \exp - \left\{ \frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta}_\tau - \xi v_i)^2}{4\sigma v_i} - \zeta v_i \right\} dv_i.$$

### 2.2. Quantile mixed-effects models

In the context of the general linear mixed-effects models (LMMs) for longitudinal data, represented as $(y_{ij}, \mathbf{x}_{ij}', \mathbf{z}_{ij}')$, for $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$:

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\boldsymbol{u}_i + \epsilon_{ij}, \tag{6}$$

where $y_{ij}$ is the $j$th response for subject $i$, $\mathbf{x}_{ij}'$ and $\mathbf{z}_{ij}'$ are the rows of the $\mathbf{X}_i$ and $\mathbf{Z}_i$, $\mathbf{X}_i$ is a known $n_i \times p$ matrix and $\mathbf{Z}_i$ is a known $n_i \times q$ matrix. $\boldsymbol{u}_i$ is a $q \times 1$ vector of unobservable random effects with $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$ and $\epsilon_{ij}$ is the error term with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

In the realm of Bayesian variable selection in LMMs, Kinney and Dunson (2007) introduced a Gibbs-sampling-based method for simultaneous selection of fixed and random effects. This method involves reparameterising the random-effects part of the model (6) as

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{z}_{ij}'\mathbf{D}\mathbf{A}\boldsymbol{c}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2),$$

where $\boldsymbol{c}_i = (c_{i1}, \ldots, c_{iq})', \mathbf{D} = \text{diag}(d_1, \ldots, d_q)'$ and $\mathbf{A} = (a_{lr})$ is a $q \times q$ lower triangular matrix with ones on diagonal and free off-diagonal elements representing the correlations among the random effects, where $l = 1, \ldots, q$ and $r = 1, \ldots, q(q-1)/2$. In this parameterisation, the random-effects vector is written as $\boldsymbol{u}_i = \mathbf{DA}\boldsymbol{c}_i$ with $\boldsymbol{c}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = \text{diag}(\omega_1, \ldots, \omega_q)'$, so the random-effects covariance maxtrix is $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathbf{DA}\boldsymbol{\Omega}\mathbf{A}'\mathbf{D}$. Each $d_l$ is proportional to the standard deviation of the $l$-th random effect. Within a Bayesian variable-selection scheme, assigning suitable shrinkage priors on the $d_l$ allows them to be shrunk towards (or exactly to) zero, effectively removing the corresponding random effects from the model while preserving the positive-definiteness of $\boldsymbol{\Sigma}_{\boldsymbol{u}}$. Recently, Alhamzawi and Yu (2014) applied this idea in conjunction with BLASSO to LQMMs. However, they made the assumption that $\boldsymbol{c}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Ji and Shi (2022) also adopted the same assumption while exploring various regularisation methods. Motivated by this reparameterisation, we propose the linear quantile mixed models,

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta}_\tau + \mathbf{z}_{ij}'\mathbf{DA}\boldsymbol{c}_{\tau,i} + \epsilon_{ij},$$

where $\boldsymbol{\beta}_\tau$ is a vector of coefficients that depend on $\tau$ and $\epsilon_{ij}$ are independent, with their $\tau$th quantile level assumed to be zero. These are also presumed to follow the AL distribution. Let $\epsilon_{ij} = y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta} - \mathbf{z}_{ij}'\mathbf{DA}\boldsymbol{c}_i$ and follow the scale mixture of normals, as outlined in Section 2.1, the likelihood of $\epsilon_{ij}$ can be expressed as a scale mixture of normals with exponential mixing density,

$$\prod_{i=1}^{N}\prod_{j=1}^{n_i}\int_0^\infty \frac{1}{\sigma\sqrt{4\pi\sigma v_{ij}}}\exp\left\{-\frac{(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{z}_{ij}'\mathbf{DA}\boldsymbol{c}_i - \xi v_{ij})^2}{4\sigma v_{ij}} - \zeta v_{ij}\right\}dv_{ij},$$

where $\xi = (1-2\tau)$, $\zeta = \tau(1-\tau)/\sigma$, and $v_{ij} \sim \text{Exp}(\zeta)$.

## 3. Methodology

### 3.1. Bayesian group and sparse group LASSO for quantile regression

Consider the QR model (3) consisting of the covariates $X_k$, which are partitioned into $G$ non-overlapping groups,

$$\mathbf{y} = \sum_{g=1}^{G}\mathbf{X}_g\boldsymbol{\beta}_{\tau,g} + \boldsymbol{\epsilon}. \tag{7}$$

Here, the matrix $\mathbf{X}_g$ has $m_g$ columns and $n$ rows, and it is a sub-matrix of the design matrix $\mathbf{X}$. $\boldsymbol{\beta}_{\tau,g} = (\beta_{\tau,g1}, \ldots, \beta_{\tau,gm_g})'$ is the coefficient vector corresponding to the group $g$. Note that for simplicity of notation, we omit the subscript $\tau$ for $\boldsymbol{\beta}_{\tau,g}$. We apply the Group LASSO (GL) approach to the quantile model as proposed by Li, Lin and Xi Li et al. (2010). Thus, at the $\tau$th quantile, the estimator can be obtained by

$$\min_{\boldsymbol{\beta}} \rho_\tau\left(\mathbf{y} - \sum_{g=1}^{G}\mathbf{X}_g\boldsymbol{\beta}_g\right) + \lambda\sum_{g=1}^{G}||\boldsymbol{\beta}_g||_2. \tag{8}$$

Here, $\lambda$ is the regularisation parameters, controlling coefficient shrinkage and thus the bias-variance trade-off in the model. However, the penalised objective function (8) is limited in its scope, as it solely caters to group-structured variables without considering individual levels within those group variables, thereby lacking the ability to address broader real-world issues. To overcome this limitation, the sparse Group LASSO (SGL) (Simon et al., 2013) was introduced, which uses combined $L_1$ and $L_2$ penalties in the regression loss to enable variable selection both between and within groups. Consequently, (8) can be expanded to take the following form

$$\min_{\boldsymbol{\beta}} \rho_\tau\left(\mathbf{y} - \sum_{g=1}^{G}\mathbf{X}_g\boldsymbol{\beta}_g\right) + \lambda_1||\boldsymbol{\beta}||_1 + \lambda_2\sum_{g=1}^{G}||\boldsymbol{\beta}_g||_2,$$

where $\lambda_1$ and $\lambda_2$ control the strength of the $L_1$ (within-group) and $L_2$ (group) penalties, thereby balancing selection within and between groups.

As noted by Tibshirani (1996), the LASSO estimator is equivalent to the posterior mode obtained under independent Laplace (double-exponential) priors on the regression coefficients. This link between penalised likelihood and Bayesian modelling extends to LASSO-type methods. Building on this idea, Kyung et al. (2010) formulated the GL approach in the Bayesian framework by assigning a multivariate Laplace prior to each group of coefficients,

$$p(\boldsymbol{\beta}_g) \propto \exp\left\{-\frac{\lambda}{\sigma}||\boldsymbol{\beta}_g||_2\right\}. \tag{9}$$

Similarly, Xu and Ghosh (2015) proposed the SGL prior,

$$p(\boldsymbol{\beta}) \propto \exp\left\{-\frac{\lambda_1}{2\sigma^2}||\boldsymbol{\beta}||_1 - \frac{\lambda_2}{2\sigma^2}||\boldsymbol{\beta}_g||_2\right\}. \tag{10}$$

In practice, the priors (9) and (10) can be represented as a scale mixture of normals with conjugate Gamma hyperpriors, allowing for the construction of a fully Bayesian hierarchical model and an efficient Gibbs sampler to address both GL and SGL problems (Xu and Ghosh, 2015). Nevertheless, Xu and Ghosh (2015) highlighted that estimates derived from these priors, whether they are posterior means or medians, do not precisely reach a value of zero. To address this issue, they proposed a hierarchical Bayesian sparse group LASSO model, incorporating an independent spike-and-slab type prior for both group variable selection and individual variable selection.

### 3.2. Spike-and-slab priors for regression coefficients

Spike-and-slab priors have been emerging as a prominent and widely favourited approach for Bayesian variable selection, ever since their introduction by Mitchell and Beauchamp (1988). These priors are typically formulated as mixture priors, often utilising normal mixture priors with two components (Ishwaran and Rao, 2005). One component assigns a probability mass to the regression coefficient $\beta_k$ being precisely equal to zero (spike component), while the other component allows for non-zero (slab component). This unique feature enables spike and slab priors to effectively identify and select relevant variables, making them suitable for variable selection tasks. As a result, these priors offer a flexible and powerful tool for variable selection within a Bayesian framework. In this paper, we adopt a class of continuous bimodal priors proposed by Ishwaran and Rao (2005). These priors offer a more computationally efficient option while maintaining the desired characteristics of the model. Consider the linear regression model, expressed as $y_i \sim \mathcal{N}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2)$. The specification of this class of priors is

$$\begin{aligned}
(\beta_k|\gamma_k, \varsigma_k^2) &\overset{\text{ind}}{\sim} \mathcal{N}(0, \gamma_k\varsigma_k^2), \quad k = 1, \ldots, p, \\
(\gamma_k|\nu_0, \pi) &\overset{\text{i.i.d.}}{\sim} (1-\pi)\delta_{\nu_0}(\cdot) + \pi\delta_1(\cdot), \\
(\nu_k^{-2}|a, b) &\overset{\text{i.i.d.}}{\sim} \text{Gamma}(a, b), \\
\pi &\sim \text{Beta}(b_1, b_2),
\end{aligned} \tag{11}$$

where $\delta_\nu(\cdot)$ is a discrete measure concentrated at the value $\nu$. Based on the prior hierarchy (11), it can be seen that $\gamma^* = \gamma_k\varsigma_k^2$ (hypervariance) approximates a continuous bimodal distribution, featuring a spike at $\nu_0$ and a continuous right tail, provided that a small near zero value ($\nu_0$) and the shape ($a$) and scale ($b$) parameters of the Gamma distribution are properly chosen. Illustrative prior densities for hypervariance ($\gamma^*$), under several choices of $(\nu_0, a, b)$, are shown in Appendix I (Figure 4). Indeed, integrating out $\gamma_k$, the conditional prior of $\beta_k$ in (11) can be written more compactly as

$$\beta_k \mid \varsigma_k^2, \pi \overset{\text{ind}}{\sim} (1-\pi)\,\mathcal{N}\big(0, \nu_0\varsigma_k^2\big) + \pi\,\mathcal{N}\big(0, \varsigma_k^2\big).$$

For sufficiently small $\nu_0$, the first component $\mathcal{N}(0, \nu_0\varsigma_k^2)$ behaves like a near–point mass at zero, inducing strong shrinkage of coefficients $\beta_{k,0}$ whose true effects are zero.

### 3.3. Bayesian group LASSO QR with spike-and-slab priors

Motivated by a work of Xu and Ghosh (2015), by the equivalence between the maximum likelihood estimator under the AL distribution and the standard quantile loss function, by the representation of the AL distribution as a scale mixture of normals with an exponential mixing density discussed in Section 2, and by the advantages of placing a point mass at zero on regression coefficients through spike-and-slab priors as detailed in Section 3.2, we formulate the following Bayesian group LASSO QR with spike-and-slab priors (hereafter referred to as BGLSSQR).

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ represent a response vector, $\boldsymbol{\beta}_g$ is a coefficient vector of length $m_g$, $\mathbf{X}_g$ is an $m_g \times n$ covariate matrix corresponding to the factor $\boldsymbol{\beta}_g (g = 1, 2, \ldots, G)$, and $\mathbf{V} = \text{diag}(v_1^{-1}, \ldots, v_n^{-1})$. Hence, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_G)'$ is a coefficient vector of length $p$ ($p = \sum_{g=1}^{G} m_g$) and $\mathbf{X}$ is an $n \times p$ design (predictors) matrix corresponding to $\boldsymbol{\beta}$. Here, we define $\boldsymbol{\Psi}_g$ as a known $m_g \times m_g$ positive definite matrix, and $\pi_0$ as the probability of $\boldsymbol{\beta}_g = 0$. Therefore, the Bayesian hierarchical QR model can be expressed as

$$
\begin{aligned}
\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{v}, \sigma &\sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta} + (1-2\tau)\mathbf{V}^{-1}\mathbf{1}_n, 2\sigma\mathbf{V}), \\
\boldsymbol{v}|\sigma &\sim \text{Exp}(\sigma^{-1}\tau(1-\tau)), \\
\boldsymbol{\beta}_g|\eta_g^2 &\overset{\text{ind}}{\sim} (1-\pi_0)\mathcal{N}_{m_g}(\mathbf{0}, \eta_g^2\boldsymbol{\Psi}_g^{-1}) + \pi_0\delta_0(\boldsymbol{\beta}_g), \\
\eta_g^2|\lambda^2 &\overset{\text{ind}}{\sim} \text{Gamma}\Big(\frac{m_g+1}{2}, \frac{\lambda^2}{2}\Big), \\
\sigma &\sim \text{InvGamma}(g_1 = 0.1, g_2 = 0.1), \\
\pi_0 &\sim \text{Beta}(a_1 = 1.0, a_2 = 1.0).
\end{aligned}
\tag{12}
$$

We propose an efficient Gibbs sampler to draw samples from the joint posterior distribution of the parameters of interest in the proposed Bayesian model (12). A detailed description of this sampler is provided in Appendix II.

### 3.4. Bayesian sparse group LASSO QR with spike-and-slab priors

To handle sparsity at the group level and within the group level, we adopt the reparameterisation of regression coefficients (Xu and Ghosh, 2015) to QR:

$$\boldsymbol{\beta}_{\tau,g} = \mathbf{K}_{\tau,g}^{1/2}\boldsymbol{b}_{\tau,g},$$

where $\mathbf{K}_{\tau,g}^{1/2} = \text{diag}\{\theta_{g1}, \ldots, \theta_{gm_g}\}, \theta_{gj} \geq 0, g = 1, \ldots, G; j = 1, \ldots, m_g$. Here, $\boldsymbol{b}_{\tau,g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m_g})$ in case where these are not zeros. This reparameterisation, employed in this context, plays a crucial role in utilising the diagonal elements of $\mathbf{K}_{\tau,g}^{1/2}$ to regulate the magnitude of the elements of $\boldsymbol{\beta}_{\tau,g}$. To maintain the simplicity of notation, the subscript $\tau$ for all pertinent parameters will be omitted.

A methodology for the selection of group variables is based on the assumption that each $\boldsymbol{b}_g$ follows the multivariate spike-and-slab prior:

$$\boldsymbol{b}_g \overset{\text{ind}}{\sim} (1-\pi_0)\mathcal{N}_{m_g}(\mathbf{0}, \mathbf{I}_{m_g}) + \pi_0\delta_0(\boldsymbol{b}_g),$$

where $\pi_0$ is defined as the probability of $\boldsymbol{b}_{\tau,g} = 0$. Consequently, if $\theta_{gj}$ is zero, regardless of whether $b_{gj}$ is zero or not, $\beta_{gj}$ can be dropped from the model. In a similar way to the selection of group variables, the spike-and-slab prior is utilised for the within-group level, assuming that each $\theta_{gj}$ also adheres to this prior distribution:

$$\theta_{gj} \overset{\text{ind}}{\sim} (1-\pi_1)\mathcal{N}^{+}(0, s^2) + \pi_1\delta_0(\theta_{gj}).$$

Here, $\pi_1$ denotes the probability that $\theta_{gj} = 0$, and $\mathcal{N}^{+}(0, s^2)$ represents a truncated normal distribution with mean 0 and variance $s^2$, truncated from below at 0.

