

An Expectation-Maximization algorithm for noncrossing Bayesian quantile regression analysis under informative sampling

Marcus L. Nascimento^{*a,b} and Kelly C. M. Gonçalves^c

^a*Escola de Matemática Aplicada, Fundação Getulio Vargas, Rio de Janeiro, Brazil*

^b*Departamento de Pesquisa, Fundação José Luiz Egydio Setúbal, São Paulo, Brazil*

^c*Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil*

Abstract

When quantiles are fitted separately, the resultant regression lines may cross, violating the basic probabilistic rule that quantiles are monotonic functions and possibly causing problems for inference and interpretation in practice. This article introduces a method for handling crossing issues regarding the analysis of complex survey data under informative sampling. Using the location-scale mixture representation of the asymmetric Laplace distribution, we write a joint posterior density function for the quantile levels of interest and develop a constrained Expectation-Maximization algorithm. A model-based simulation study is proposed, and data from the Brazilian National Demographic and Health Survey of Women and Children is analyzed to verify and illustrate the algorithm's effectiveness.

Keywords: Complex survey data, maximum a posteriori, monotonicity, child growth standards.

*Marcus L. Nascimento can be contacted via e-mail: marcus.lavagnole@fgv.br

1 Introduction

Frequently, data analysis considers the relation among random variables through the mean. However, the interest in exploring this relation in quantiles might be more informative in a myriad of applications as quantile regression describes it more comprehensively. Austin et al. (2005) demonstrated that richer inferences can be drawn regarding gender dissimilarities in the delivery of thrombolysis in patients with an acute myocardial infarction. Gilpin (2012) found discrepancies between mean and quantile regression results when investigating the association between salaries and scholastic aptitude for full-time public high school teachers. Beyond its applicability in medicine and education, quantile regression is also employed in the environmental sciences (Villarini et al., 2011), finance (Baur and Schulze, 2005), and public health (Winkelmann, 2006), to mention a few.

Extending the quantile regression analysis to complex survey data strengthens the potential of this class of models, as complex surveys are of fundamental importance for government agencies and policymakers in collecting strategic data to support decisions and for scientific advancement in different areas. A good example is the Programme for International Student Assessment (PISA) promoted by the Organisation for Economic Co-operation and Development (OECD) in various countries worldwide. The primary examples of complex surveys in public health rely on the Demographic and Health Surveys (DHS) Program, which collects data about childhood and maternal mortality, fertility, HIV, malaria, and tuberculosis in over 90 countries.

In this regard, Geraci (2016) introduced a survey-weighted estimator based on the asymmetric check loss function (Koenker and Bassett Jr., 1978) with data missing at random. Chen and Zhao (2019) and Nascimento and Gonçalves (2024) focused on informative designs and explored methods based on the survey-weighted estimator and the estimating equations. Zhao et al. (2020), Wang, Kim, and Yang (2018), and Savitsky and Toth (2016) developed general frameworks that encompass the quantile regression as particular cases. Zhao et al. (2020) introduced a Bayesian empirical likelihood approach defined through estimating equations. The authors demonstrated the design consistency of their estimator for general sampling designs and the asymptotic validity

of the Bayesian credible intervals. Wang, Kim, and Yang (2018) proposed a Bayesian approach that applies the sampling distribution of a summary statistic to obtain the posterior distribution of the parameters of interest. Savitsky and Toth (2016) constructed a pseudo-posterior distribution that utilizes sampling weights based on the marginal inclusion probabilities to exponentiate the likelihood contribution of each sampled unit.

A drawback of most approaches cited above is that since quantiles are fitted separately, the resultant regression lines may cross, violating the basic probabilistic rule that quantiles are monotonic functions and possibly causing problems for inference in practice. An alternative to overcome this problem is implementing a joint estimation and imposing constraints that ensure the monotonicity assumption. Considering the case in which data is randomly collected, Bondell, Reich, and Wang (2010) and Reich, Fuentes, and Dunson (2011) applied this solution. The former employed a constrained minimization problem to a weighted function, and the latter described the quantile process using Bernstein basis polynomials. He (1997) attained the noncrossing property restricting the space of solutions for the conditional quantiles. Chernozhukov, Fernández-Val, and Galichon (2010) suggested rearranging the estimated quantiles to obtain monotonic results.

More recently, Rodrigues and Fan (2017) introduced a two-stage approach, where the standard Bayesian quantile regression model (Kozumi and Kobayashi, 2011; Yu and Moyeed, 2001) is applied, succeeded by a Gaussian process regression adjustment, which monotonizes the quantile function. Rodrigues, Dortet-Bernadet, and Fan (2019a) proposed a new methodology for joint quantile regression modeling based on quantile pyramids, where the monotonicity rule is guaranteed. Rodrigues, Dortet-Bernadet, and Fan (2019b) extended this idea to the context of a spline regression setting. Zhou, Wang, and Feng (2020) presented a novel method to obtain noncrossing quantile estimates within the distributional reinforcement learning framework. Shen et al. (2024) proposed a penalty function to enforce the noncrossing of quantile regression curves in a nonseparable model using rectifier quadratic unit-activated deep neural networks.

Proposing a Bayesian approach that addresses crossing issues is challenging, as employing traditional Monte Carlo Markov Chain (MCMC) algorithms in the inference process of multiple

quantiles without violating the probabilistic rule that quantiles are monotonic functions can be computationally costly. In scenarios like this, Expectation-Maximization (EM) algorithms emerge as an appealing alternative (Caron and Doucet, 2012; Ročková and George, 2014; Ročková and McAlinn, 2021; Souza et al., 2024). In this paper, we develop a novel EM algorithm for non-crossing Bayesian quantile regression analysis under informative sampling. To our knowledge, we are the first to introduce an EM algorithm that handles crossing issues and also to investigate noncrossing methods for complex survey data analysis. Our approach differs from that of Zhao and Lian (2016) in three distinct ways. First, we incorporate survey design in the inference procedure and define the quantile regression coefficients at the super-population level. Second, we perform inference for multiple conditional quantiles of an arbitrary number by writing a joint posterior density function for the quantile levels. Third, we handle crossing issues by introducing a simple constraint in the maximization step. Our methodology exploits a conjugate prior formulation to find the posterior modes and encompasses an alternative to the traditional Monte Carlo Markov Chain (MCMC) algorithm for Bayesian inference that can be computationally costly when analyzing multiple quantiles and considering a constrained parameter space.

The remainder of the paper is organized as follows. Section 2 summarizes the Bayesian approach for quantile regression analysis under informative sampling. Section 3 presents the approach to handle crossing issues, detailing the Expectation-Maximization algorithm derivations. Simulations are performed in Section 4 to compare the performance of the proposed method with some established solutions. In Section 5, we analyze child growth standards using data from the Brazilian National Demographic and Health Survey of Women and Children. Section 6 concludes the work with final remarks and future directions.