We combine the Bayesian sparse group LASSO method with the mixture representation of the AL distribution and the aforementioned spike-and-slab priors to develop the Bayesian sparse group LASSO for QR (BSGSSQR). This innovative approach is structured as a Bayesian hierarchical model, integrating the mentioned components to achieve efficient variable selection within the context of QR. Let $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_G)$, and $\mathbf{K}^{1/2} = \text{diag}(\mathbf{K}_1^{1/2}, \ldots, \mathbf{K}_G^{1/2})$, then the Bayesian hierarchical QR model is expressed as follows:

$$\begin{aligned}
\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{v}, \sigma &\sim \mathcal{N}(\mathbf{X}\mathbf{K}^{1/2}\boldsymbol{b} + (1-2\tau)\mathbf{V}^{-1}\mathbf{1}_n, 2\sigma\mathbf{V}),\\
\boldsymbol{v}|\sigma &\sim \text{Exp}(\sigma^{-1}\tau(1-\tau)),\\
\boldsymbol{b}_g &\overset{\text{ind}}{\sim} (1-\pi_0)\mathcal{N}_{m_g}(\mathbf{0}, \mathbf{I}_{m_g}) + \pi_0\delta_0(\boldsymbol{b}_g),\\
\theta_{gj} &\overset{\text{ind}}{\sim} (1-\pi_1)\mathcal{N}^+(0, s^2) + \pi_1\delta_0(\theta_{gj}),\\
s^2|t &\overset{\text{ind}}{\sim} \text{InvGamma}(1, t),\\
\sigma &\sim \text{InvGamma}(g_1 = 0.1, g_2 = 0.1),\\
\pi_0 &\sim \text{Beta}(a_1 = 1.0, a_2 = 1.0),\\
\pi_1 &\sim \text{Beta}(c_1 = 1.0, c_2 = 1.0).
\end{aligned} \tag{13}$$

The full specification of the efficient Gibbs sampler used to draw samples from the joint posterior distribution of the parameters of interest in the proposed Bayesian model (13) is provided in Appendix II.

### 3.5. Bayesian sparse group LASSO mixed QR

Based on the BSGSSQR method detailed in Section 3.4 and the reparameterisation of random effects mentioned in Section 2.2, we propose the following $l_1$-penalised check function

$$\min_{\boldsymbol{\beta}_\tau, \boldsymbol{c}_\tau} \sum_{i=1}^{N}\sum_{j=1}^{n_i} \rho_\tau\big(\epsilon_{ij}\big) + \lambda_1||\boldsymbol{\beta}_\tau||_1 + \lambda_2\sum_{g=1}^{G}||\boldsymbol{\beta}_{g,\tau}||_2 + \lambda_3\sum_{i=1}^{N}\sum_{l=1}^{q}|c_{il\tau}|, \tag{14}$$

where $\boldsymbol{c}_\tau = (\boldsymbol{c}'_{1\tau}, \ldots, \boldsymbol{c}'_{N\tau})$ and $\lambda_3 \geq 0$. In this case, we follow Alhamzawi (2013) to restrict $\lambda_3$ to being 1. Here, it can be observed that the function (14) yields the sparse group LASSO estimates on $\boldsymbol{\beta}_{g,\tau}$ and the LASSO estimates on $\boldsymbol{c}_\tau$. These estimates can be interpreted as posterior mode estimates when $\boldsymbol{\beta}_{g,\tau}$ and $\boldsymbol{c}_\tau$ have independent and identical Laplace priors. Therefore, the Laplace prior $(1/2)\exp\{-|\boldsymbol{c}_{il\tau}|\}$ on $\boldsymbol{c}_{il\tau}, l = 1, \ldots, q$ (random effects) (Alhamzawi, 2013), can be employed, while the prior on $\boldsymbol{\beta}_{g,\tau}$ (fixed effects) can be specified in Section 3.4. In practice, a scale mixture of normals of the Laplace prior is exploited in the Gibbs sampler for the Bayesian analysis (Alhamzawi, 2013). This mixture can be represented as

$$\frac{1}{2}\exp\{-|\boldsymbol{c}_{il\tau}|\} = \int_0^\infty \mathcal{N}(\boldsymbol{c}_{il\tau}; 0, \boldsymbol{\omega}_l)\text{Exp}\Big(\boldsymbol{\omega}_l; \frac{1}{2}\Big)d\boldsymbol{\omega}_l.$$

Henceforth, this approach is referred to as Bayesian sparse group LASSO-mixed quantile model with Spike-and-Slab priors (BSGSSMQR). In this context, the prior for $\boldsymbol{\omega}_l$ can be assumed to follow a Gamma(1,2) distribution, which is equivalent to a $\text{Exp}\Big(\boldsymbol{\omega}_l; 1/2\Big)$ distribution. This prior is informative and convenient for prior elicitation (Alhamzawi, 2013). For $d_l$, it is specified to follow a zero-inflated half-normal prior, $\text{ZI} - \mathcal{N}^+(\mu_{d_l} = 0, \sigma^2_{d_l} = 1, p_{l0})$, where $p_{l0} = \Pr(d_l = 0)$ (the prior probability that random effects $l$ is excluded). For $\boldsymbol{a} = (a_{lr} : l = 2, \ldots, q; r = 1, \ldots, l-1)$, it is assumed to follow the prior $p(\boldsymbol{a}|\boldsymbol{d}) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{a}}) \cdot 1(\boldsymbol{a} \in \mathcal{R}_{\boldsymbol{d}})$, where $\boldsymbol{\Sigma}_{\boldsymbol{a}}$ is the variance-covariance matrix of $\boldsymbol{a}$, $1(\cdot)$ is an indicator function and $\mathcal{R}_{\boldsymbol{d}}$ enforced each element of $\boldsymbol{a}$ to be zero corresponding to random effects that are zero (Kinney and Dunson, 2007). The Gibbs sampler for this approach is provided in Appendix II. The code to implement the BSGSSMQR method was written in the statistical package R (R Core Team, 2024) and is available on request from the first author.

### 3.6. Specification for the tuning parameters

For the BGLSSQR approach, the hyperparameter $\lambda$ holds significant importance within any Bayesian LASSO model type, as it plays a pivotal role in the coefficient shrinkage process, thus exerting a substantial influence on its overall effectiveness and predictive capabilities (Park and Casella, 2008; Xu and Ghosh, 2015). One commonly employed approach is to assign a conjugate prior, such as the Gamma$(c_1, c_2)$ prior, to this hyperparameter, which allows for straightforward computation and incorporation of prior beliefs or knowledge regarding coefficient sparsity. However, an alternative method for estimating $\lambda$ exists. We use a Monte Carlo EM (Park and Casella, 2008; Xu and Ghosh, 2015) to estimate $\lambda$ by maximising the marginal likelihood; the $E$-step computes $E[\eta_g^2|\mathbf{y}]$ from the Gibbs samples and the $M$-step updates $\lambda$ via (13). The $k$th update of the EM algorithm for $\lambda$ is represented by

$$\lambda^{(k)} = \sqrt{\frac{p+G}{\sum_{g=1}^{G} E_{\lambda^{k-1}}[\eta_g^2|\mathbf{y}]}},$$

where $p$ denotes the number of covariates and $G$ represents the number of group variables. During this EM update, the value of $\eta_g^2$ is substituted by the sample average of $\eta_g^2$, which is obtained through the Gibbs sampler using the $\lambda$ from the previous iteration $(k-1)$. Additionally, hyperparameters such as the shape and scale parameters of an inverse Gamma prior for the scale parameter, $\sigma$, are chosen to be relatively small: $g_1 = 0.1$ and $g_2 = 0.1$. This decision is based on allowing the data to speak for themselves in determining the posterior distribution and ensuring computational stability (Gelman, 2006, 2014). Also, two hyperparameters of a Beta prior for $\pi_0$ are set to $a_1 = a_2 = 1$ as default values. This is equivalent to assuming a standard uniform distribution, resulting in a prior mean of $1/2$ and allowing a prior spread (Xu and Ghosh, 2015).

Analogous to the estimation of $\lambda$, the hyperparameter $t$ for BSGSSQR and BSGSSMQR, assumes a crucial role in the process of coefficient shrinkage. To estimate this parameter, the Monte Carlo EM algorithm is adopted, as described by Casella (2001) and Park and Casella (2008). For the $k$th EM update,

$$t^{(k)} = \frac{1}{E_{t^{(k-1)}}[\frac{1}{s^2}|\mathbf{y}]},$$

where the posterior expectation of $s^2$ is substituted by the sample average of $s^2$. This average is obtained through the Gibbs sampler using the $t$ from the previous iteration $(k-1)$.

### 3.7. Variable selection process

The Bayesian hierarchical QR models discussed earlier typically yield posterior mean estimators. Nevertheless, as highlighted by Xu and Ghosh (2015), these estimators are not inherently effective in sufficiently shrinking the regression coefficients toward zero. Consequently, these posterior mean estimators are less suited for the specific purpose of variable selection. While methods such as the utilisation of posterior credible intervals (Zhang et al., 2014) can potentially address this limitation, they do introduce additional complexity to the analysis. To avoid this complexity, an alternative way is to employ the posterior median (Xu and Ghosh, 2015) as an adaptive thresholding estimator. This approach presents distinct merits in terms of both selection and estimation. Xu and Ghosh (2015) showed that the use of posterior median thresholding possesses an oracle property under orthogonal designs, meaning that this thresholding correctly identifies which predictors should be included in a model and which should be excluded. Therefore, within this work, we advocate for the adoption of the posterior median estimator as the preferred method for the variable selection process within the domain of QR.

## 4. Simulation

Here, our proposed approaches are assessed through two main simulation studies: fixed effects selection and simultaneous selection (fixed and random effect).

### 4.1. Simulation study for fixed effects selection

To assess the efficacy of the proposed methodologies in Sections 3.3 and 3.4, a series of simulation studies were conducted. In this context, the number of group variables ($G$) was adjusted based on the specific contextual requirements of each simulation scenario. For each study, a sample size of 100 was employed. While larger samples could enhance estimation precision, the primary objective at this stage is to evaluate and validate the model's behaviour, rather than to generalise the findings to a broader population. Moreover, this sample size is considered sufficient to represent LCGD in realistic settings, particularly for preliminary methodological investigations. Subsequently, the model was fitted at three different quantiles, $\tau \in \{0.10, 0.50, 0.90\}$, which represent the lower, central, and upper locations of the response distribution (e.g. weight in the LCGD). Concerning the prior specifications of our proposed approaches (BGLSSQR and BSGSSQR), we chose to set $a_1 = a_2 = c_1, = c_2 = 1.0$ for the beta priors on both $\pi_0$ and $\pi_1$. Additionally, the inverse Gamma priors applied to $\sigma$ were configured with $g_1 = g_2 = 0.1$. For all Bayesian approaches, a Gibbs sampler was employed, running for a total of 50,000 iterations. To account for convergence issues, an initial burn-in period of 10,000 iterations was implemented.

As is commonly known, LASSO introduces bias into parameter estimates by shrinking the coefficients of some variables toward zero. Thus, we need to investigate this property in our proposed methods as well. Consequently, the first two simulation studies aim to examine the estimation performance of two proposed approaches (BGLSSQR and BSGSSQR) in comparison with both existing frequentist and Bayesian methods. In terms of the existing frequentist approaches, we consider the classical quantile regression (CQR) as introduced by Koenker (2005), and both LASSO methods for quantile regression, i.e. LASSO (LASSOQR) and group LASSO (GLASSOQR) approaches, as presented by Sherwood et al. (2023). The CQR was implemented using the `rq` function from the R package `quantreg` (Koenker, 2021), while the LASSOQR and GLASSOQR approaches were implemented using the `rq.pen(penalty = "LASSO")` function and `rq.group.pen(penalty = "gLASSO")` function, respectively, from the R package `rqPen` (Sherwood et al., 2023). It is important to note that the tuning parameter ($\lambda$) for both the LASSO and GLASSOQR methods was chosen based on the Bayesian information criterion (BIC) using the `qic.select(method="BIC")` function (Sherwood et al., 2023). Regarding the existing Bayesian approaches, we consider the Bayesian LASSO quantile regression (BLASSOQR) as reported in Alhamzawi and Ali (2020). The `BLqr` function from the R package `Brq` (Alhamzawi, 2020) was employed to perform the BLASSOQR approach. For both studies, three distinct error $\boldsymbol{\epsilon}$ distributions were considered: a standard normal distribution, a Student's t-distribution with $\nu = 3$ ($t_{\nu=3}$) and a Chi-squared distribution with $\nu = 3$ ($\chi^2_{\nu=3}$).

**Scenario I:** We treat model (7), which includes grouped continuous predictors, as the true data-generating model in our initial investigation. Specifically, we restrict attention to three variable groups ($G = 3$) to examine the behaviour of parameter estimates from the proposed methods relative to existing approaches. Among these groups, the first one was designed as the "active group", where each coefficient within the group took a value of one, indicating its significance. In contrast, the last group was designed as the "inactive group", wherein each coefficient within the group took a value of zero, indicating its lack of contribution to the response variable. Meanwhile, the second group was designed to contain at least one inactive coefficient. More specifically, the active coefficients were established with magnitudes categorised as weak ($\beta = 0.30$), moderate ($\beta = 0.50$) and strong ($\beta = 1.00$). Thus, the true regression coefficients were set to

$$\boldsymbol{\beta} = \Big( \underbrace{(1.00, 1.00, 1.00)}_{X_1}, \underbrace{(0.50, 0.30, 0.00)}_{X_2}, \underbrace{(0.00, 0.00, 0.00)}_{X_3} \Big).$$

A group structure with a length of 3 was chosen as the minimal structure capable of accommodating three magnitudes of coefficients. In this particular scenario, we assume that each predictor is

independent, following the i.i.d. standard normal distribution. Additionally, we aim to determine the number of updates and iterations for estimating the hyperparameters $\lambda$ and $t$ for both BGLSSQR and BSGSSQR through a Monte Carlo EM algorithm in two distinct scenarios: 100 and 1000.

**Scenario II:** We adopt one of the simulation designs of Yuan and Lin (2006) to specify a more complex model with polynomial terms, linear terms, and dummy predictors, allowing for multicollinearity among some covariates. We then extend the number of variable groups to four ($G = 4$): a group of polynomial basis functions for the continuous covariate $(X_1)$, two groups of continuous predictors $(X_2, X_3)$, and a group of dummy variables for the multi-level categorical covariate $(X_4)$. Analogous to Scenario I, we classify the magnitude of the active coefficients into categories of weak ($\beta \in \{0.30, 0.33\}$), moderate ($\beta \in \{0.50, 0.67\}$) and strong ($\beta \in \{1.00, 2.00\}$). Thus, the true model was

$$\mathbf{y} = 0.67X_1 - X_1^2 + 0.33X_1^3 + 0.50X_{2,1} + 0.30X_{2,2} + 0X_{3,1} + 0X_{3,2} + 0X_{3,3} + 2I(X_4 = 1) + I(X_4 = 2).$$

Here, $X_1$ and $X_3$ were generated by $X_p = \frac{Z_p + W}{\sqrt{2}}$, $Z_p \sim \mathcal{N}(0, 1)$, and $W \sim \mathcal{N}(0, 1)$. This implies that both $X_1$ and $X_3$ were correlated with each other. Each predictor of the $X_2$ (i.e. $X_{2,1}$, $X_{2,2}$, and $X_{2,3}$) was generated to follow the i.i.d standard normal distribution. $X_4$ was generated by trichotomising $X_1$ as 0, 1, 2 under the conditions that $X_4 = 0$ if $X_1 < \Phi^{-1}(1/3), X_4 = 1$ if $X_1 > \Phi^{-1}(2/3)$, and $X_4 = 2$ otherwise; thus, $I(X_4 = 1)$ and $I(X_4 = 2)$ indicate whether $X_4$ is either 1 or 2. Therefore, the true regression coefficients of this model were

$$\boldsymbol{\beta} = \Bigg( \underbrace{(0.67, -1, 0.33)}_{X_1}, \underbrace{(0.50, 0.30, 0.00)}_{X_2}, \underbrace{(0.00, 0.00, 0.00)}_{X_3}, \underbrace{(2.00, 1.00)}_{X_4} \Bigg).$$

To summarise the performance of estimators, we consider mean bias errors (MBE) and root-mean-square errors (RMSE) for $\boldsymbol{\beta}_\tau$ given by $\text{MBE}(\beta_\tau) = (1/500) \sum_{r=1}^{500} (\hat{\beta}_{\tau,r} - \beta_{\tau,r})$, $\text{RMSE}(\beta_{\tau,h}) = \sqrt{(1/500) \sum_{r=1}^{500} (\hat{\beta}_{\tau,r} - \beta_{\tau,r})^2}$.

Our focus now shifts to the evaluation of subset selection and predictive performances in the remaining studies, where we compare our proposed approaches with existing methods. In particular, we begin with two existing frequentist approaches, i.e. LASSOQR and GLASSOQR. Subsequently, we delve into the comparison of relevant existing Bayesian variable selection methods. One such existing method is the model selection based on credible intervals (MSCI) employed for the BLASSOQR approach. The MSCI method was implemented using the `model` function provided by the R package `Brq` (Alhamzawi, 2020). In these simulation studies, we restrict our study to the standard normal distribution for the errors. The assessment of each approach was based on two rates and one score: the true positive rate (TPR), the false positive rate (FPR) and $F_1$ score. These matrices were calculated using the information provided in the confusion matrix, as presented in Table 1. Additionally, to assess the predictive performance, we employ the mean weighted absolute errors (MAE), $\text{MAE}_\tau = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \epsilon^*_{ij,\tau} (\tau - I(\epsilon^*_{ij,\tau} < 0)) / (\sum_i^N n_i)$, where $\epsilon^*_{ij,\tau} = y_{ij} - \hat{Q}_{y_{ij}|t_{ij}}(\tau)$ and $I(\cdot)$ is an indicator function, to measure the variation in the predicted values around the observations. Smaller MAE values indicate more accurate predictions.

**Scenario III:** We extend the design to four variable groups $(G = 4)$ and increase the number of predictors within each group to five, as an initial setting for studying variable selection. The true regression coefficients were denoted as

$$\boldsymbol{\beta} = \Big( \underbrace{(0.30, -1.00, 0.00, 0.50, 0.01)}_{X_1}, \underbrace{(0, 0, 0, 0, 0)}_{X_2}, \underbrace{(0.80, 0.80, 0.80, 0.80, 0.80)}_{X_3}, \underbrace{(0, 0, 0, 0, 0)}_{X_4} \Big).$$

**Table 1** Confusion Matrix

| | # $(\beta_p \neq 0)$ | # $(\beta_p = 0)$ |
|---|---|---|
| # $(\hat{\beta}_p \neq 0)$ | True Positive (TP) | False Positive (FP) |
| # $(\hat{\beta}_p = 0)$ | False Negative (FN) | True Negative (TN) |
| $\textbf{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$ | $\textbf{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$ | $F_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$ |

Here, $X_1$ and $X_3$ are treated as active groups, and we consider two scenarios for $\boldsymbol{\beta}$: (i) a mix of strong ($\beta = -1.00$), moderate ($\beta = 0.50$) and weak ($\beta = 0.30$) effects with both positive and negative signs, and (ii) effects of uniformly strong positive magnitude ($\beta = 0.80$). The Predictors were generated to follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance $\boldsymbol{\Sigma}_x$, where off-diagonal elements of $\boldsymbol{\Sigma}_x$ are defined as $0.5^{|i-j|}$ for $i \neq j$. This implies that predictors closer in sequence have a higher correlation.

**Scenario IV:** For the high sparsity design, we expand the number of variable groups to ten ($G = 10$). The true regression coefficients of this model were

$$\boldsymbol{\beta} = \Bigg( \underbrace{(0.67, -1.00, 0.33)}_{X_1}, \underbrace{(0.5, 0.30, 0.00)}_{X_2}, \underbrace{\mathbf{0}}_{X_3}, \underbrace{(2,1)}_{X_4}, \underbrace{\mathbf{0}}_{X_5}, \underbrace{\mathbf{0}}_{X_6}, \underbrace{\mathbf{0}}_{X_7}, \underbrace{\mathbf{0}}_{X_8}, \underbrace{\mathbf{0}}_{X_9}, \underbrace{\mathbf{0}}_{X_{10}} \Bigg).$$

We treat $X_1$, $X_2$, and $X_4$ as active groups: $X_1$ contains mixed effect sizes, $X_2$ includes both active and inactive coefficients, and $X_4$ has strong effects. Here, $\mathbf{0}$ denotes a zero vector of length 3 for continuous variables and length 2 for categorical variables (dummy variables). For the third-degree polynomial of $X_1, X_5, X_7$, we generate these variables using $X_p = \frac{Z_p + W}{\sqrt{2}}$, $Z_p \sim \mathcal{N}(0,1)$, and $W \sim \mathcal{N}(0,1)$. Each predictor of $X_2$, $X_3$, $X_9$ and $X_{10}$ was generated from $\mathcal{N}_4(\mathbf{0}, \boldsymbol{\Sigma}_{X.})$, where $\boldsymbol{\Sigma}_{X.} = \text{diag}(1,1,1,1)$. For each dummy predictor of $X_4$, $X_6$ and $X_8$, we trichotomised $X_1, X_5, X_7$ to meet the conditions that $X_p < \Phi^{-1}(1/3) = 0, X_p > \Phi^{-1}(2/3) = 1$, and otherwise equal to 2. Subsequently, we set dummy predictors as $I(X_4 = 1), I(X_4 = 2), I(X_6 = 1), I(X_6 = 2), I(X_8 = 1)$ and $I(X_8 = 2)$.

### 4.2. Simulation study for simultaneous selection

Subsequently, our attention shifted towards evaluating the effectiveness of simultaneous selection using BSGSSMQR in Section 3.5 through a comprehensive simulation study. We thoroughly investigate the performance in prediction and subset selection, with specific emphasis on both fixed and random effects. Additionally, we compare this proposed method with AQMM. The simulated data were generated by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + u_{i1}z_{ij1} + u_{i2}z_{ij2} + u_{i3}z_{ij3} + u_{i4}z_{ij4} + \epsilon_{ij}, \quad i = 1, \ldots, 50, j = 1, \ldots, 10,$$

where the fixed effect component follows the specifications, as outlined in Scenario IV. The residuals $\epsilon_{ij}$ were generated from $\mathcal{N}(0, \sigma_\epsilon^2)$ with $\sigma_\epsilon^2$ set to two scenarios: $\sigma_\epsilon^2 = 1.00$ and $\sigma_\epsilon^2 = 9.00$. Specifically, $z_{ij1}$ took the form of a column vector of ones with dimension $\mathbf{1}_n$. Additionally, $z_{ijl}$, with $l$ ranging from 2 to 4, were generated from a uniform distribution $\mathcal{U}(-2, 2)$. We set $\boldsymbol{u} = (u_{i1}, u_{i2}, u_{i3}, u_{i4})' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$, where

$$\boldsymbol{\Sigma}_{\boldsymbol{u}} = \begin{pmatrix} 0.90 & 0.40 & 0.06 & 0.00 \\ 0.40 & 0.50 & 0.10 & 0.00 \\ 0.06 & 0.10 & 0.20 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{pmatrix}.$$

For the Beta priors for $\pi_0$ and $\pi_1$, and inverse Gamma priors for $\sigma$, we follow the same prior specifications as mentioned in Section 4.1. The number of updates and iterations for estimating the hyperparameter $t$ of $s^2$ through a Monte Carlo EM algorithm was designated as 100. Furthermore, we fix $p_{l0}$ of $d_l$ to 0.5, while the prior mean and variance of $\boldsymbol{a}$ were set as $\mathbf{0}$ and $0.5\mathbf{I}$. We employ a Gibbs sampler for a total of 25,000 iterations, with an initial burn-in period of 5,000 iterations, considering a total of 200 simulations.

Regarding the additive quantile mixed model (AQMM), it should be noted that the `aqmm` function from the R package `aqmm` allows only for fitting of non-linear terms through the `smooth.terms` specification in the `mgcv` package. Consequently, variables associated with a third-degree polynomial, such as $X_1, X_5$, and $X_7$, were modelled as cubic P-spines on 40 equidistant knots distributed across the range of each variable. Other variables were assumed to be linear predictors. To select the variables or predictors, the $p$-values for a two-tailed $t$-test from the bootstrap method with 50 replications were employed, on the understanding that if this $p$-value is less than the significance level of 0.05, the variable will be included in the model. To ensure fairness in comparing BSGSSMQR and AQMM, we exclude predictors from the group variables $X_1, X_5$ and $X_7$ when measuring subset selection performance. This decision was made because the bootstrap testing in AQMM does not support hypothesis tests of smooth terms.

## 5. Results

### 5.1. Results of simulation studies

In Scenario I, classical quantile regression (CQR) consistently produced estimates closer to the true parameters than the other methods, as shown in Figure 1, where the mean bias error (MBE) values are close to zero. This pattern held across three distinct quantile levels, multiple error distributions, and varying coefficient magnitudes (weak, medium, and strong). In contrast, both frequentist and Bayesian LASSO methods—including our proposed approaches (BGLSSQR and BSGSSQR)—tended to consistently underestimate the coefficients. These methods also exhibited sensitivity to non-negative error distributions ($\chi^2$-error), resulting in biased estimates with high variability. Nevertheless, they demonstrated robustness against heavy-tailed errors ($t_3$-error). Furthermore, there was no discernible difference in performance between 100 and 1000 updates and iterations when estimating $\lambda$ and $t$ via the Monte Carlo EM method (results not shown). A similar pattern was observed in Scenario II, which introduced multicollinearity in some predictors, as presented in Figure 2.

Tables 2 and 3 summarise the variable selection metrics (i.e. mean weighted absolute error (MAE), true positive rate (TPR), false positive rate (FPR), and $F_1$ score) in Scenarios III and IV, comparing the performance seven methods on the simulated data. The BLASSOQR method achieved the best predictive performance across all quantile levels, closely followed by BGLSSQR and BSGSSQR methods. In contrast, LASSOQR and GLASSOQR exhibited higher prediction errors, particularly in the extreme quantile models. In terms of model selection accuracy, the proposed methods outperformed other approaches across various quantile levels and simulation studies. BSGSSQR achieved a favourable balance by maintaining a lower FPR alongside a superior $F_1$ score, whereas BGLSSQR tended to produce a higher FPR. The frequentist methods, LASSOQR and GLASSOQR, failed to effectively control the FPR, resulting in unacceptably low $F_1$ scores.

In the simultaneous selection study, Table 4 presents the mean of the mean weighted absolute errors (MAE) under two error variance scenarios ($\sigma_\epsilon^2$). The results show that both the posterior mean and posterior median estimators yielded similar prediction errors across the two levels of error variance. However, both estimators exhibited slightly higher prediction errors compared to the AQMM. Regarding selection performance, summarised in Table 5, the posterior median estimator proved to be a robust choice, effectively serving as a thresholding rule for selecting or excluding predictors. As expected, the performance of BSGSSMQR declined when the data contained high levels of noise. In contrast, AQMM consistently underperformed across all selection accuracy metrics.