2 A Bayesian approach for quantile regression analysis under informative sampling

Let $\mathbf{x} = (x_1, \dots, x_p)'$, $\mathbf{x} \in \mathbb{R}^p$, denote a vector of explanatory variables and y denote the response variable. Suppose a finite population $\mathcal{F}_N = \{(y_i, \mathbf{x}_i'), i = 1, \dots, N\}$ generated from a super-population model \mathcal{F} for which the true underlying distribution function is unknown. Through a linear quantile regression model, one can consider that for a fixed quantile level $\tau \in (0, 1)$ the super-population τ -th conditional quantile function of the response variable follows a parametric form as below:

$$Q_y(\tau|\mathbf{x}) = \beta_0(\tau) + \sum_{j=1}^p x_j \beta_j(\tau), \quad (1)$$

where $Q_y(\tau|\mathbf{x}) \equiv F_y^{-1}(\tau; \mathbf{x})$, $F_y(\tau; \mathbf{x})$ is the unknown conditional cumulative distribution function evaluated at τ .

A complex survey sample $S = \{i_1, \dots, i_n\}$ with size $|S| = n$ in which the units are sampled with inclusion probability $\pi_i = \mathbb{P}(I_i = 1)$, where $I_i \in \{0, 1\}$ is a sampling indicator such that $I_i = 1$ if unit i is selected and $I_i = 0$ otherwise, is drawn from a informative sampling design when π_i , $i \in S$, depends on the covariates (\mathbf{x}) and study variable (y). In this case, not incorporating the survey design in the estimation process may produce biased results for the finite population parameters of interest. Geraci (2016) and Chen and Zhao (2019) described a survey-weighted (Chen and Zhao, 2019; Geraci, 2016) estimator for the coefficients in (1) as the solution of the minimization problem:

$$\underset{\boldsymbol{\beta}(\tau) \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i \in S} w_i \rho_\tau(y_i - \mathbf{z}_i' \boldsymbol{\beta}(\tau)), \quad (2)$$

where $\mathbf{z}_i = (1, \mathbf{x}_i')'$, $\boldsymbol{\beta}(\tau) = (\beta_0(\tau), \beta_1(\tau), \dots, \beta_p(\tau))'$, and $\rho_\tau(\ell) = \ell(\tau - \mathbb{1}(\ell < 0))$ is the asymmetric check loss function (Koenker and Bassett Jr., 1978). As (Savitsky and Toth, 2016), we are taking normalized sampling weights $w_i = n\pi_i^{-1} / \sum_{i \in S} \pi_i^{-1}$.

From the previous expression, Nascimento and Gonçalves (2024) followed a similar argument to that of Yu and Moyeed (2001) and wrote the minimization problem in (2) as a maximization of a likelihood function reached from independently distributed asymmetric Laplace densities with location $\mathbf{z}'_i\boldsymbol{\beta}(\tau)$, scale w_i^{-1} , and skewness τ :

$$\begin{aligned}
\underset{\boldsymbol{\beta}(\tau) \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i \in S} w_i \rho_\tau(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau)) &= \underset{\boldsymbol{\beta}(\tau) \in \mathbb{R}^p}{\operatorname{argmax}} \exp \left\{ - \sum_{i \in S} w_i \rho_\tau(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau)) \right\} \\
&= \underset{\boldsymbol{\beta}(\tau) \in \mathbb{R}^p}{\operatorname{argmax}} \prod_{i \in S} \exp \{ -w_i \rho_\tau(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau)) \} \\
&= \underset{\boldsymbol{\beta}(\tau) \in \mathbb{R}^p}{\operatorname{argmax}} \prod_{i \in S} w_i \tau (1 - \tau) \\
&\quad \times \exp \{ -w_i \rho_\tau(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau)) \}.
\end{aligned} \tag{3}$$

Incorporating a scale parameter σ , Nascimento and Gonçalves (2024) proposed a Bayesian quantile regression for complex survey data based on the asymmetric Laplace distribution (ALD), where $y_i \sim \text{ALD}(\mathbf{z}'_i \boldsymbol{\beta}(\tau), w_i^{-1} \sigma, \tau)$. Observe that the authors do not assume the finite population is generated from a super-population model with an ALD likelihood function. Instead, they are using it as a “working likelihood” for Bayesian quantile inference, deeming it an efficient regression parameter estimation instrument.

Exploring the location-scale mixture representation for the ALD (Kozumi and Kobayashi, 2011), the model has a hierarchical structure as below:

$$\begin{aligned}
y_i | \mathbf{z}_i, \nu_{i\tau}, w_i, \boldsymbol{\beta}(\tau), \sigma(\tau) &\sim N(\mathbf{z}'_i \boldsymbol{\beta}(\tau) + \theta_\tau \nu_{i\tau}, \gamma_\tau^2 \nu_{i\tau} \sigma(\tau) / w_i) \\
\nu_{i\tau} | w_i, \sigma(\tau) &\sim \text{Exp}(\sigma(\tau) / w_i),
\end{aligned} \tag{4}$$

where $\theta_\tau = \frac{1-2\tau}{\tau(1-\tau)}$, $\gamma_\tau^2 = \frac{2}{\tau(1-\tau)}$, and $\text{Exp}(\zeta)$ denotes an exponential distribution with mean ζ . Assuming prior independence between $\boldsymbol{\beta}(\tau)$ and $\sigma(\tau)$, $p(\boldsymbol{\beta}(\tau), \sigma(\tau)) = p(\boldsymbol{\beta}(\tau))p(\sigma(\tau))$, and specifying conjugate priors, $\boldsymbol{\beta}(\tau) \sim N(\boldsymbol{\mu}_0, \Sigma_0)$ and $\sigma(\tau) \sim IG(a_0, b_0)$, the following full conditional densities with closed form are obtained:

$$\bullet \boldsymbol{\beta}(\tau) | \mathbf{y}, \mathbf{z}, \boldsymbol{\nu}_\tau, \mathbf{w}, \sigma(\tau) \sim N(\boldsymbol{\mu}_1, \Sigma_1), \Sigma_1 = \left[\sum_{i \in S} \frac{w_i}{\gamma_\tau^2 \sigma(\tau) \nu_{i\tau}} \mathbf{z}_i \mathbf{z}'_i + \Sigma_0^{-1} \right]^{-1} \text{ and}$$

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \Sigma_1 \left[\sum_{i \in \mathcal{S}} \frac{w_i(y_i - \theta_\tau \nu_{i\tau})}{\gamma_\tau^2 \sigma(\tau) \nu_{i\tau}} \mathbf{z}_i + \Sigma_0^{-1} \boldsymbol{\mu}_0 \right]; \\
\bullet \quad \sigma(\tau) | \mathbf{y}, \mathbf{z}, \boldsymbol{\nu}_\tau, \mathbf{w}, \boldsymbol{\beta}(\tau) &\sim IG \left(a_0 + \frac{3n}{2}, b_0 + \sum_{i \in \mathcal{S}} \left[\frac{w_i(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau) - \theta_\tau \nu_{i\tau})^2}{2\gamma_\tau^2 \nu_{i\tau}} + w_i \nu_{i\tau} \right] \right); \\
\bullet \quad \nu_{i\tau} | \mathbf{y}, \mathbf{z}, \mathbf{w}, \boldsymbol{\beta}(\tau), \sigma(\tau) &\sim GIG \left(\frac{1}{2}, \frac{w_i(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau))^2}{\gamma_\tau^2 \sigma(\tau)}, \frac{w_i \theta_\tau^2}{\gamma_\tau^2 \sigma(\tau)} + \frac{2w_i}{\sigma(\tau)} \right), i \in \mathcal{S},
\end{aligned}$$