In addition, Table 6 presents the mean TPR, FPR and $F_1$ score obtained from a sensitivity analysis of BSGSSMQR, in which the hyperparameters $\pi_0, \pi_1$, and $d_{l'}$ were varied across nine distinct

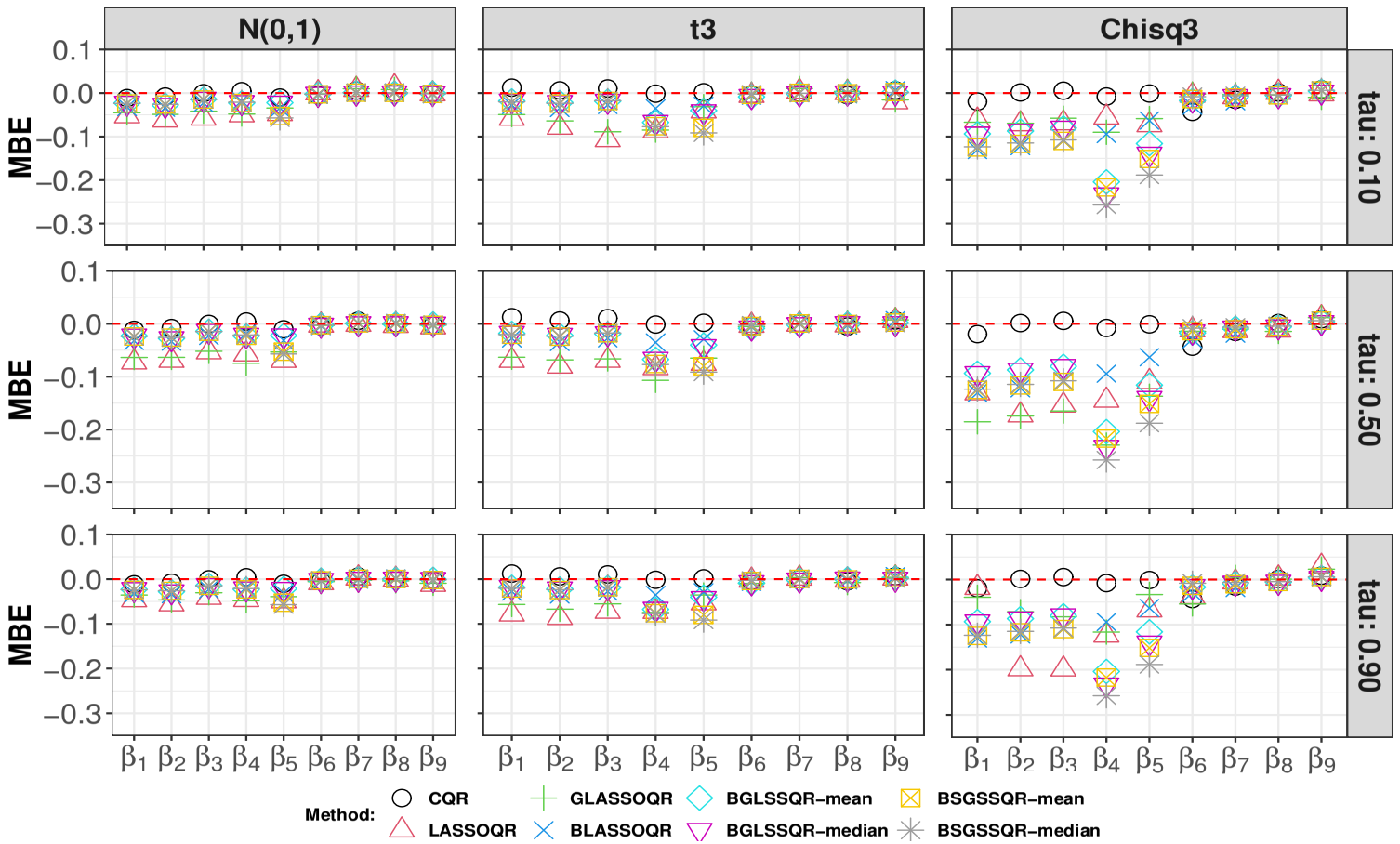



**Figure 1** Mean Bias errors (MBE) were obtained from eight different methods across three quantiles (**tau**) and three error distribution scenarios—standard normal distribution (**N(0,1)**), Student's t-distribution with $\nu = 3$ (**t3**), and chi-squared distribution with $\nu = 3$ (**Chisq3**)—under the Scenario I setting

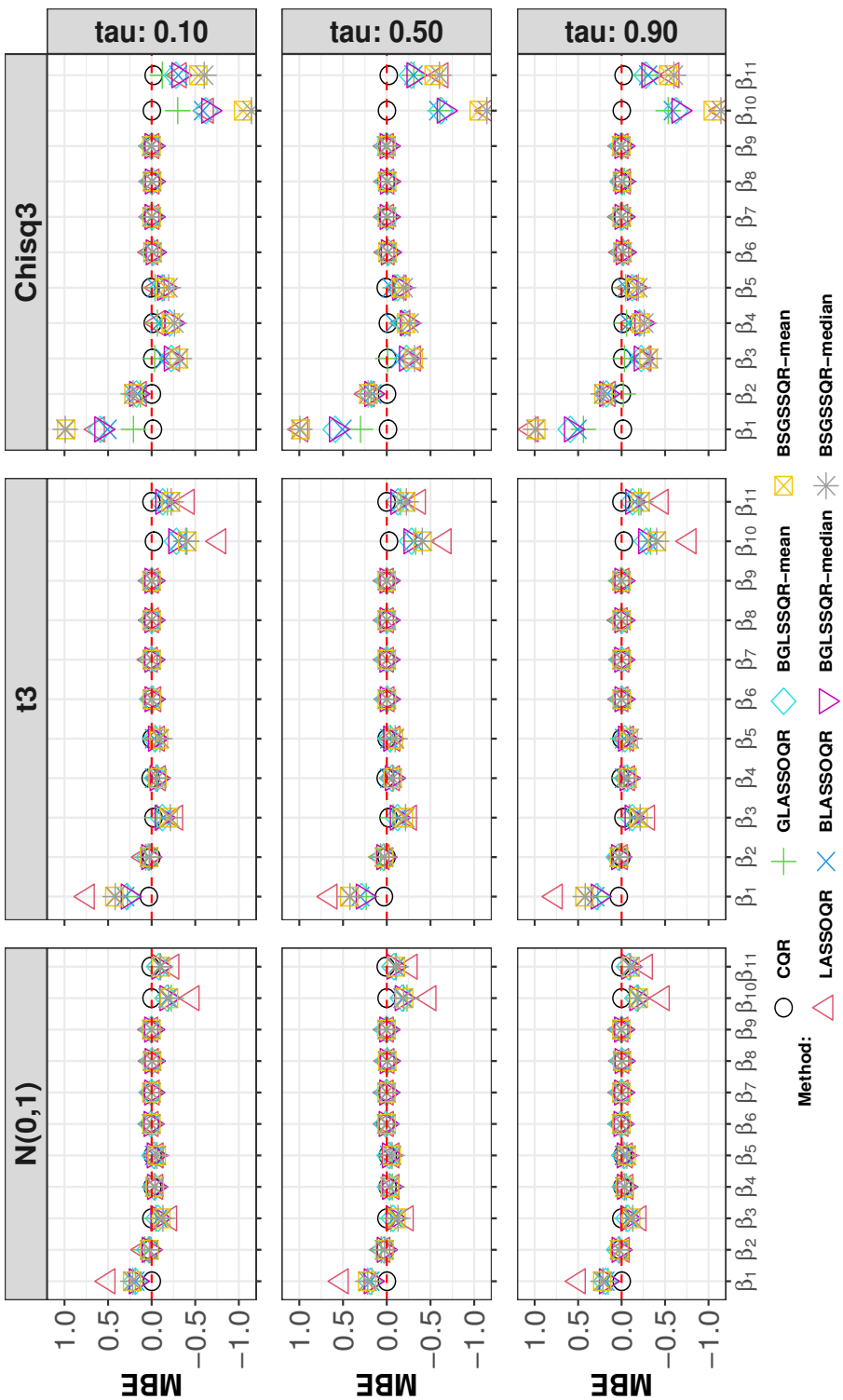



**Figure 2** Mean Bias errors (MBE) were obtained from eight different methods across three quantiles (**tau**) and three error distribution scenarios—standard normal distribution (**N(0,1)**), Student's t-distribution with $\nu = 3$ (**t3**), and chi-squared distribution with $\nu = 3$ (**Chisq3**)—under the Scenario II setting

**Table 2** Mean MAE, true positive (TPR) rate, false positive (FPR) rate and $F_1$ score for five different methods under Scenario III

| $\tau$ | | LASSOQR | GLASSOQR | BLASSOQR | BGLSSQR | | BSGSSQR | |
|---|---|---|---|---|---|---|---|---|
| | | | | | mean | median | mean | median |
| **0.10** | MAE (SD) | 0.42 (0.04) | 0.41 (0.04) | **0.35 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.94 | **1.00** | 0.83 | **1.00** | | 0.94 | |
| | FPR | 0.55 | 0.78 | **0.04** | 0.11 | | 0.06 | |
| | $F_1$ | 0.73 | 0.70 | 0.88 | **0.94** | | **0.94** | |
| **0.50** | MAE (SD) | 0.36 (0.03) | 0.37 (0.04) | **0.35 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.93 | **1.00** | 0.83 | **1.00** | | 0.94 | |
| | FPR | 0.37 | 0.39 | **0.04** | 0.11 | | 0.06 | |
| | $F_1$ | 0.79 | 0.83 | 0.88 | **0.94** | | **0.94** | |
| **0.90** | MAE (SD) | 0.42 (0.04) | 0.41 (0.04) | **0.35 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.94 | **1.00** | 0.83 | **1.00** | | 0.94 | |
| | FPR | 0.57 | 0.76 | **0.04** | 0.11 | | 0.06 | |
| | $F_1$ | 0.72 | 0.70 | 0.88 | **0.94** | | **0.94** | |

**Table 3** Mean MAE, true positive (TPR) rate, false positive (FPR) rate and $F_1$ score for five different methods under Scenario IV

| $\tau$ | | LASSOQR | GLASSOQR | BLASSOQR | BGLSSQR | | BSGSSQR | |
|---|---|---|---|---|---|---|---|---|
| | | | | | mean | median | mean | median |
| **0.10** | MAE (SD) | 0.42 (0.04) | 0.42 (0.05) | **0.33 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.87 | 0.98 | 0.77 | **0.99** | | 0.92 | |
| | FPR | 0.48 | 0.50 | **0.03** | 0.08 | | **0.03** | |
| | $F_1$ | 0.55 | 0.62 | 0.83 | 0.90 | | **0.91** | |
| **0.50** | MAE (SD) | 0.36 (0.04) | 0.39 (0.04) | **0.33 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.84 | 0.98 | 0.77 | **0.99** | | 0.92 | |
| | FPR | 0.23 | 0.20 | **0.03** | 0.07 | | **0.03** | |
| | $F_1$ | 0.65 | 0.79 | 0.83 | 0.90 | | **0.91** | |
| **0.90** | MAE (SD) | 0.42 (0.04) | 0.42 (0.04) | **0.33 (0.03)** | 0.36 (0.03) | 0.37 (0.03) | 0.36 (0.03) | 0.37 (0.03) |
| | TPR | 0.87 | **0.99** | 0.77 | **0.99** | | 0.92 | |
| | FPR | 0.50 | 0.55 | **0.03** | 0.08 | | **0.03** | |
| | $F_1$ | 0.53 | 0.60 | 0.83 | 0.90 | | **0.91** | |

**Table 4** Mean MAE (SD) for AQMM and BSGSSMQR with two posterior estimators in two error variances for the simulation study in Section 4.2

| $\tau$ | | **AQMM** | **BSGSSMQR**$_{mean}$ | **BSGSSMQR**$_{median}$ |
|---|---|---|---|---|
| **0.10** | $\sigma_\epsilon^2 = 1$ | 0.15 (0.01) | 0.23 (0.02) | 0.24 (0.02) |
| | $\sigma_\epsilon^2 = 9$ | 0.49 (0.03) | 0.95 (0.05) | 0.97 (0.05) |
| **0.50** | $\sigma_\epsilon^2 = 1$ | 0.33 (0.02) | 0.35 (0.02) | 0.36 (0.02) |
| | $\sigma_\epsilon^2 = 9$ | 1.08 (0.06) | 1.01 (0.04) | 1.02 (0.04) |
| **0.90** | $\sigma_\epsilon^2 = 1$ | 0.15 (0.01) | 0.23 (0.02) | 0.24 (0.02) |
| | $\sigma_\epsilon^2 = 9$ | 0.48 (0.03) | 0.95 (0.04) | 0.97 (0.05) |

**Table 5** Mean true positive rate (TPR), false positive rate (FPR) and $F_1$ score for AQMM and BSGSSMQR in two error variances for the simulation study in Section 4.2

| | $\sigma_\epsilon^2 = 1$ | | | $\sigma_\epsilon^2 = 9$ | | |
|---|---|---|---|---|---|---|
| | **AQMM** | **BSGSSMQR**$_f$ | **BSGSSMQR**$_a$ | **AQMM** | **BSGSSMQR**$_f$ | **BSGSSMQR**$_a$ |
| $\tau = \mathbf{0.10}$ | | | | | | |
| TPR | 0.53 | 1.00 | 1.00 | 0.27 | 0.80 | 0.85 |
| FPR | 0.14 | 0.02 | 0.03 | 0.15 | 0.09 | 0.12 |
| $F_1$ | 0.54 | 0.97 | 0.97 | 0.43 | 0.75 | 0.81 |
| $\tau = \mathbf{0.50}$ | | | | | | |
| TPR | 0.66 | 1.00 | 1.00 | 0.34 | 0.80 | 0.85 |
| FPR | 0.06 | 0.02 | 0.03 | 0.05 | 0.09 | 0.12 |
| $F_1$ | 0.71 | 0.97 | 0.97 | 0.49 | 0.75 | 0.80 |
| $\tau = \mathbf{0.90}$ | | | | | | |
| TPR | 0.77 | 1.00 | 1.00 | 0.45 | 0.80 | 0.88 |
| FPR | 0.07 | 0.02 | 0.03 | 0.11 | 0.09 | 0.15 |
| $F_1$ | 0.77 | 0.97 | 0.97 | 0.54 | 0.75 | 0.80 |

Note: The term $f$ represents only fixed effects were considered and the term $a$ represents both fixed and random effects were considered.

scenarios. The results indicate that BSGSSMQR exhibited strong stability across all three metrics for most of the variables and across three quantile models.

### 5.2. Results of illustrative analysis

We conduct an illustrative analysis using BSGSSMQR on two simulated datasets corresponding to two levels of error variance, $\sigma_\epsilon^2 = 1$ and $\sigma_\epsilon^2 = 9$, as described in Section 4.2. The results for the case of $\sigma_\epsilon^2 = 1$ are presented in Table 7. We observe that the BSGSSMQR method correctly identifies both fixed and random effects when the posterior median is used as the thresholding estimator. However, the method shows increased sensitivity to large error variance ($\sigma_\epsilon^2 = 9$), leading to a slight decline in overall performance and a more pronounced impact on the selection of random effects (see Table 10 in Appendix III). In addition, we assess the computational efficiency of the method under these settings. Using a simulated dataset with 500 observations, 27 predictors, 2 random effects, and 25,000 Gibbs sampler iterations, the BSGSSMQR approach required approximately 120 minutes of computational time.

**Table 6** Sensitivity analysis for BSGSSMQR

| $\tau$ | $p_{l0}$ | **Hyperprior** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $a_1, a_2, c_1, c_2 = 0.50$ | | | $a_1, a_2, c_1, c_2 = 1.00$ | | | $a_1, a_2, c_1, c_2 = 1.50$ | | |
| | | TPR | FPR | $F_1$ | TPR | FPR | $F_1$ | TPR | FPR | $F_1$ |
| **0.10** | 0.30 | 0.96 | 0.04 | 0.94 | 0.96 | 0.04 | 0.94 | 0.96 | 0.04 | 0.94 |
| | 0.50 | 0.96 | 0.03 | 0.95 | 0.96 | 0.03 | 0.95 | 0.96 | 0.02 | 0.95 |
| | 0.80 | 0.96 | 0.02 | 0.96 | 0.96 | 0.02 | 0.96 | 0.96 | 0.02 | 0.96 |
| **0.50** | 0.30 | 0.97 | 0.03 | 0.95 | 0.97 | 0.03 | 0.95 | 0.97 | 0.03 | 0.95 |
| | 0.50 | 0.97 | 0.03 | 0.96 | 0.98 | 0.03 | 0.96 | 0.97 | 0.02 | 0.96 |
| | 0.80 | 0.97 | 0.02 | 0.97 | 0.97 | 0.02 | 0.97 | 0.97 | 0.02 | 0.97 |
| **0.90** | 0.30 | 0.97 | 0.04 | 0.95 | 0.97 | 0.04 | 0.95 | 0.97 | 0.04 | 0.95 |
| | 0.50 | 0.97 | 0.03 | 0.96 | 0.97 | 0.03 | 0.96 | 0.97 | 0.02 | 0.96 |
| | 0.80 | 0.97 | 0.02 | 0.96 | 0.97 | 0.02 | 0.97 | 0.97 | 0.02 | 0.97 |

## 6. Application to longitudinal child growth data in Scotland

We apply the BSGSSMQR approach to the Growing up in Scotland (GUS) data (Bradshaw et al., 2007), which were collected in eleven sweeps between 2005/06 and 2021/23 from Scottish children (born in 2002/03, 2004/05, and 2010/11) participating in the family longitudinal survey. The GUS data are available via the UK Data Service (https://ukdataservice.ac.uk/) under special licence access. However, no physical growth measurements were collected during the 2nd, 3rd, or 5th sweep. Consequently, the data used in this paper consist of measurements from only seven sweeps. After removing data from these three sweeps and conducting data cleaning, we are left with a sample of 4,563 children, including 2,326 males and 2,237 females. In this paper, we focus the analysis exclusively on identifying risk factors associated with weight measurements at two distinct locations ($\tau = 0.10$ and $0.90$), representing the lower and upper bounds of weight growth for children, respectively. These quantiles help identify children who fall in the lower or higher ends of the weight distribution, often used in growth monitoring and assessments of underweight and overweight categories. We focus on this growth measure because childhood obesity has become a concern in Scotland. This choice was motivated by the Scottish Health Survey Report highlighted an increase in the prevalence of childhood obesity in 2021 compared to the 2017 figures (Birtwistle et al., [ISBN: 9781835216569]. 2023). Also, we aim to determine the appropriate variability of individual linear trends (intercepts and age slopes) through the selection of random effects. The potential risk factors (independent variables) were selected by mapping variables in GUS to the framework outlines in "Multiple Risk Factors in Young Children's Development" by Sabates and Dex (2015). This framework utilises an ecological model of child development, helping to pinpoint risk factors associated with child development. Furthermore, we consider other potential variables relating to children and their mother demographic data. The list of these variables is presented in Table 8.

Within this analytical framework, age was exclusively incorporated as a fixed smooth component in each quantile model as growth patterns tended to follow non-linear trajectories (see Figure 3). This was achieved by using the cubic B-spline method to establish the bases for this smooth term. The determination of the number of knots ($K$) adhered to the guidelines of Ruppert et al. (2003), setting $K$ as $\min\{40,$ the number of unique $x/4\}$. These knots were strategically placed at appropriate quantiles of the variable age. All other independent variables in each model were assumed to be linear. Additionally, raw weight was centred by subtracting its mean, and all covariates, including the basis functions for age, were standardised. Furthermore, a design matrix of random effects, $\mathbf{Z} = [\mathbf{1}, X_1]$, wherein $X_1$ represents the variable of age, was established. This configuration serves as an instrument to account for the inherent potential heterogeneity among children by incorporating random intercepts

**Table 7** Posterior mean, posterior median for both the fixed (FEs) and random (REs) effects, and computational times (in minutes) of three quantile models under the scenario $\sigma_\epsilon^2 = 1$.

| | | $\tau = 0.10$ | | | $\tau = 0.50$ | | | $\tau = 0.90$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **True** | **Mean** | **Median** | **SD** | **Mean** | **Median** | **SD** | **Mean** | **Median** | **SD** |
| **FEs** | | | | | | | | | | |
| $\beta_1$ | 0.67 | 0.83 | **0.84** | 0.30 | 0.75 | **0.75** | 0.28 | 0.75 | **0.74** | 0.29 |
| $\beta_2$ | -1.00 | -1.00 | **-1.00** | 0.10 | -1.00 | **-1.00** | 0.10 | -0.99 | **-0.99** | 0.10 |
| $\beta_3$ | 0.33 | 0.28 | **0.30** | 0.20 | 0.32 | **0.34** | 0.18 | 0.33 | **0.34** | 0.20 |
| $\beta_4$ | 0.50 | 0.49 | **0.49** | 0.09 | 0.48 | **0.48** | 0.08 | 0.48 | **0.48** | 0.08 |
| $\beta_5$ | 0.30 | 0.36 | **0.36** | 0.08 | 0.36 | **0.36** | 0.08 | 0.37 | **0.37** | 0.08 |
| $\beta_6$ | 0.00 | 0.00 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 |
| $\beta_7$ | 0.00 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 |
| $\beta_8$ | 0.00 | -0.05 | 0.00 | 0.07 | -0.04 | 0.00 | 0.06 | -0.03 | 0.00 | 0.06 |
| $\beta_9$ | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| $\beta_{10}$ | 2.00 | 1.87 | **1.86** | 0.20 | 1.93 | **1.93** | 0.18 | 1.93 | **1.93** | 0.19 |
| $\beta_{11}$ | 1.00 | 0.94 | **0.94** | 0.12 | 0.97 | **0.97** | 0.11 | 0.98 | **0.98** | 0.12 |
| $\beta_{12}$ | 0.00 | 0.05 | 0.00 | 0.11 | 0.06 | 0.00 | 0.11 | 0.06 | 0.00 | 0.11 |
| $\beta_{13}$ | 0.00 | 0.02 | 0.00 | 0.05 | 0.02 | 0.00 | 0.05 | 0.02 | 0.00 | 0.05 |
| $\beta_{14}$ | 0.00 | -0.05 | 0.00 | 0.10 | -0.06 | 0.00 | 0.11 | -0.05 | 0.00 | 0.10 |
| $\beta_{15}$ | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.05 |
| $\beta_{16}$ | 0.00 | 0.04 | 0.00 | 0.07 | 0.03 | 0.00 | 0.06 | 0.03 | 0.00 | 0.06 |
| $\beta_{17}$ | 0.00 | -0.04 | 0.00 | 0.09 | -0.04 | 0.00 | 0.09 | -0.04 | 0.00 | 0.09 |
| $\beta_{18}$ | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| $\beta_{19}$ | 0.00 | -0.02 | 0.00 | 0.06 | -0.01 | 0.00 | 0.06 | -0.02 | 0.00 | 0.07 |
| $\beta_{20}$ | 0.00 | 0.02 | 0.00 | 0.06 | 0.03 | 0.00 | 0.07 | 0.02 | 0.00 | 0.07 |
| $\beta_{21}$ | 0.00 | -0.02 | 0.00 | 0.06 | -0.02 | 0.00 | 0.06 | -0.02 | 0.00 | 0.05 |
| $\beta_{22}$ | 0.00 | 0.01 | 0.00 | 0.04 | 0.01 | 0.00 | 0.04 | 0.01 | 0.00 | 0.04 |
| $\beta_{23}$ | 0.00 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 |
| $\beta_{24}$ | 0.00 | 0.00 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 |
| $\beta_{25}$ | 0.00 | -0.01 | 0.00 | 0.04 | -0.01 | 0.00 | 0.04 | -0.02 | 0.00 | 0.05 |
| $\beta_{26}$ | 0.00 | 0.04 | 0.00 | 0.06 | 0.04 | 0.00 | 0.06 | 0.04 | 0.00 | 0.06 |
| $\beta_{27}$ | 0.00 | 0.02 | 0.00 | 0.05 | 0.02 | 0.00 | 0.05 | 0.03 | 0.00 | 0.06 |
| **REs** | | | | | | | | | | |
| $\sigma_{11}$ | 0.90 | 0.87 | **0.84** | 0.21 | 0.80 | **0.77** | 0.20 | 0.88 | **0.86** | 0.21 |
| $\sigma_{22}$ | 0.50 | 0.47 | **0.46** | 0.11 | 0.47 | **0.46** | 0.11 | 0.48 | **0.47** | 0.11 |
| $\sigma_{33}$ | 0.20 | 0.12 | **0.11** | 0.05 | 0.11 | **0.10** | 0.05 | 0.12 | **0.12** | 0.05 |
| $\sigma_{44}$ | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 |
| **Time** | | | 124.00 | | | 132.00 | | | 133.00 | |

and slopes. Regarding the selected prior specifications, we opt for the same specifications of the Beta priors and the inverse Gamma priors, as outlined in Section 4.2. The hyperparameter $t$ of $s^2$ was estimated through a Monte Carlo EM algorithm with 100 updates and iterations. To ensure adequate convergence and accuracy, we ran the Gibbs sampler for 15,000 iterations, with a burn-in period of 5000 iterations. Furthermore, we ran three independent MCMC chains and assessed convergence by inspecting trace plots and computing the Gelman-Rubin diagnostic ($\hat{R}_{GR}$) (Gelman and Rubin, 1992). We also evaluated the independence of the samples using the effective sample size ($n_{ESS}$) (Gelman, 2014).

Table 9 presents the posterior mean and median estimates of the basis functions of age and selected linear fixed effects related to centred raw weight across two quantile models. The results indicate that the posterior median estimator produced zero estimates for the basis function, S(Age22),

**Table 8** List of independent variables

| Variable | Name | Variable | Name |
|---|---|---|---|
| *age* | Age of child | *nhhs* | Number of people in household |
| *sex* | Sex of child | *equi* | Equivalised income |
| *lbwe* | Low birth weight | *depr* | Scottish Index of Multiple Deprivation |
| *ceth* | Ethnicity of a child | *pjob* | Current job status of parent(s) |
| *chea* | Child's health in general | *phea* | Health problem in a year of parent(s) |
| *nacj* | No of accidents/injuries of child | *pche* | Current health status of parent(s) |
| *cbor* | Child's birth order | *mral* | Drinking alcohol during pregnancy |
| *mmrs* | Mother's marital status | *msmo* | Smoking cigarettes during pregnancy |
| *urcl* | Urban-rural classification | *pral* | Alcohol drink status of parent(s) |
| *mage* | Age of mother at first child's birth | | |

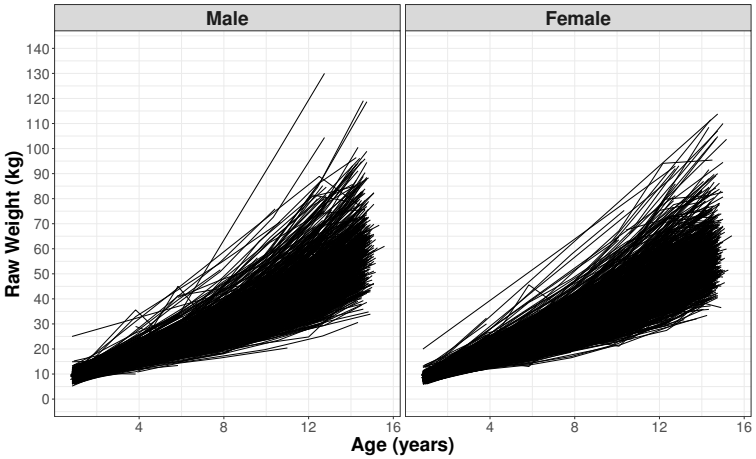



**Figure 3** Individual raw weight trajectories from the GUS data by sex

across both quantile models. Another interesting result is that the two quantile models appeared to have similar estimates of basis functions. The linear (predictor) fixed effects, including low birth weight, deprivation, and equivalised income, were selected for both 0.10th and 0.90th quantile models. Regarding the random effects, both the random intercept for age and the random slope of age were selected in all two quantile models, with the former exhibiting higher variability than the latter. This indicates that there was variability in individual linear trends (intercepts and temporal slopes) across the two quantiles.

With regard to the assessment of the MCMC samples for both quantile models (see Figure 5 in Appendix III), the trace plots indicated convergence, with all chains fluctuating around a stable mean and exhibiting no discernible trends over time. In particular, the plots for the three linear fixed predictors (risk factors) showed slight pseudo floor or ceiling effects approaching zero, suggesting that the slab component of the spike-and-slab prior predominantly influenced the selection of each fixed variable. All $\hat{R}_{GR}$ values were equal to 1, indicating that the chains have likely converged to the same posterior distribution and are well mixed. Moreover, each $n_{ESS}$ value exceeded 1,000, which generally reflects a sufficient number of effectively independent samples, thereby suggesting that the MCMC sampling process was robust and that the resulting parameter estimates are likely to be reliable.

**Table 9** Posterior mean and posterior median for fixed and random effects across two quantile models for the **centred raw weight**

| | $\tau = 0.10$ | | | $\tau = 0.90$ | | |
|---|---|---|---|---|---|---|
| | **Mean** | **Median** | **SD** | **Mean** | **Median** | **SD** |
| **Fixed effects** | | | | | | |
| S(Age1) | -3.28 | -3.23 | 0.62 | -3.25 | -3.20 | 0.63 |
| S(Age2) | 3.42 | 3.38 | 0.45 | 3.43 | 3.38 | 0.48 |
| S(Age3) | 0.83 | 0.84 | 0.34 | 0.69 | 0.72 | 0.38 |
| S(Age4) | 1.71 | 1.71 | 0.25 | 1.78 | 1.77 | 0.27 |
| S(Age5) | 0.48 | 0.47 | 0.35 | 0.46 | 0.45 | 0.35 |
| S(Age6) | 3.14 | 3.13 | 0.23 | 3.10 | 3.08 | 0.24 |
| S(Age7) | 2.63 | 2.66 | 0.25 | 2.55 | 2.58 | 0.26 |
| S(Age8) | 1.82 | 1.82 | 0.37 | 1.81 | 1.81 | 0.37 |
| S(Age9) | 0.58 | 0.66 | 0.38 | 0.55 | 0.62 | 0.38 |
| S(Age10) | 0.49 | 0.51 | 0.23 | 0.48 | 0.50 | 0.24 |
| S(Age11) | 1.16 | 1.39 | 0.47 | 1.25 | 1.36 | 0.47 |
| S(Age12) | 4.58 | 4.57 | 0.24 | 4.50 | 4.51 | 0.24 |
| S(Age13) | 4.15 | 4.16 | 0.25 | 4.06 | 4.07 | 0.25 |
| S(Age14) | 3.43 | 3.45 | 0.36 | 3.39 | 3.43 | 0.36 |
| S(Age15) | 0.99 | 1.17 | 0.51 | 0.97 | 1.15 | 0.51 |
| S(Age16) | 1.13 | 1.14 | 0.23 | 1.10 | 1.10 | 0.24 |
| S(Age17) | 2.07 | 2.17 | 0.50 | 2.05 | 2.14 | 0.49 |
| S(Age18) | 4.74 | 4.74 | 0.36 | 4.68 | 4.68 | 0.36 |
| S(Age19) | 5.54 | 5.55 | 0.23 | 5.42 | 5.43 | 0.24 |
| S(Age20) | 4.24 | 4.26 | 0.34 | 4.20 | 4.22 | 0.34 |
| S(Age21) | 1.23 | 1.33 | 0.48 | 1.20 | 1.31 | 0.48 |
| S(Age22) | 0.16 | 0.00 | 0.20 | 0.16 | 0.00 | 0.20 |
| S(Age23) | 0.50 | 0.50 | 0.17 | 0.50 | 0.49 | 0.17 |
| *lbwe* | -0.35 | -0.36 | 0.12 | -0.34 | -0.35 | 0.12 |
| *depr* (Quintile 1[b]) [a] | | | | | | |
| *depr* (Quintile 2) | 0.01 | 0.00 | 0.04 | 0.01 | 0.00 | 0.04 |
| *depr* (Quintile 3) | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| *depr* (Quintile 4) | 0.13 | 0.15 | 0.10 | 0.13 | 0.14 | 0.09 |
| *depr* (Quintile 5) | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.04 |
| *equi* | 0.10 | 0.10 | 0.09 | 0.09 | 0.10 | 0.09 |
| **Random effects** | | | | | | |
| $\hat{\sigma}_0$ (Intercept age) | 2.79 | 2.79 | 0.25 | 2.79 | 2.79 | 0.25 |
| $\hat{\sigma}_1$ (Slope of age) | 1.91 | 1.91 | 0.14 | 1.92 | 1.92 | 0.14 |

[a] indicates the reference categories.
[b] indicates this quintile is least deprived.