where $\mathbf{y} = (y_{i_1}, \dots, y_{i_n})$, $\mathbf{z} = (\mathbf{z}'_{i_1}, \dots, \mathbf{z}'_{i_n})$, $\boldsymbol{\nu}_\tau = (\nu_{i_1\tau}, \dots, \nu_{i_n\tau})$, $\mathbf{w} = (w_{i_1}, \dots, w_{i_n})$. In addition, $IG(a, b)$ denotes the Inverse Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$, and $GIG(\delta, a, b)$ is the generalized inverse Gaussian distribution whose density is given by

$$f(x|\delta, a, b) = \frac{(b/a)^\delta}{2K_\delta(ab)} x^{\delta-1} \exp \left\{ -\frac{1}{2} (a^2 x^{-1} + b^2 x) \right\}, \quad x > 0, \quad -\infty < \delta < \infty, \quad a, b \geq 0,$$

and $K_\delta(\cdot)$ is the modified Bessel function of third kind.

From the previous full conditional densities, a simple MCMC algorithm with only Gibbs steps can be employed for posterior analysis, but each quantile of interest is evaluated separately. When we are interested in q quantile levels such that $\tau_1 < \dots < \tau_q$, fitting separate models may result in crossing issues, implying conditional quantile curves that will not be a monotonically increasing function of the quantile. In the subsequent section, we introduce a method based on the Expectation-Maximization (EM) algorithm that addresses crossing issues.

3 An EM algorithm for noncrossing Bayesian quantile regression

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}(\tau_1)', \dots, \boldsymbol{\beta}(\tau_q)')$ and $\boldsymbol{\sigma} = (\sigma(\tau_1), \dots, \sigma(\tau_q))$ be the parameters of interest. Following the hierarchical representation for the ALD in (4), assuming prior independence, $p(\boldsymbol{\beta}, \boldsymbol{\sigma}) = \prod_{k=1}^q p(\boldsymbol{\beta}(\tau_k))p(\sigma(\tau_k))$, and specifying conjugate priors, $\boldsymbol{\beta}(\tau_k) \sim N(\boldsymbol{\mu}_{0k}, \Sigma_{0k})$ and $\sigma(\tau_k) \sim IG(a_{0k}, b_{0k})$ for $k = 1, \dots, q$, we can write the posterior density function of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, and the la-

latent variables conditional on the observed data as below:

$$\begin{aligned}
p(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\nu} | \mathbf{y}, \mathbf{z}, \mathbf{w}) &\propto \prod_{k=1}^q f(\mathbf{y} | \mathbf{z}, \boldsymbol{\nu}_{\tau_k}, \mathbf{w}, \boldsymbol{\beta}(\tau_k), \sigma(\tau_k)) f(\boldsymbol{\nu}_{\tau_k} | \mathbf{w}, \sigma(\tau_k)) p(\boldsymbol{\beta}(\tau_k)) p(\sigma(\tau_k)) \\
&\propto \prod_{k=1}^q \sigma(\tau_k)^{-3n/2} \exp \left\{ - \sum_{i \in \mathcal{S}} \left[\frac{w_i (y_i - \mathbf{z}_i' \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} \nu_{i\tau_k})^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) \nu_{i\tau_k}} + \frac{w_i \nu_{i\tau_k}}{\sigma(\tau_k)} \right] \right\} \quad (5) \\
&\times \exp \left\{ - \frac{1}{2} (\boldsymbol{\beta}(\tau_k) - \boldsymbol{\mu}_{0k})' \Sigma_{0k}^{-1} (\boldsymbol{\beta}(\tau_k) - \boldsymbol{\mu}_{0k}) \right\} \\
&\times \sigma(\tau_k)^{-a_0-1} \exp \left\{ - \frac{b_{0k}}{\sigma(\tau_k)} \right\},
\end{aligned}$$

where $\boldsymbol{\nu} = (\boldsymbol{\nu}_{\tau_1}, \dots, \boldsymbol{\nu}_{\tau_q})$, $\theta_{\tau_k} = \frac{1-2\tau_k}{\tau_k(1-\tau_k)}$ and $\gamma_{\tau_k}^2 = \frac{2}{\tau_k(1-\tau_k)}$.

The EM algorithm finds the posterior modes of the parameter posterior by indirectly maximizing $p(\boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{y}, \mathbf{z}, \mathbf{w})$, and proceeding in respect of the “complete-data” log posterior, $\log(p(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\nu} | \mathbf{y}, \mathbf{z}, \mathbf{w}))$, where the latent variables $\boldsymbol{\nu}$ are treated as “missing data”. As the log posterior is not observable, the E-step replaces it with the conditional expectation given current parameter estimates and observed data. Afterward, the M-step maximizes the expected log posterior with respect to $(\boldsymbol{\beta}, \boldsymbol{\sigma})$. By iterating between E and M steps, the EM algorithm produces a sequence of parameter estimates, which converge toward a local maximum of $p(\boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{y}, \mathbf{z}, \mathbf{w})$.

Objectively, our EM algorithm indirectly maximizes $p(\boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{y}, \mathbf{z}, \mathbf{w})$ by iteratively maximizing

the following objective function:

$$\begin{aligned}
Q(\boldsymbol{\beta}, \boldsymbol{\sigma}) &= \mathbb{E}_{\boldsymbol{\nu}} [\log(p(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\nu} | \mathbf{y}, \mathbf{z}, \mathbf{w}))] \\
&= C - \mathbb{E}_{\boldsymbol{\nu}} \left[\sum_{k=1}^q \left(\frac{3n + a_{0k} + 1}{2} \log(\sigma(\tau_k)) + \sum_{i \in \mathcal{S}} \frac{w_i (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} \nu_{i\tau_k})^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) \nu_{i\tau_k}} \right. \right. \\
&\quad \left. \left. + \sum_{i \in \mathcal{S}} \frac{w_i \nu_{i\tau_k}}{\sigma(\tau_k)} + \frac{b_{0k}}{\sigma(\tau_k)} + \frac{1}{2} \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\beta}(\tau_k) - \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\mu}_{0k} \right) \right] \\
&= C - \sum_{k=1}^q \left(\frac{3n + a_{0k} + 1}{2} \log(\sigma(\tau_k)) + \mathbb{E}_{\boldsymbol{\nu}} \left[\sum_{i \in \mathcal{S}} \frac{w_i (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} \nu_{i\tau_k})^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) \nu_{i\tau_k}} \right] \right. \\
&\quad \left. + \mathbb{E}_{\boldsymbol{\nu}} \left[\sum_{i \in \mathcal{S}} \frac{w_i \nu_{i\tau_k}}{\sigma(\tau_k)} \right] + \frac{b_{0k}}{\sigma(\tau_k)} + \frac{1}{2} \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\beta}(\tau_k) - \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\mu}_{0k} \right) \\
&= C - \sum_{k=1}^q \left(\frac{3n + a_{0k} + 1}{2} \log(\sigma(\tau_k)) + \sum_{i \in \mathcal{S}} \frac{w_i (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} / \mathbb{E}[\nu_{i\tau_k}^{-1}])^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) / \mathbb{E}[\nu_{i\tau_k}^{-1}]} \right. \\
&\quad \left. + \sum_{i \in \mathcal{S}} \frac{w_i \theta_{\tau_k}^2 (\mathbb{E}[\nu_{i\tau_k}] - \mathbb{E}[\nu_{i\tau_k}^{-1}]^{-1})}{2\gamma_{\tau_k}^2 \sigma(\tau_k)} + \sum_{i \in \mathcal{S}} \frac{w_i \mathbb{E}[\nu_{i\tau_k}]}{\sigma(\tau_k)} + \frac{b_{0k}}{\sigma(\tau_k)} + \frac{1}{2} \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\beta}(\tau_k) \right. \\
&\quad \left. - \boldsymbol{\beta}(\tau_k)' \Sigma_{0k}^{-1} \boldsymbol{\mu}_{0k} \right), \tag{6}
\end{aligned}$$

where C is a constant. Observe that the term $\mathbb{E}_{\boldsymbol{\nu}} [-\sum_{k=1}^q \sum_{i \in \mathcal{S}} \log(\nu_{i\tau_k}/2)]$ was absorbed into the constant C as it does not influence the subsequent M-step. In the third equality in (6), we used

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\nu}} \left[\sum_{i \in \mathcal{S}} \frac{w_i (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} \nu_{i\tau_k})^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) \nu_{i\tau_k}} \right] \\
&= \mathbb{E}_{\boldsymbol{\nu}} \left[\sum_{i \in \mathcal{S}} w_i \left(\frac{(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k))^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) \nu_{i\tau_k}} - \frac{(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k)) \theta_{\tau_k}}{\gamma_{\tau_k}^2 \sigma(\tau_k)} + \frac{\theta_{\tau_k}^2 \nu_{i\tau_k}}{2\gamma_{\tau_k}^2 \sigma(\tau_k)} \right) \right] \\
&= \sum_{i \in \mathcal{S}} w_i \left(\frac{(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k))^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) / \mathbb{E}[\nu_{i\tau_k}^{-1}]} - \frac{(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k)) \theta_{\tau_k}}{\gamma_{\tau_k}^2 \sigma(\tau_k)} + \frac{\theta_{\tau_k}^2 \mathbb{E}[\nu_{i\tau_k}]}{2\gamma_{\tau_k}^2 \sigma(\tau_k)} \right) \\
&= \sum_{i \in \mathcal{S}} w_i \left(\frac{(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k))^2 - 2(y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k)) \theta_{\tau_k} / \mathbb{E}[\nu_{i\tau_k}^{-1}] + \theta_{\tau_k}^2 / \mathbb{E}^2[\nu_{i\tau_k}^{-1}]}{2\gamma_{\tau_k}^2 \sigma(\tau_k) / \mathbb{E}[\nu_{i\tau_k}^{-1}]} \right) \\
&\quad + \sum_{i \in \mathcal{S}} w_i \left(\frac{\theta_{\tau_k}^2 \mathbb{E}[\nu_{i\tau_k}]}{2\gamma_{\tau_k}^2 \sigma(\tau_k)} - \frac{\theta_{\tau_k}^2 / \mathbb{E}[\nu_{i\tau_k}^{-1}]}{2\gamma_{\tau_k}^2 \sigma(\tau_k)} \right) \\
&= \sum_{i \in \mathcal{S}} \frac{w_i (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} / \mathbb{E}[\nu_{i\tau_k}^{-1}])^2}{2\gamma_{\tau_k}^2 \sigma(\tau_k) / \mathbb{E}[\nu_{i\tau_k}^{-1}]} + \sum_{i \in \mathcal{S}} \frac{w_i \theta_{\tau_k}^2 (\mathbb{E}[\nu_{i\tau_k}] - \mathbb{E}[\nu_{i\tau_k}^{-1}]^{-1})}{2\gamma_{\tau_k}^2 \sigma(\tau_k)}.
\end{aligned}$$

3.1 The E-step

The E-step proceeds by computing the conditional expectations $E[\nu_{i\tau_k}]$ and $E[\nu_{i\tau_k}^{-1}]$ for $Q(\beta, \sigma)$. Given that we saw in the previous section that the full conditional distribution of $\nu_{i\tau_k}$ follows a generalized inverse Gaussian (GIG) distribution, we have that

$$E[\nu_{i\tau_k}^{-1}] = \frac{a_1}{b_1} \text{ and } E[\nu_{i\tau_k}] = \frac{b_1}{a_1} \times \frac{K_{3/2}(a_1 b_1)}{K_{1/2}(a_1 b_1)},$$

where $a_1 = \sqrt{\frac{(y_i - z_i' \beta(\tau_k))^2}{\gamma_{\tau_k}^2 \sigma(\tau_k)}}$ and $b_1 = \sqrt{\frac{\theta_{\tau_k}^2}{\gamma_{\tau_k}^2 \sigma(\tau_k)} + \frac{2}{\sigma(\tau_k)}}$. The equalities follow from the fact that $E[\nu_{i\tau_k}^r] = \left(\frac{a_1}{b_1}\right)^r \frac{K_{1/2+r}(a_1 b_1)}{K_{1/2}(a_1 b_1)}$ (Karlis, 2002) and $K_{1/2}(\cdot) = K_{-1/2}(\cdot)$ (Abramowitz and Stegun, 1968).

3.2 The M-step

To maximize the expected log-posterior, we first maximize $Q(\beta, \sigma)$ with respect to β under a restriction that alleviates the crossing issue. For that, we need to solve the following optimization problem:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \sum_{k=1}^q -\frac{1}{2} \beta(\tau_k)' (z' \Sigma_{\tau_k}^{-1} z + \Sigma_{0k}^{-1}) \beta(\tau_k) + \beta(\tau_k)' (z' \Sigma_{\tau_k}^{-1} \mathbf{y}_* + \Sigma_{0k}^{-1} \mu_{0k}), \quad (7)$$

where $\Sigma_{\tau_k} = \gamma_{\tau_k}^2 \sigma(\tau_k) \operatorname{diag}((E[w_{i_1} \nu_{i_1 \tau_k}^{-1}])^{-1}, \dots, (E[w_{i_n} \nu_{i_n \tau_k}^{-1}])^{-1})$, and $\mathbf{y}_* = \mathbf{y} - \theta_{\tau_k} / E[\nu_{\tau_k}^{-1}]$, subject to a constraint that guarantees monotonicity among the quantiles in a computationally efficient way.