## 7. Discussion

In this paper, we propose a novel Bayesian group LASSO approach with spike-and-slab priors for variable selection in the context of QR. Two main aspects were considered: (i) variable selection within the QR model, and (ii) variable selection including both fixed and random effects within the QMM. The first aspect aimed to identify a more effective method for selecting variables in QR models with only fixed effects. Two novel Bayesian variable selection methods were proposed. These methods were developed by integrating Bayesian group LASSO methods with spike-and-slab priors on QR coefficients, similar to the work of Xu and Ghosh (2015). The first method, BGLSSQR, is based on the BGLASSO, which facilitates the selection of variables exhibiting group structures. The second method, BSGSSQR, extends the first, allowing for the selection of a group of predictors and

individual within a group using BSGLASSO. The performance of each method was assessed through simulation studies and compared with existing methods. The results indicate that these two proposed methods generally yield biased estimators, similar to other regularisation methods, whether frequentist or Bayesian. Notably, they were superior to both existing frequentist and Bayesian approaches and maintained their predictive performance and model selection accuracy in both simple and complex scenarios. This included situations with high sparsity, either independence or correlation among predictors, and a variety of variable groups and predictors. However, BSGSSQR offers an advantage over BGLSSQR as it also provides the ability to select individual levels within a group variable. Therefore, BSGSSQR was exclusively chosen for incorporation into the quantile mixed model as the proposed method for the second aspect mentioned earlier. This proposed method, BSGSSMQR, incorporates elements of BSGSSQR and introduces a decomposition for the covariance matrix of random effects, enabling the simultaneous selection of both fixed and random effects. Unlike the additive quantile mixed models (AQMM) (Geraci, 2019), BSGSSMQR estimates the smooth function through spline coefficients estimated in a penalised form. In AQMM, however, the smooth function is estimated by treating a set of spline coefficients as random variables, which are then predicted in a linear mixed model framework. In this aspect, a simulation study was also conducted to assess the proposed method in comparison to AQMM. Regarding the fixed effect selection, the proposed method demonstrates good performance compared with the bootstrap method implemented in AQMM. Moreover, it facilitates the simultaneous selection of random effects. Furthermore, we apply our approach to the analysis of longitudinal child growth data in Scotland, thus providing practical insights into its real-world applicability. The selected risk factors are similar between the 0.10th and 0.90th quantiles. Both random intercepts and random slopes were selected, implying variability in the individual linear trends.

One limitation of the BSGSSQR and BSGSSMQR approaches is the assumption that group structures are non-overlapping—for example, groups formed from multiple categorical variables or basis functions representing continuous covariates. However, in real-world applications such as child growth studies, risk factors often do not conform to this disjoint structure. Instead, some variables may belong to multiple conceptual domains. For instance, factors such as family income, parental education, and housing conditions may simultaneously relate to both economic and environmental domains. Further extensions of the current methodology could address this limitation by developing models capable of accommodating overlapping group structures. Another methodological limitation lies in the treatment of random effect selection in the BSGSSMQR framework. The penalty term associated with random effects is fixed at one, which may result in misspecification of the random effect structure and limit the model's capacity to fully capture the variability present in the real data. Future research could explore more flexible regularisation strategies. One potential direction is to extend the Bayesian group LASSO framework by incorporating truncated normal priors for random effects with group structures, similar to the strategy used for fixed effects selection. In addition, a notable challenge in implementing the proposed Bayesian approach using traditional MCMC methods is the substantial computational burden, which escalates with increasing sample size and the number of random effects. To address this, future work may consider more computationally efficient alternatives to MCMC. One promising solution is the Integrated Nested Laplace Approximation (INLA), which has been successfully applied to Bayesian QR models in recent studies (e.g. Yue and Rue (2011)). A final concern is that the proposed methods rely on independently estimating each conditional quantile using a likelihood-based approach via the AL distribution. As a result, they may violate the quantile monotonicity property, potentially leading to crossing quantile curves—especially at extreme quantiles. This issue is particularly important when constructing reference child growth curves for children. Future research should aim to address this limitation, for example by incorporating techniques that enforce quantile non-crossing.

## 8. Conclusions

This study introduced a novel Bayesian variable selection framework for quantile mixed models, facilitating the identification of both fixed and random effects associated with distributional outcomes in longitudinal data. Through comprehensive simulation studies, the proposed approach demonstrated superior performance over existing methods in both variable selection and predictive accuracy. Its application to real child growth data further highlighted its practical utility by identifying critical risk factors associated with weight extremes. By accommodating both categorical and continuous predictors, as well as selecting relevant random effects, the method provides a powerful and flexible tool for uncovering important determinants of longitudinal outcomes—such as raw weight—in the study of child growth and development. Future research could extend this framework to more complex longitudinal settings, including non-linear random effects, overlapping group structures, covariate interactions, and high-dimensional or multivariate outcome data, while also exploring computational strategies to further improve scalability for large cohort studies.

**Acknowledgements**

The authors gratefully acknowledge the use of data from the Growing Up in Scotland (GUS) study, accessed via the UK Data Service. The data were originally collected by Scottish Centre for Social Research. The UK Data Service and the original data creators bear no responsibility for the analysis or interpretations presented in this paper. We would like to thank the referees for their comments and suggestions on the manuscript.

## References

Alhamzawi R. Brq: Bayesian Analysis of Quantile Regression Models; 2020, https://CRAN.R-project.org/package=Brq, R package version 3.0.

Alhamzawi R, Ali HTM. The Bayesian adaptive lasso regression. Math Biosci. 2018; 303: 75–82.

Alhamzawi R, Ali HTM. Brq: an R package for Bayesian quantile regression. Metron. 2020; 78(3): 313–328.

Alhamzawi R, Yu K. Conjugate priors and variable selection for Bayesian quantile regression. Comput Stat Data Anal. 2013; 64: 209–219.

Alhamzawi R, Yu K. Bayesian Lasso-mixed quantile regression. J Stat Comput Simul. 2014; 84(4): 868–880.

Alhamzawi R. Prior elicitation and variable selection for bayesian quantile regression. PhD [dissertation] . London, Brunel University;2013.

Andrews DF, Mallows CL. Scale mixtures of normal distributions. J R Stat Soc Ser B Stat Methodol. 1974; 36(1): 99–102.

Aniley TT, Debusho LK, Nigusie ZM, Yimer WK, Yimer BB. A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients. BMC Med Res Methodol. 2019; 19(1): 13.

Berk LE. Child development. Boston: Pearson; 2013.

Berk LE, Meyers AB. Infants, children, and adolescents. Boston: Pearson; 2016.

Birtwistle S, Deakin E, Whitford R, Hinchliffe S, Daniels-Creasey A, Rule S, The Scottish Health Survey: 2022 edition; [ISBN: 9781835216569]. 2023. [cited 2023 Nov 29]. Available from: https://www.gov.scot/publications/scottish-health-survey-2022-volume-1-main-report/pages/2/.

Black RE, Allen LH, Bhutta ZA, et al. Maternal and child undernutrition: global and regional exposures and health consequences. Lancet. 2008; 371(9608): 243–260.

Bradshaw P, Tipping S, Marryat L, Corbett J. Growing up in Scotland Sweep 1 – 2005: User guide. Edinburgh, UK: Scottish Centre for Social Research; 2007.

Bukatko D, Daehler MW. Child development: a thematic approach. Boston: Houghton Mifflin; 1995.

Casella G. Empirical bayes gibbs sampling. Biostatistics. 2001; 2(4): 485–500.

Chen CWS, Dunson DB, Reed C, Yu K. Bayesian variable selection in quantile regression. Stat Its Interface. 2013; 6(2): 261–274.

Chen Z, Dunson DB. Random effects selection in linear mixed models. Biometrics. 2003; 59(4): 762–769.
Cole TJ. Growth charts for both cross-sectional and longitudinal data. Stat Med. 1994; 13(23-24): 2477–2492.
Cole TJ, Donaldson MDC, Ben-Shlomo Y. SITAR—a useful instrument for growth curve analysis. Int J Epidemiol. 2010; 39(6): 1558–1566.
Denham SA. Social-emotional competence as support for school readiness: what is it and how do we assess it? EE&D. 2006; 17(1): 57–89.
Durbán M, Harezlak J, Wand MP, Carroll RJ. Simple fitting of subject-specific curves for longitudinal data. Stat Med. 2005; 24(8): 1153–1167.
Fitzmaurice GM, Ravichandran C. A primer in longitudinal data analysis. Circ. 2008; 118(19): 2005–2010.
Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on article by browne and draper). Bayesian Anal. 2006; 1(3): 515–534.
Gelman A. Bayesian data analysis. Boca Raton: CRC Press; 2014.
Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci. 1992; 7(4).
George EI, McCulloch RE. Approaches for bayesian variable selection. Stat Sin. 1997; 7(2): 339–373.
Geraci M. Additive quantile regression for clustered data with an application to children's physical activity. J R Stat Soc Ser C Appl Stat. 2019; 68(4): 1071–1089.
Geraci M, Bottai M. Linear quantile mixed models. Stat Comput. 2014; 24(3): 461–479.
Grammer JK, Coffman JL, Ornstein PA, Morrison FJ. Change over time: conducting longitudinal studies of children's cognitive development. J Cogn Dev. 2013; 14(4): 515–528.
Ishwaran H, Rao JS. Spike and slab variable selection: frequentist and bayesian strategies. Ann Stat. 2005; 33(2): 730–773.
Ji Y, Shi H. Shrinkage estimation of fixed and random effects in linear quantile mixed models. J Appl Stat. 2022; 49(14): 3693–3716.
Kinney SK, Dunson DB. Fixed and random effects selection in linear and logistic models. Biometrics. 2007; 63(3): 690–698.
Koenker R. Quantile regression. Cambridge: Cambridge University Press; 2005.
Koenker R. quantreg: Quantile Regression; 2021, https://CRAN.R-project.org/package=quantreg, R package version 5.86.
Koenker R, Machado JAF. Goodness of Fit and Related Inference Processes for Quantile Regression. J Am Stat Assoc. 1999; 94(448): 1296–1310.
Kozumi H, Kobayashi G. Gibbs sampling methods for Bayesian quantile regression. J Stat Comput Simul. 2011; 81(11): 1565–1578.
Kyung M, Gill J, Ghosh M, Casella G. Penalized regression, standard errors, and Bayesian lassos. Bayesian Anal. 2010; 5(2): 369–411.
Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38(4): 963.
Li Q, Lin N, Xi R. Bayesian regularized quantile regression. Bayesian Anal. 2010; 5(3): 533–556.
Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. J Am Stat Assoc. 1988; 83(404): 1023–1032.
Muggeo VM, Atkins DC, Gallop RJ, Dimidjian S. Segmented mixed models with random change-points: a maximum likelihood approach with application to treatment for depression study. Stat Model. 2014; 14(4): 293–313.
Park T, Casella G. The Bayesian lasso. J Am Stat Assoc. 2008; 103(482): 681–686.
R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria; 2024, https://www.R-project.org/.
Royal College of Physicians and Surgeons of Glasgow, Obesity Action Scotland | Providing leadership and advocacy on preventing & reducing obesity & overweight in Scotland | 5 years on, are we on track to halve childhood obesity in Scotland by 2030?; 2023. [cited 2023 Nov 29]. Available from: https://www.obesityactionscotland.org/blogs/5-years-on-are-we-on-track-to-halve-childhood-obesity-in-scotland-by-2030.

Ruppert D, Wand MP, Carroll RJ. Semiparametric regression. Cambridge: Cambridge University Press; 2003.
Sabates R, Dex S. The impact of multiple risk factors on young children's cognitive and behavioural development. Child Soc. 2015; 29(2): 95–108.
Sherwood B, Maidman A, Li S. rqPen: Penalized Quantile Regression; 2023, https://CRAN.R-project.org/package=rqPen, R package version 3.1.3.
Shonkoff JP, Phillips DA, National Research Council (U S ), editors. From neurons to neighborhoods: the science of early child development. Washington (DC): National Academy Press; 2000.
Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. J Comput Graph Stat. 2013; 22(2): 231–245.
Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q. 2005; 83(3): 457–502.
Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol. 1996; 58(1): 267–288.
Tsionas EG. Bayesian quantile inference. J Stat Comput Simul. 2003; 73(9): 659–674.
Weiss ST, Ware JH. Overview of issues in the longitudinal analysis of respiratory data. Am J Respir Crit Care Med. 1996; 154(6_pt_2): S208–S211.
Wood SN. Generalized additive models: an introduction with R. Boca Raton: Chapman & Hall/CRC; 2006.
Xu X, Ghosh M. Bayesian variable selection and estimation for group lasso. Bayesian Anal. 2015; 10(4): 909–936.
Yu K, Moyeed RA. Bayesian quantile regression. Stat Probab Lett. 2001; 54(4): 437–447.
Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B Stat Methodol. 2006; 68(1): 49–67.
Yue YR, Rue H. Bayesian inference for additive mixed quantile regression models. Comput Stat Data Anal. 2011; 55(1): 84–96.
Zhang L, Baladandayuthapani V, Mallick BK, et al. Bayesian hierarchical structured variable selection methods with application to molecular inversion probe studies in breast cancer. J R Stat Soc Ser C Appl Stat. 2014; 63(4): 595–620.