Applying $z_i' \beta(\tau_k) \geq z_i' \beta(\tau_{k-1})$, for all $i \in \mathcal{S}$ and $k = 2, \dots, q$, results in a large number of constraints. We instead employ the ideas of Bondell, Reich, and Wang (2010), diminishing the number of constraints to a total of $q-1$. It means that we reduce the domain of interest to $\mathbf{x} \in [0, 1]^p$ and consider the transformation from $\beta(\tau_1), \dots, \beta(\tau_q)$ to $\delta(\tau_1), \dots, \delta(\tau_q)$, where for $k = 2, \dots, q$ $\delta_{\tau_1} = (\delta_0(\tau_1), \delta_1(\tau_1), \dots, \delta_p(\tau_1))' = \beta(\tau_1)$ and $\delta(\tau_k) = (\delta_0(\tau_k), \delta_1(\tau_k), \dots, \delta_p(\tau_k))' = \beta(\tau_k) - \beta(\tau_{k-1})$. Breaking each $\delta_j(\tau_k)$ into its positive and negative parts, so that $\delta_j(\tau_k) = \delta_j^+(\tau_k) - \delta_j^-(\tau_k)$,

where both $\delta_j^+(\tau_k)$ and $\delta_j^-(\tau_k)$ are nonnegative, and only one may be nonzero, we simply need solve (7) subject to

$$\delta_0(\tau_k) - \sum_{j=1}^p \delta_j^-(\tau_k) \geq 0 \quad (k = 2, \dots, q).$$

Regarding the scale parameters, we take the first derivative of the objective function with respect to the parameter $\sigma(\tau_k)$, $k = 1, \dots, q$, and equal to 0, obtaining

$$\begin{aligned} \hat{\sigma}(\tau_k) = & \left[2b_{0k} + \sum_{i \in \mathcal{S}} w_i E[\nu_{i\tau_k}^{-1}] (y_i - \mathbf{z}'_i \boldsymbol{\beta}(\tau_k) - \theta_{\tau_k} / E[\nu_{i\tau_k}^{-1}])^2 + \sum_{i \in \mathcal{S}} w_i \theta_{\tau_k}^2 (E[\nu_{i\tau_k}] - E[\nu_{i\tau_k}^{-1}]^{-1}) \right. \\ & \left. + 2\gamma_{\tau_k}^2 \sum_{i \in \mathcal{S}} w_i E[\nu_{i\tau_k}] \right] \times [(3n + a_{0k} + 1)\gamma_{\tau_k}^2]^{-1}. \end{aligned}$$

4 Model-based simulation study

This simulation study intends to evaluate the noncrossing method proposed (NCBQR) in Section 3 more comprehensively, comparing it to the Bayesian quantile regression approach (BQR) proposed by Nascimento and Gonçalves (2024). For that, we draw 1,000 finite populations with size $N = 5,000$ from four different data-generating processes (DGPs):

DGP 1: $\beta_0(\tau) = \log[\tau/(1 - \tau)]$, $\beta_1(\tau) = 2$, $\beta_2(\tau) = 1$;

DGP 2: $\beta_0(\tau) = \text{sign}(0.5 - \tau) \log(1 - 2|0.5 - \tau|)$, $\beta_1(\tau) = 2\tau$, $\beta_2(\tau) = 1$;

DGP 3: $\beta_0(\tau) = \Phi^{-1}(\tau)$, $\beta_1(\tau) = 2\min\{\tau - 0.5, 0\}$, $\beta_2(\tau) = 1$;

DGP 4: $\beta_0(\tau) = 2\Phi^{-1}(\tau)$, $\beta_1(\tau) = 2\min\{\tau - 0.5, 0\}$, $\beta_2(\tau) = 2\tau$.

For each DGP, finite populations are drawn by generating $U_i \stackrel{iid}{\sim} (0, 1)$, $X_{ij} \stackrel{iid}{\sim} (0, 1)$, and $Y_i = \beta_0(U_i) + \sum_j \beta_j(U_i)X_{ij}$ for $i = 1, 2, \dots, N$ and $j = 1, 2$. These specifications and the preceding DGPs are similar to those analyzed by Reich and Smith (2013) and Rodrigues and Fan (2017). For each finite population, we employ probability proportional to size (PPS) samplings with expected size equal to $n = 250$ and inclusion probabilities $\pi_i = nk_i / \sum_{j=1}^N k_j$, where $k_i = \{1 + \exp(2.5 -$

$0.5z_i)\}^{-1}$ and $z_i \sim N(1 + y_i, 0.25)$. These particular choices for k_i and z_i were also assumed by Chen and Zhao (2019). In our exercise, we consider three different PPS schemes: Poisson, stratified, and systematic.

We assume the Bayesian quantile regression under informative sampling proposed by Nascimento and Gonçalves (2024), which does not consider crossing issues, as a benchmark. Hereafter, it will be referred to as BQR, and the methodology presented in Section 3 as NCBQR. Nine quantiles ($q = 9$) are fitted to data, where $\tau \in \{0.05, 0.10, 0.20, 0.35, 0.50, 0.65, 0.80, 0.90, 0.95\}$. The number of draws in which at least one crossing occurs and the root mean integrated squared error averaged across \mathbf{X} , $\text{RMISE} = \sqrt{1/N \sum_{i=1}^N \{\hat{Q}_{y_i}(\tau|\mathbf{x}_i) - Q_{y_i}(\tau|\mathbf{x}_i)\}^2}$, are computed for comparison purposes. Note that the quantile curves are fitted regarding the samples, while the RMISE is calculated considering the finite populations. For both methods, we assume non-informative priors for the coefficients by taking $\boldsymbol{\mu}_{0k} = \mathbf{0}_3$, $\Sigma_{0k} = 1,000I_3$, and $a_{0k} = b_{0k} = 0.001$ for all $k = 1, 2, \dots, 9$, where $\mathbf{0}_3$ is a 3-dimensional vector of zeros and I_3 is an identity matrix with dimension 3×3 .

Table 1: Number of draws out of the 1,000 generated that which at least one cross is observed in the finite population domain

	Poisson	Stratified	Systematic
DGP 1	677	379	676
DGP 2	678	413	680
DGP 3	561	281	474
DGP 4	833	416	826

Table 1 presents the number of draws out of the 1,000 generated datasets in which at least one cross is observed in the finite population domain. Although we are adjusting both models using the samples, we check crossing issues referred to the finite population since it is the object of interest of our inference procedure. Additionally, reporting crossing issues is relevant as there is a trade-off between this quantity and the RMISE, for example. In many real applications, even if noncrossing methods are better regarding RMISE, researchers and practitioners might not like using them when quantiles cross.

Figure 1: Root mean integrated squared error (RMISE) results over the 1,000 samples for data-generating processes 1 to 4 and Poisson, stratified, and systematic sampling designs. Points refer to the median, and line ranges refer to the interval between quantiles 0.05 and 0.95.