## Appendix I

**Illustrative prior densities of the hypervariance $(\gamma^*)$ under the spike-and-slab prior specification in (11)**

## Appendix II

**Gibbs sampler for BGLSSQR**

The full conditional distribution of $\boldsymbol{\beta}_g$ follows a multivariate spike-and-slab distribution, given by

$$\boldsymbol{\beta}_g|\text{rest} \sim (1-q_g)\mathcal{N}_{m_g}(\boldsymbol{\mu}_g, 2\sigma\boldsymbol{\Sigma}_g) + q_g\delta_0(\boldsymbol{\beta}_g), \quad g = 1, \ldots, G,$$

where

$$\boldsymbol{\mu}_g = \boldsymbol{\Sigma}_g\mathbf{X}_g'\mathbf{V}(\mathbf{y} - \mathbf{X}_{(g)}\boldsymbol{\beta}_{(g)} - \xi\boldsymbol{v}) \text{ and } \boldsymbol{\Sigma}_g = \left(\mathbf{X}_g'\mathbf{V}\mathbf{X}_g + \frac{1}{\eta_g^2}\mathbf{I}_{m_g}\right)^{-1}.$$

Here, $q_g$ is the probability which controls the spike-and-slab of the distribution, computed as:

$$\begin{aligned} q_g &= P(\boldsymbol{\beta}_g = \mathbf{0}|\text{rest}) \\ &= \frac{\pi_0}{\pi_0 + (1-\pi_0)(\eta_g^2)^{-\frac{m_g}{2}}|\boldsymbol{\Sigma}_g|^{\frac{1}{2}}\exp\left\{\frac{1}{4\sigma}||\boldsymbol{\Sigma}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}(\mathbf{y} - \mathbf{X}_{(g)}\boldsymbol{\beta}_{(g)} - \xi\boldsymbol{v})||_2^2\right\}}. \end{aligned}$$

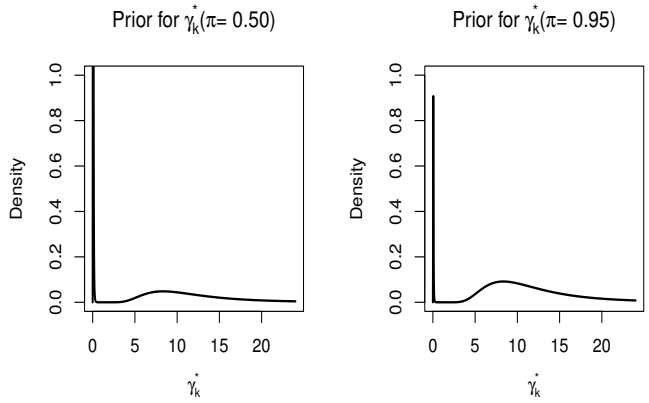



**Figure 4** Densities of the hypervariance $\gamma^* = \gamma_k \varsigma_k^2$ with $\nu_0 = 0.005$, $a = 5$, and $b = 50$, for two values of $\pi$: $\pi = 0.50$ (left) and $\pi = 0.50$ (right)

It is worth noting that the term $\mathbf{y} - \mathbf{X}_{(g)}\boldsymbol{\beta}_{(g)} - \xi\boldsymbol{v}$ represents the residual vector obtained by excluding the $g$th factor $\boldsymbol{\beta}_g$ in our quantile regression model. As a result, the elements in $\mathbf{X}_g'(\mathbf{y} - \mathbf{X}_{(g)}\boldsymbol{\beta}_{(g)} - \xi\boldsymbol{v})$ are proportional to the correlation between each predictor in the $g$th group and this residual vector.

Rather than directly defining the full conditional of $\eta_g^2$, an alternative approach is adopted by introducing the parameter $\alpha_g^2 = 1/\eta_g^2$. The full conditional of $\alpha_g^2$ is subsequently defined under two distinct conditions. In the case when $\boldsymbol{\beta}_g = 0$, the full conditional distribution of $\alpha_g^2$ is

$$\alpha_g^2|\text{rest} \sim \text{InvGamma}\left(s_1 = \frac{m_g + 1}{2}, s_2 = \frac{\lambda^2}{2}\right),$$

where $s_1$ and $s_2$ are the shape and scale parameters of an inverse Gamma distribution (InvGamma), respectively. On the other hand, when $\boldsymbol{\beta}_g \neq 0$,

$$\alpha_g^2|\text{rest} \sim \text{InvGaussian}\left(\mu' = \frac{\sqrt{\lambda^2\sigma}}{||\beta_g||_2}, \lambda' = \lambda^2\right).$$

The full conditional distribution of each $v_i$ is then an inverse Gaussian distribution (InvGaussian),

$$v_i|\text{rest} \sim \text{InvGaussian}\left(\mu' = \frac{1}{|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|}, \lambda' = \frac{1}{2\sigma}\right).$$

The full conditional distribution of $\sigma$ is

$$\sigma|\text{rest} \sim \text{InvGamma}\left(\frac{3n}{2} + \frac{k}{2} + g_1, \frac{(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}(\tilde{\mathbf{y}} - \mathbf{X}\boldsymbol{\beta})}{2} + \tau(1-\tau)\sum_{i=1}^{n} v_i + \boldsymbol{\beta}'\mathbf{S}\boldsymbol{\beta} + g_2\right),$$

where $\tilde{\mathbf{y}} = \mathbf{y} - \xi\boldsymbol{v}$, $k = \sum_{g=1}^{G} m_g \mathbf{1}_{\{\boldsymbol{\beta}_g \neq 0\}}$, and $\mathbf{S} = \text{diag}(\eta_1^{-2}, \ldots, \eta_G^{-2})$.

The full conditional distribution of $\pi_0$ is a Beta distribution,

$$\pi_0|\text{rest} \sim \text{Beta}\left(a_1 + \sum_{g=1}^{G} \mathbf{1}_{\{\boldsymbol{\beta}_g \neq 0\}}, a_2 + \sum_{g=1}^{G} m_g - \sum_{g=1}^{G} \mathbf{1}_{\{\boldsymbol{\beta}_g \neq 0\}}\right).$$

**Gibbs sampler for BSGLSSQR**

The full conditional distribution of $\boldsymbol{b}_g$ is then a multivariate spike-and- slab distribution

$$\boldsymbol{b}_g|\text{rest} \sim (1-q_g)\mathcal{N}_{m_g}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + q_g\delta_0(\boldsymbol{b}_g),$$

where

$$\boldsymbol{\mu}_g = \left(\frac{1}{2\sigma}\boldsymbol{\Sigma}_g^{\frac{1}{2}}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}(\mathbf{y}-\mathbf{X}_{(g)}\mathbf{K}_{(g)}^{\frac{1}{2}}\boldsymbol{b}_{(g)}^{\frac{1}{2}} - \xi\boldsymbol{v})\right),$$

$$\boldsymbol{\Sigma}_g = \left(\frac{1}{2\sigma}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}\mathbf{X}_g\mathbf{K}_g^{\frac{1}{2}} + \mathbf{I}_{m_g}\right)^{-1},$$

and $q_g$ be the posterior probability of $\boldsymbol{b}_g$ being zero given the remaining parameters and can be computed by

$$\begin{aligned} q_g &= P(\boldsymbol{b}_g = \mathbf{0}|\text{rest}) \\ &= \frac{\pi_0}{\pi_0 + (1-\pi_0)|\boldsymbol{\Sigma}|^{\frac{1}{2}}\exp\left\{\frac{1}{4\sigma^2}||\boldsymbol{\Sigma}_g^{\frac{1}{2}}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}(\mathbf{y}-\mathbf{X}_{(g)}\mathbf{K}_{(g)}^{\frac{1}{2}}\boldsymbol{b}_{(g)}^{\frac{1}{2}} - \xi\boldsymbol{v})||_2^2\right\}}. \end{aligned}$$

The full conditional distribution of $\theta_{gj}$ is then a spike-and-slab mixture, where the slab component is taken to be a normal distribution truncated to the positive value,

$$\theta_{gj}|\text{rest} \sim (1-r_{gj})N^+(u_{gj}, v_{gj}^2) + r_{gj}\delta_0(\theta_{gj}),$$

where

$$u_{gj} = \frac{1}{2\sigma}v_{gj}^2(\mathbf{y}-\mathbf{X}_{(gj)}\boldsymbol{\beta}_{(gj)} - \xi\boldsymbol{v})'\mathbf{V}\mathbf{X}_{gj}b_{gj},$$

$$v_{gj}^2 = \left(\frac{1}{s^2} + \frac{1}{2\sigma}\mathbf{X}_{gj}'\mathbf{V}\mathbf{X}_{gj}b_{gj}^2\right)^{-1},$$

and

$$r_{gj} = p(\theta_{gj}=0|\text{rest}) = \frac{\pi_1}{\pi_1 + 2(1-\pi_1)(s^2)^{-\frac{1}{2}}(v_{gj}^2)^{\frac{1}{2}}\exp\left\{\frac{u_{gj}^2}{2v_{gj}^2}\right\}\left[\Phi(\frac{u_{gj}}{v_{gj}})\right]}.$$

The full conditional distribution of each $v_i$ is

$$v_i|\text{rest} \sim \text{InvGaussian}\left(\mu' = \frac{1}{|y_i - \boldsymbol{x}_i'\boldsymbol{\beta}|}, \lambda' = \frac{1}{2\sigma}\right).$$

The full conditional distribution of $\sigma$ is given by

$$\sigma|\text{rest} \sim \text{InvGamma}\left(\frac{3n}{2} + g_1, \frac{1}{4}(\boldsymbol{\epsilon}-\xi\boldsymbol{v})'\mathbf{V}(\boldsymbol{\epsilon}-\xi\boldsymbol{v}) + \tau(1-\tau)\sum_{i=1}^{n}v_i + g_2\right),$$

where $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$.

The full conditional distribution of $\pi_0$ and $\pi_1$ are given by

$$\begin{aligned} \pi_0|\text{rest} &\sim \text{Beta}(\#(\boldsymbol{b}_g = 0) + a_1, \#(\boldsymbol{b}_g \neq 0) + a_2), \\ \pi_1|\text{rest} &\sim \text{Beta}(\#(\theta_{gj} = 0) + c_1, \#(\theta_{gj} \neq 0) + c_2). \end{aligned}$$

Note that $\#(\cdot)$ denotes the cardinality or count of the set of elements for which the condition in $(\cdot)$ holds true.

The full conditional distribution of $s^2$ is given by

$$s^2|\text{rest} \sim \text{InvGamma}\left(1 + \frac{1}{2}\#(\theta_{gj}=0), t + \frac{1}{2}\sum_{g,j}\theta_{gj}^2\right).$$

**Gibbs sampler for BSGSSMQR**

Let $i = 1, \ldots, N$ index individuals and $j = 1, \ldots, n_i$ index repeated observations for individual $i$. Define

$$\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})', \qquad \mathbf{y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{N1}, \ldots, y_{Nn_N})',$$
$$\boldsymbol{v}_i = (v_{i1}, \ldots, v_{in_i})', \qquad \boldsymbol{v} = (v_{11}, \ldots, v_{1n_1}, \ldots, v_{N1}, \ldots, v_{Nn_N})',$$

and

$$\mathbf{V} = \mathrm{diag}(\boldsymbol{v}^{-1}),$$

where $\boldsymbol{v}^{-1}$ denotes the elementwise inverse of $\boldsymbol{v}$. For the random-effects part, let

$$R_{ij} = \mathbf{z}_{ij}'\mathbf{D}\mathbf{A}\boldsymbol{c}_i, \qquad \mathbf{R} = (R_{11}, \ldots, R_{1n_1}, \ldots, R_{N1}, \ldots, R_{Nn_N})',$$

and define

$$\boldsymbol{r}_{1ij} = \big(c_{i\ell}d_{m'}z_{ijm'} : \ell = 1, \ldots, q,\ m' = \ell + 1, \ldots, q\big)',$$
$$\boldsymbol{r}_{2ij} = \Big(z_{ij\ell}\big(c_{i\ell} + \sum_{m'=1}^{\ell-1} c_{im'}a_{m'\ell}\big) : \ell = 1, \ldots, q\Big)',$$

with $n = \sum_{i=1}^{N} n_i$.

The full conditional distribution of $\boldsymbol{b}_g$ is then a multivariate spike-and- slab distribution

$$\boldsymbol{b}_g|\text{rest} \sim (1 - q_g)\mathcal{N}_{m_g}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) + q_g\delta_0(\boldsymbol{b}_g),$$

where

$$\boldsymbol{\mu}_g = \left(\frac{1}{2\sigma}\boldsymbol{\Sigma}_g^{\frac{1}{2}}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}(\mathbf{y} - \mathbf{X}_{(g)}\mathbf{K}_{(g)}^{\frac{1}{2}}\boldsymbol{b}_{(g)}^{\frac{1}{2}} - \mathbf{R} - \xi\boldsymbol{v})\right),$$
$$\boldsymbol{\Sigma}_g = \left(\frac{1}{2\sigma}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}\mathbf{X}_g\mathbf{K}_g^{\frac{1}{2}} + \mathbf{I}_{m_g}\right)^{-1},$$

and $q_g$ be the posterior probability of $\boldsymbol{b}_g$ being zero given the remaining parameters and can be computed by

$$q_g = P(\boldsymbol{b}_g = \mathbf{0}|\text{rest}) = \frac{\pi_0}{\pi_0 + (1 - \pi_0)|\boldsymbol{\Sigma}_g|^{\frac{1}{2}}\exp\left\{\frac{1}{4\sigma^2}||\boldsymbol{\Sigma}_g^{\frac{1}{2}}\mathbf{K}_g^{\frac{1}{2}}\mathbf{X}_g'\mathbf{V}(\mathbf{y} - \mathbf{X}_{(g)}\mathbf{K}_{(g)}^{\frac{1}{2}}\boldsymbol{b}_{(g)}^{\frac{1}{2}} - \mathbf{R} - \xi\boldsymbol{v})||_2^2\right\}}.$$

The full conditional distribution of $\theta_{gj}$ is then a spike-and-slab mixture, where the slab component is taken to be a normal distribution truncated to the positive value,

$$\theta_{gj}|\text{rest} \sim (1 - r_{gj})N^+(u_{gj}, v_{gj}^2) + r_{gj}\delta_0(\theta_{gj}),$$

where

$$u_{gj} = \frac{1}{2\sigma}v_{gj}^2(\mathbf{y} - \mathbf{X}_{(gj)}\boldsymbol{\beta}_{(gj)} - \mathbf{R} - \xi\boldsymbol{v})'\mathbf{V}\mathbf{X}_{gj}b_{gj},$$
$$v_{gj}^2 = \left(\frac{1}{s^2} + \frac{1}{2\sigma}\mathbf{X}_{gj}'\mathbf{V}\mathbf{X}_{gj}b_{gj}^2\right)^{-1},$$

and

$$r_{gj} = p(\theta_{gj} = 0|\text{rest}) = \frac{\pi_1}{\pi_1 + 2(1 - \pi_1)(s^2)^{-\frac{1}{2}}(v_{gj}^2)^{\frac{1}{2}}\exp\left\{\frac{u_{gj}^2}{2v_{gj}^2}\right\}\left[\Phi(\frac{u_{gj}}{v_{gj}})\right]}.$$

The full conditional distribution of $\boldsymbol{c}_i$ is a multivariate normal distribution

$$\boldsymbol{c}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{c}_i}, \boldsymbol{\Sigma}_{\boldsymbol{c}_i}),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{c}_i} = \frac{\boldsymbol{\Sigma}_{\boldsymbol{c}_i}\mathbf{A}'\mathbf{D}\mathbf{Z}_i'\mathbf{V}_i}{2\sigma}\Big(\mathbf{y}_i - \mathbf{x}_i'\boldsymbol{\beta} - \xi\boldsymbol{v}_i\Big) \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{c}_i} = \left(\frac{\mathbf{A}'\mathbf{D}\mathbf{Z}_i'\mathbf{V}_i\mathbf{Z}_i\mathbf{D}\mathbf{A}}{2\sigma} + \boldsymbol{\Omega}^{-1}\right)^{-1}.$$

The full conditional distribution of $\boldsymbol{a}$ is a multivariate normal distribution

$$\boldsymbol{a} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{a}}, \boldsymbol{\Sigma}_{\boldsymbol{a}}) \cdot 1(\boldsymbol{a} \in \mathcal{R}_{\boldsymbol{d}}),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{a}} = \boldsymbol{\Sigma}_{\boldsymbol{a}}\left(\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\boldsymbol{r}_{1ij}(y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta} - \xi v_{ij})}{2\sigma v_{ij}}\right) \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{a}} = \left(\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{\boldsymbol{r}_{1ij}\boldsymbol{r}_{1ij}'}{2\sigma v_{ij}} + \mathbf{A}_0^{-1}\right)^{-1}.$$

The full conditional distribution of each $d_l (l = 1, \ldots, q)$ is a zero-inflated half-normal distribution

$$d_l \sim \text{ZI} - N^+(\hat{p}_l, \mu_{d_l}, \sigma_{d_l}^2),$$

where

$$\mu_{d_l} = \sigma_{d_l}^2\left(\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{r_{2ij}(y_{ij} - \mathbf{x}_{ij}'\boldsymbol{\beta} - \sum_{s\neq l'} r_{2ijs}d_s - \xi v_{ij})}{2\sigma v_{ij}}\right),$$

$$\sigma_{d_l}^2 = \left(\sum_{i=1}^{N}\sum_{j=1}^{n_i}\frac{r_{2ij}^2}{2\sigma v_{ij}} + 1\right)^{-1},$$

$$\hat{p}_l = \left(1 + \frac{(1-p_{l0})N(0;0,1)(1-\Phi(0;\mu_{d_l},\sigma_{d_l}^2))}{p_{l0}N(0;\mu_{d_l},\sigma_{d_l}^2)(1-\Phi(0;0,1)}\right)^{-1},$$

and $\Phi(\cdot)$ is the cumulative normal distribution function.