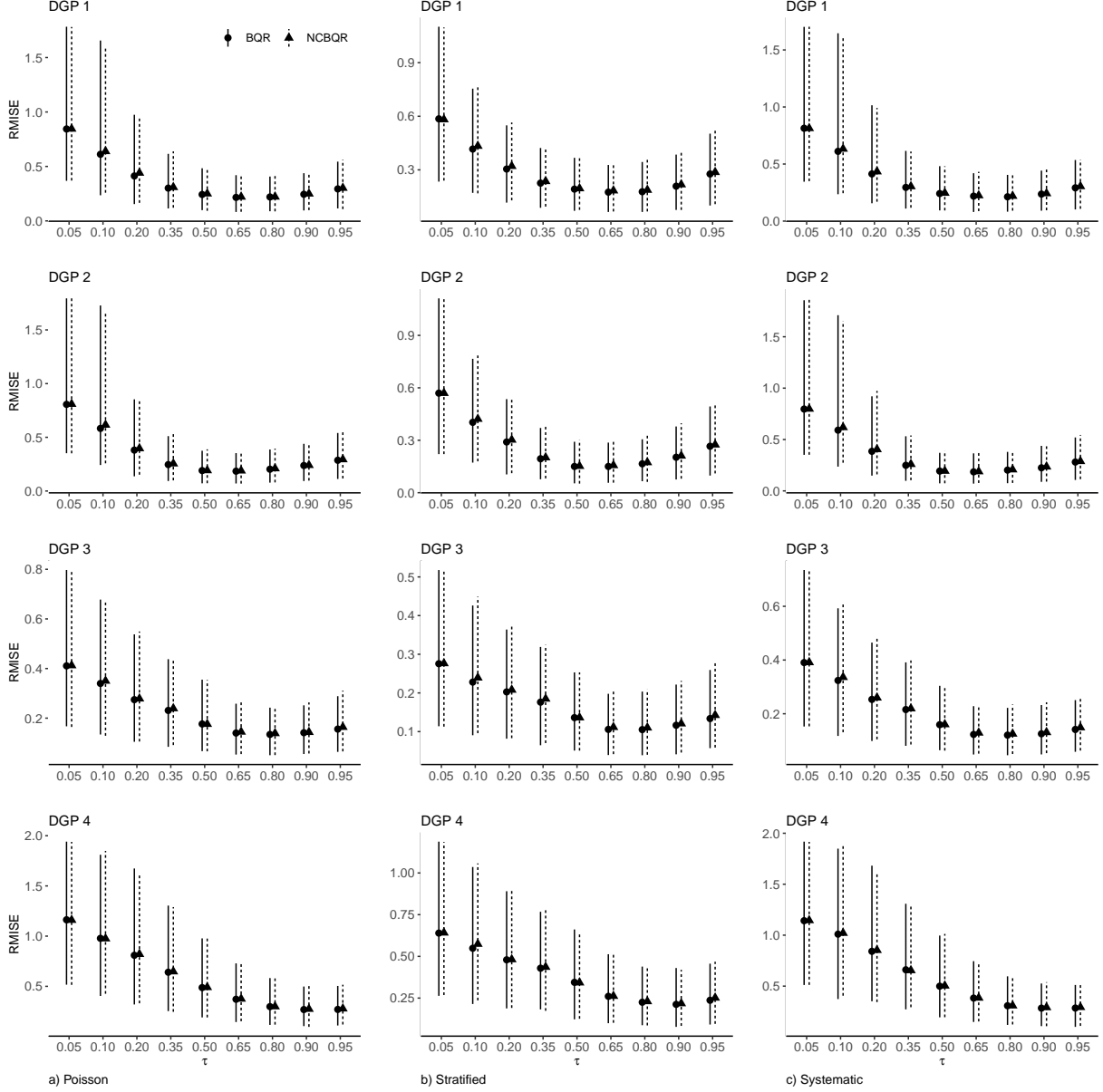


Figure 1 summarizes the RMISE from all 1,000 generated data regarding the data-generating processes, the three sampling schemes, and the quantiles under analysis. We observe that regardless of the scenario, the results are very similar, indicating that the NCBQR handles crossing issues with no or almost no impact on the performance in terms of the RMISE. These results might be interpreted in favor of the NCBQR when the trade-off between crossing and performance is con-

sidered.

When dealing with data collected from a simple random sampling scheme, quantities such as RMISE and estimator bias and variance are expected to increase as we deviate from the median and go to the right and left tails in similar magnitudes. However, in Figure 1, we note that RMISE for quantiles in the left tail is higher and varies more for lower quantiles than the upper ones. Regarding DGP 3 and 4, it is possible to observe that it also occurs for the median. These effects reflect the PPS designs under analysis, where a lower response variable implies lower probability sampling. It means that making inference regarding the left tail quantiles in the finite population gets even more challenging.

5 An application to the 2006 Brazilian NDHS

The 2006 Brazilian National Demographic and Health Survey of Women and Children (NDHS) intended to describe women of reproductive age (15-49 years) population and children under the age of 5 according to demographic, socioeconomic, and cultural factors; identify marital, parental, and fertility patterns; identify childhood morbidity and mortality and breastfeeding profiles; evaluate the nutritional state; assess food insecurity and the iodine content in the household context; and assess the availability of medical services. The study universe comprised household residences stratified into ten independent strata - the five Brazilian macro-regions (North, Northeast, Central-West, Southeast, and South) as well as urban and rural areas.

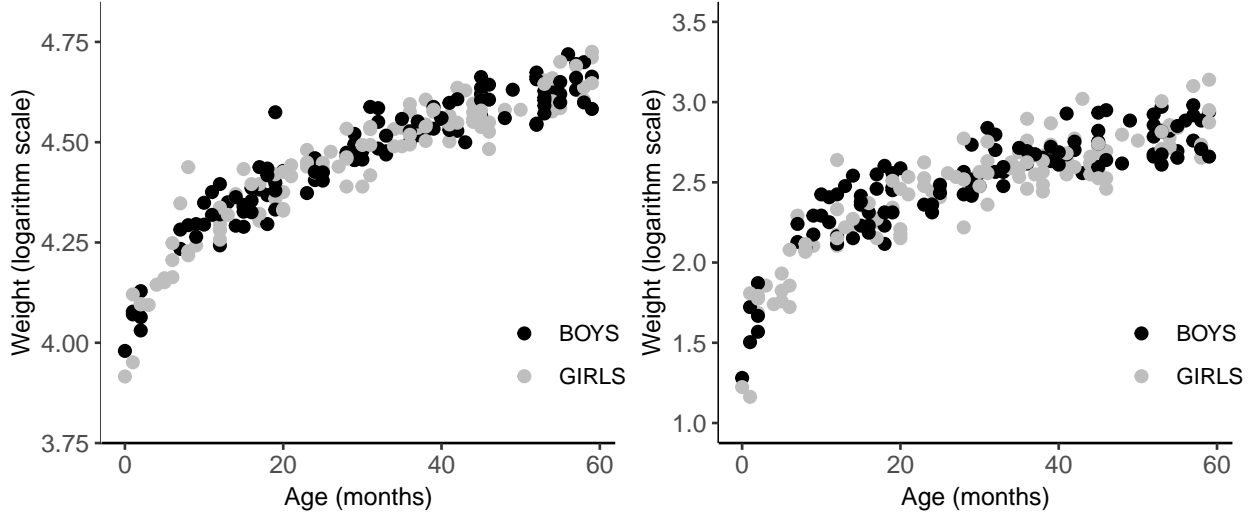
The sampling units were selected in two stages: primary sampling units (PSUs) comprised census tracts, and secondary sampling units (SSUs) consisted of households. Tract selection in each stratum was designed to ensure a minimum number of blood samples, based on the prevalence of vitamin A deficiency in children. Information was obtained through various methods, including face-to-face questionnaires, document presentations, anthropometric measurements, identification of iodine concentration in cooking salt available at the domicile, and blood sample collection and laboratory analysis of vitamin A and hemoglobin levels.