The full conditional distribution of each $v_i$ is

$$v_i|\text{rest} \sim \text{InvGaussian}\left(\mu' = \sqrt{\frac{1}{(y_i - \boldsymbol{x}_{ij}'\beta - \boldsymbol{z}_{ij}'\mathbf{D}\mathbf{A}\boldsymbol{c}_i)^2}}, \lambda' = \frac{1}{2\sigma}\right).$$

The full conditional distribution of $\sigma$ is

$$\sigma|\text{rest} \sim \text{InvGamma}\left(\frac{3n}{2} + g_1, \frac{1}{4}(\boldsymbol{\epsilon} - \xi\boldsymbol{v})'\mathbf{V}(\boldsymbol{\epsilon} - \xi\boldsymbol{v}) + \tau(1-\tau)\sum_{i=1}^{n} v_i + g_2\right),$$

where $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{R}$.

The full conditional distribution of each $\omega_l (l = 1, \ldots, q)$ is then a generalised inverse Gaussian distribution (GIG),

$$\omega_l \sim \text{GIG}(\phi, \kappa, \psi),$$

where $\psi = -(N+2)/2, \kappa^2 = \sum_{i=1}^{N} c_{il}^2$, and $\psi^2 = N$.

The full conditional distribution of $\pi_0$ and $\pi_1$ is given by

$$\pi_0|\text{rest} \sim \text{Beta}(\#(\boldsymbol{b}_g = 0) + a_1, \#(\boldsymbol{b}_g \neq 0) + a_2),$$
$$\pi_1|\text{rest} \sim \text{Beta}(\#(\theta_{gj} = 0) + c_1, \#(\theta_{gj} \neq 0) + c_2).$$

The full conditional distribution of $s^2$ is given by

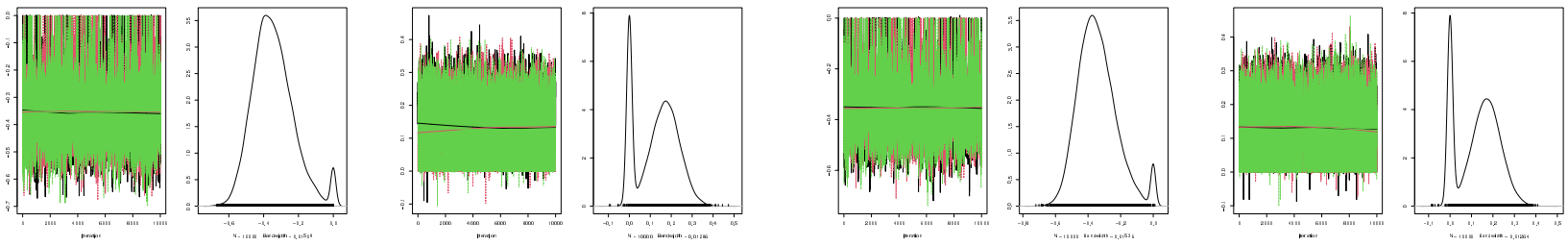
$$s^2|\text{rest} \sim \text{InvGamma}\left(1 + \frac{1}{2}\#(\theta_{gj} = 0), t + \frac{1}{2}\sum_{g,j}\theta_{gj}^2\right).$$

## Appendix III

## Additional results

**Table 10** Posterior mean, posterior median for both the fixed (FEs) and random (REs) effects, and computational times (in minutes) of three quantile models under the scenario $\sigma_\epsilon^2 = 9$

| | | $\tau = 0.10$ | | | $\tau = 0.50$ | | | $\tau = 0.90$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **True** | **Mean** | **Median** | **SD** | **Mean** | **Median** | **SD** | **Mean** | **Median** | **SD** |
| **FEs** | | | | | | | | | | |
| $\beta_1$ | 0.67 | 0.45 | **0.39** | 0.50 | 0.44 | **0.37** | 0.50 | 0.43 | **0.35** | 0.51 |
| $\beta_2$ | -1.00 | -1.08 | **-1.08** | 0.25 | -1.08 | **-1.08** | 0.25 | -1.08 | **-1.08** | 0.25 |
| $\beta_3$ | 0.33 | 0.16 | 0.00 | 0.29 | 0.16 | 0.00 | 0.29 | 0.17 | 0.01 | 0.30 |
| $\beta_4$ | 0.50 | 0.61 | **0.62** | 0.24 | 0.61 | **0.62** | 0.24 | 0.61 | **0.62** | 0.24 |
| $\beta_5$ | 0.30 | 0.29 | **0.29** | 0.22 | 0.29 | **0.29** | 0.23 | 0.28 | **0.28** | 0.23 |
| $\beta_6$ | 0.00 | -0.10 | 0.00 | 0.17 | -0.10 | 0.00 | 0.17 | -0.10 | 0.00 | 0.17 |
| $\beta_7$ | 0.00 | -0.12 | 0.00 | 0.17 | -0.11 | 0.00 | 0.17 | -0.11 | 0.00 | 0.17 |
| $\beta_8$ | 0.00 | -0.07 | 0.00 | 0.15 | -0.08 | 0.00 | 0.15 | -0.08 | 0.00 | 0.15 |
| $\beta_9$ | 0.00 | 0.04 | 0.00 | 0.13 | 0.03 | 0.00 | 0.14 | 0.03 | 0.00 | 0.14 |
| $\beta_{10}$ | 2.00 | 2.36 | **2.39** | 0.41 | 2.37 | **2.40** | 0.42 | 2.38 | **2.42** | 0.42 |
| $\beta_{11}$ | 1.00 | 1.03 | **1.04** | 0.30 | 1.04 | **1.05** | 0.30 | 1.04 | **1.06** | 0.30 |
| $\beta_{12}$ | 0.00 | -0.09 | 0.00 | 0.21 | -0.09 | 0.00 | 0.21 | -0.10 | 0.00 | 0.21 |
| $\beta_{13}$ | 0.00 | 0.15 | 0.00 | 0.21 | 0.15 | 0.00 | 0.21 | 0.15 | 0.00 | 0.21 |
| $\beta_{14}$ | 0.00 | -0.07 | 0.00 | 0.19 | -0.07 | 0.00 | 0.19 | -0.07 | 0.00 | 0.20 |
| $\beta_{15}$ | 0.00 | -0.01 | 0.00 | 0.12 | -0.01 | 0.00 | 0.12 | -0.01 | 0.00 | 0.12 |
| $\beta_{16}$ | 0.00 | 0.03 | 0.00 | 0.13 | 0.03 | 0.00 | 0.13 | 0.04 | 0.00 | 0.13 |
| $\beta_{17}$ | 0.00 | -0.13 | 0.00 | 0.23 | -0.12 | 0.00 | 0.22 | -0.13 | 0.00 | 0.23 |
| $\beta_{18}$ | 0.00 | 0.07 | 0.00 | 0.16 | 0.07 | 0.00 | 0.16 | 0.07 | 0.00 | 0.16 |
| $\beta_{19}$ | 0.00 | 0.06 | 0.00 | 0.18 | -0.05 | 0.00 | 0.18 | 0.05 | 0.00 | 0.18 |
| $\beta_{20}$ | 0.00 | -0.04 | 0.00 | 0.15 | -0.05 | 0.00 | 0.15 | -0.04 | 0.00 | 0.15 |
| $\beta_{21}$ | 0.00 | -0.16 | 0.00 | 0.20 | -0.16 | 0.00 | 0.20 | -0.16 | 0.00 | 0.21 |
| $\beta_{22}$ | 0.00 | -0.05 | 0.00 | 0.13 | -0.05 | 0.00 | 0.13 | -0.05 | 0.00 | 0.13 |
| $\beta_{23}$ | 0.00 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 | 0.11 |
| $\beta_{24}$ | 0.00 | 0.03 | 0.00 | 0.12 | 0.03 | 0.00 | 0.12 | 0.03 | 0.00 | 0.12 |
| $\beta_{25}$ | 0.00 | -0.04 | 0.00 | 0.13 | -0.04 | 0.00 | 0.13 | -0.04 | 0.00 | 0.13 |
| $\beta_{26}$ | 0.00 | 0.05 | 0.00 | 0.13 | 0.05 | 0.00 | 0.13 | 0.05 | 0.00 | 0.13 |
| $\beta_{27}$ | 0.00 | -0.11 | 0.00 | 0.16 | -0.11 | 0.00 | 0.16 | -0.11 | 0.00 | 0.16 |
| **REs** | | | | | | | | | | |
| $\sigma_{11}$ | 0.90 | 1.13 | **1.08** | 0.35 | 1.10 | **1.06** | 0.35 | 1.10 | **1.06** | 0.35 |
| $\sigma_{22}$ | 0.50 | 0.53 | **0.52** | 0.24 | 0.54 | **0.52** | 0.25 | 0.55 | **0.54** | 0.25 |
| $\sigma_{33}$ | 0.20 | 0.44 | **0.41** | 0.23 | 0.47 | **0.43** | 0.26 | 0.47 | **0.42** | 0.25 |
| $\sigma_{44}$ | 0.00 | 0.19 | 0.13 | 0.20 | 0.21 | 0.14 | 0.21 | 0.20 | 0.14 | 0.20 |
| **Time** | | | 114.00 | | | 128.00 | | | 120.00 | |



(a) Low birth weight (Yes)



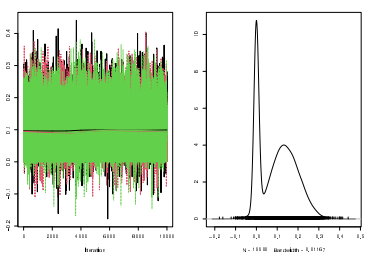
(b) Deprivation quintile (4)



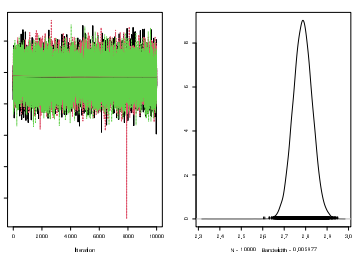
(f) Low birth weight (Yes)



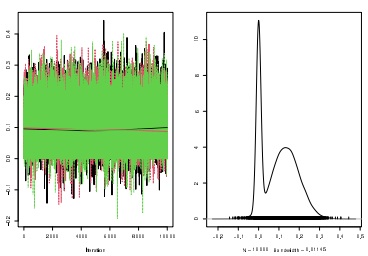
(g) Deprivation quintile (4)



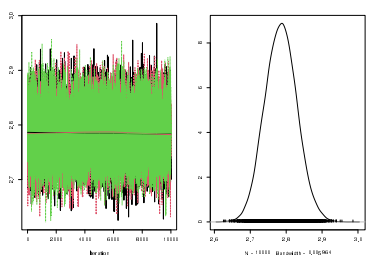
(c) Equivalised income



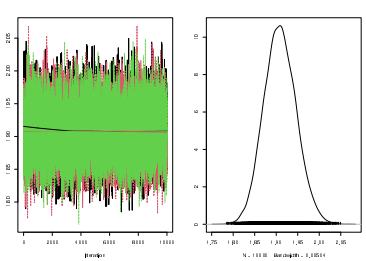
(d) SD of random intercepts



(h) Equivalised income



(i) SD of random intercepts



(e) SD of random slopes

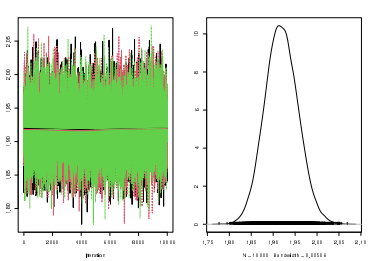
(a) $\hat{R}_{GR} = 1.00, n_{ESS} = 7970.29$

(b) $\hat{R}_{GR} = 1.00, n_{ESS} = 5163.04$

(c) $\hat{R}_{GR} = 1.00, n_{ESS} = 7408.98$

(d) $\hat{R}_{GR} = 1.00, n_{ESS} = 19135.60$

(e) $\hat{R}_{GR} = 1.00, n_{ESS} = 6144.25$



(j) SD of random slopes

(f) $\hat{R}_{GR} = 1.00, n_{ESS} = 6474.53$

(g) $\hat{R}_{GR} = 1.00, n_{ESS} = 5342.68$

(h) $\hat{R}_{GR} = 1.00, n_{ESS} = 7814.62$

(i) $\hat{R}_{GR} = 1.00, n_{ESS} = 23326.68$

(j) $\hat{R}_{GR} = 1.00, n_{ESS} = 5404.25$

**Figure 5** Trace plots of MCMC samples for each selected linear predictor and random effect from the 0.10th quantile model ((a)-(e)) and the 0.90th quantile model ((f)-(j)) for the centred raw weight, alongside the Gelman-Rubin diagnostic ($\hat{R}_{GR}$) and the effective sample size ($n_{ESS}$)