We consider children's height and weight as study variables and age (x_1) and an indicator variable for girls (x_2) as predictors. The sampling mechanism is informative in that height and weight correlate with vitamin A deficiency, a risk factor for child development. Vitamin A deficiency is observed in young children and pregnant women in low-income countries and causes delayed growth and development in children. Our analysis focuses on the strata from the rural zone of the Northeast region, comprising 224 observations after excluding non-collected data due to nonresponse and equipment difficulties during anthropometric measurements.

Figure 2 illustrates the dataset under analysis. Due to the observed smooth variation of weight and height concerning age, we assume a quadratic model in age. Yu and Moyeed (2001), Kottas and Krnjajic (2009), and Rodrigues and Fan (2017) employed a similar specification when applying their quantile regression models to the Immunoglobulin G data (Isaacs et al., 1987). We also include an indicator variable for sex, as child growth curves are frequently presented separately by gender. Hence, we have:

$$Q_y(\tau|\mathbf{x}) = \beta_0(\tau) + \beta_1(\tau)x_1 + \beta_2(\tau)x_1^2 + \beta_3x_2.$$

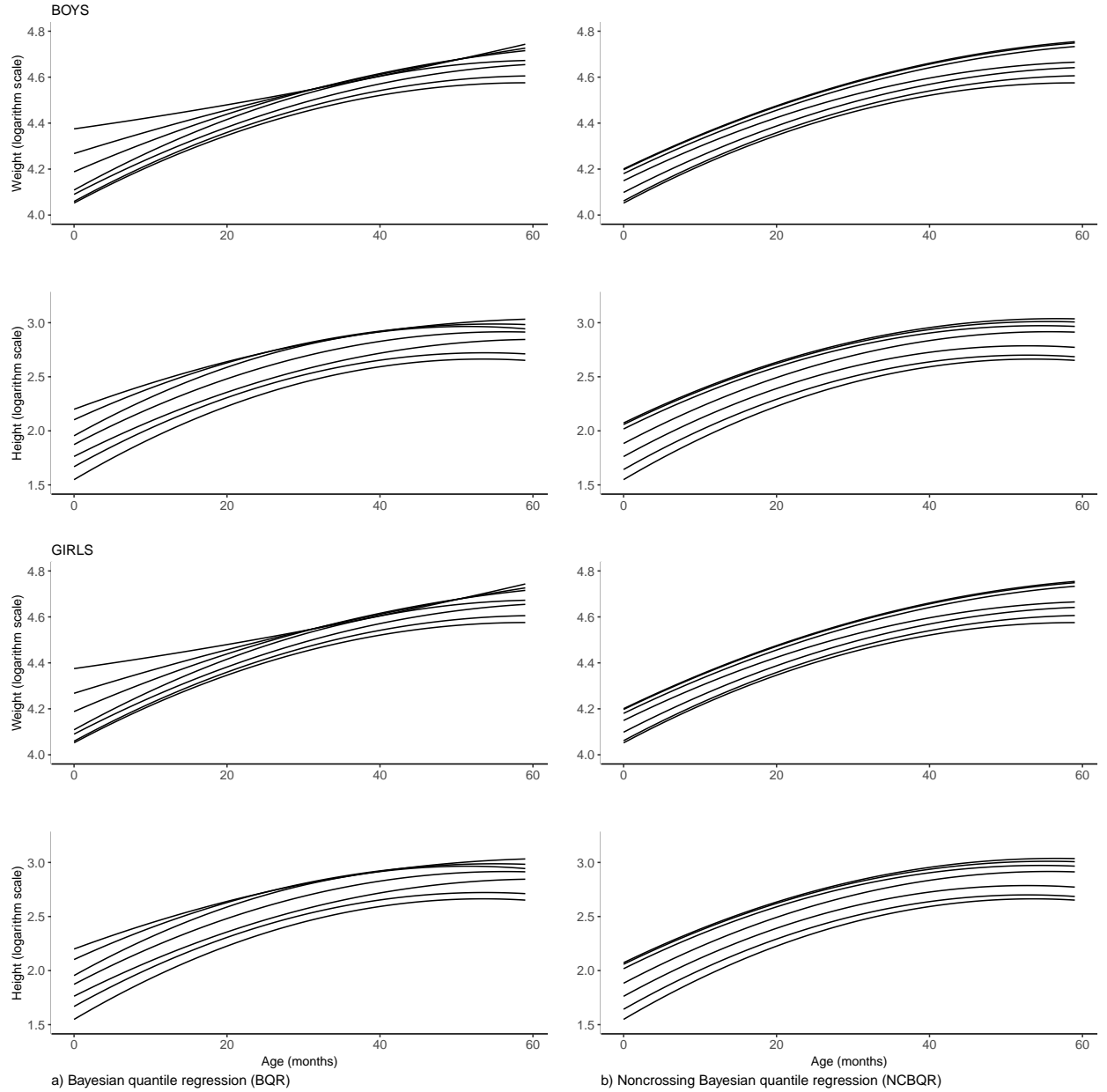
Figure 2: Children's height and weight: Birth to 5 years.



In our analysis, we estimate the conditional quantile levels $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90, 0.95, 0.975\}$.

We specify non-informative priors for the coefficients and the scale parameters by taking $\boldsymbol{\mu}_{0k} = \underline{0}_4$, $\Sigma_{0k} = 1,000I_4$, and $a_{0k} = b_{0k} = 0.001$ for all $k = 1, 2, \dots, 9$. Figure 3 presents the BQR and the NCBQR results. We note that the independent fitting of the BQR implies crossing curves, particularly at $\tau = 0.90$, $\tau = 0.95$, and $\tau = 0.975$, for both response variables under analysis (weight and height) and both sexes. In contrast, the same does not occur with the NCBQR, and the monotonicity constraint is respected.

Figure 3: Quantile regression of children's height and weight: Birth to 5 years.



6 Final remarks

When quantiles are fitted separately, the resultant regression lines may cross, violating the basic probabilistic rule that quantiles are monotonic functions and possibly causing problems for inference in practice. This issue is well reported in the literature, and different solutions have been

proposed Bondell, Reich, and Wang (2010); Chernozhukov, Fernández-Val, and Galichon (2010); He (1997); Rodrigues and Fan (2017). However, to our knowledge, none of these methods have explored the case in which data are collected through complex surveys, particularly under informative designs. Therefore, our main contribution in this paper was to introduce a method for multiple quantile levels that handles crossing issues and complex survey data under informative sampling. The proposed EM algorithm can also be applied to handle crossing quantiles in the usual simple random sampling, as it is encompassed in our methodology as a particular case ($w_i = 1$ for all $i \in \mathcal{S}$).

The proposed method consisted of an Expectation-Maximization algorithm, and its fundamental ingredients were based on consolidated ideas from both Bayesian quantile regression and non-crossing literature. First, we used the location-scale mixture representation of the asymmetric Laplace distribution (Kozumi and Kobayashi, 2011) and its conjugate formulation. Second, we employed the constrained term in Bondell, Reich, and Wang (2010). From these, we introduced a new methodology that performs inference for multiple quantile levels, handles crossing issues, and incorporates survey design. In a model-based simulation study, the algorithm was compared to a benchmark that ignores crossing issues, and it was verified that the method handles crossing curves with no negative impact on the estimation performance under different data-generating processes and different sampling schemes. A real data analysis illustrated how our approach can be effective in practice.

Extensions of the proposed EM algorithm to frameworks beyond linear regression provide interesting directions for new methodological developments. For example, using the latent variable inferential approach of Albert and Chib (1993), a straightforward probit extension for the classification of binary responses can be implemented with an additional E-step to calculate the expected values of the latent continuous data. Following an analogous idea, extending our method to ordinal data (Rahman, 2016) could also be a direction. Further interesting directions are to consider variable selection (Alhamzawi and Yu, 2013). Finally, as full posterior inference has been lost in place of computational feasibility, the uncertainty description is another important avenue for

future research.

References

- Abramowitz, Milton, and Irene A Stegun. 1968. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Vol. 55. US Government printing office.
- Albert, James H., and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88 (422): 669–679.
- Alhamzawi, Rahim, and Keming Yu. 2013. "Conjugate priors and variable selection for Bayesian quantile regression." *Computational Statistics & Data Analysis* 64: 209–219.
- Austin, Peter C., Jack V. Tu, Paul A. Daly, and David A. Alter. 2005. "The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy." *Statistics in Medicine* 24 (5): 791–816.
- Baur, Dirk, and Niels Schulze. 2005. "Coexceedances in financial markets - a quantile regression analysis of contagion." *Emerging Markets Review* 6 (1): 21–43.
- Bondell, Howard D., Brian J. Reich, and Huixia Wang. 2010. "Noncrossing quantile regression curve estimation." *Biometrika* 97 (4): 825–838.
- Caron, François, and Arnaud Doucet. 2012. "Efficient Bayesian Inference for Generalized Bradley–Terry Models." *Journal of Computational and Graphical Statistics* 21 (1): 174–196.
- Chen, Sixia, and Yan Daniel Zhao. 2019. "Quantile Regression Analysis of Survey Data Under Informative Sampling." *Journal of Survey Statistics and Methodology* 7 (2): 157–174.
- Chernozhukov, Victor, Ivan Fernández-Val, and Alfred Galichon. 2010. "Quantile and Probability Curves Without Crossing." *Econometrica* 78 (3): 1093–1125.
- Geraci, Marco. 2016. "Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants." *Statistical Methods in Medical Research* 25 (4): 1393–1421.
- Gilpin, Gregory A. 2012. "Teacher salaries and teacher aptitude: An analysis using quantile regressions." *Economics of Education Review* 31 (3): 15–29.

- He, Xuming. 1997. "Quantile Curves without Crossing." *The American Statistician* 51 (2): 186–192.
- Isaacs, D., D. G. Altman, C. E. Tidmarsh, H. B. Valman, and A. D. Webster. 1987. "Serum immunoglobulin concentrations in preschool children measured by laser nephelometry: reference ranges for IgG, IgA, IgM." *Journal of Clinical Pathology* 36 (10): 1193–1196.
- Karlis, Dimitris. 2002. "An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution." *Statistics & Probability Letters* 57 (1): 43–52.
- Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica* 46 (1): 33–50.
- Kottas, Athanasios, and Milovan Krnjajic. 2009. "Bayesian Semiparametric Modelling in Quantile Regression." *Scandinavian Journal of Statistics* 36 (2): 297–319.
- Kozumi, Hideo, and Genya Kobayashi. 2011. "Gibbs sampling methods for Bayesian quantile regression." *Journal of Statistical Computation and Simulation* 81 (11): 1565–1578.
- Nascimento, Marcus L., and Kelly C. M. Gonçalves. 2024. "Bayesian quantile regression models for complex survey data under informative sampling." *Journal of Survey Statistics and Methodology* 12 (4): 1105–1130.
- Rahman, Mohammad Arshad. 2016. "Bayesian Quantile Regression for Ordinal Models." *Bayesian Analysis* 11 (1): 1–24.
- Reich, Brian J., Montserrat Fuentes, and David B. Dunson. 2011. "Bayesian Spatial Quantile Regression." *Journal of the American Statistical Association* 106 (493): 6–20.
- Reich, Brian J., and Luke B. Smith. 2013. "Bayesian quantile regression for censored data." *Biometrics* 69 (3): 651–660.
- Rodrigues, Thais, Jean-Luc Dortet-Bernadet, and Yanan Fan. 2019a. "Pyramid Quantile Regression." *Journal of Computational and Graphical Statistics* 28 (3): 732–746.
- Rodrigues, Thais, Jean-Luc Dortet-Bernadet, and Yanan Fan. 2019b. "Simultaneous fitting of Bayesian penalised quantile splines." *Computational Statistics & Data Analysis* 134: 93–109.
- Rodrigues, Thais, and Yanan Fan. 2017. "Regression Adjustment for Noncrossing Bayesian Quantile Regression." *Journal of Computational and Graphical Statistics* 26 (2): 275–284.
- Ročková, Veronika, and Edward I. George. 2014. "EMVS: The EM Approach to Bayesian Variable Selection." *Journal of the American Statistical Association* 109 (506): 828–846.

- Ročková, Veronika, and Kenichiro McAlinn. 2021. “Dynamic Variable Selection with Spike-and-Slab Process Priors.” *Bayesian Analysis* 16 (1): 233–269.
- Savitsky, Terrance D., and Daniell Toth. 2016. “Bayesian estimation under informative sampling.” *Electronic Journal of Statistics* 10 (1): 1677–1708.
- Shen, Guohao, Yuling Jiao, Yuanyuan Lin, Joel L. Horowitz, and Jian Huang. 2024. “Nonparametric Estimation of Non-Crossing Quantile Regression Process with Deep ReQU Neural Networks.” *Journal of Machine Learning Research* 25 (88): 1–75.
- Souza, Rebecca, Lilia Costa, Marina Paez, João Sato, and Candida Barreto. 2024. “Dynamic Graphical Models with Variable Selection for Effective Connectivity.” *Bayesian Analysis* 19 (4): 1041–1065.
- Villarini, Gabriele, James A. Smith, Mary Lynn Baeck, Renato Vitolo, David B. Stephenson, and Witold F. Krajewski. 2011. “On the frequency of heavy rainfall for the Midwest of the United States.” *Journal of Hydrology* 400 (1–2): 103–120.
- Wang, Zhonglei, Jae Kwang Kim, and Shu Yang. 2018. “Approximate Bayesian inference under informative sampling.” *Biometrika* 105 (1): 91–102.
- Winkelmann, Rainer. 2006. “Reforming health care: Evidence from quantile regressions for counts.” *Journal of Health Economics* 25 (1): 131–145.
- Yu, Keming, and Rana A. Moyeed. 2001. “Bayesian quantile regression.” *Statistics & Probability Letters* 55 (4): 437–447.
- Zhao, Kaifeng, and Heng Lian. 2016. “The Expectation–Maximization approach for Bayesian quantile regression.” *Computational Statistics and Data Analysis* 96: 1–11.
- Zhao, Puying, Malay Ghosh, J. N. K. Rao, and Changbao Wu. 2020. “Bayesian empirical likelihood inference with complex survey data.” *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 82 (1): 155–174.
- Zhou, Fan, Jianing Wang, and Xingdong Feng. 2020. “Non-Crossing Quantile Regression for Distributional Reinforcement Learning.” In *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, Vol. 33, 15909–15919. Curran Associates, Inc.