Flexible Bayesian quantile curve fitting with shape restrictions under the Dirichlet process mixture of the generalized asymmetric Laplace distribution

Genya KOBAYASHI[1], Taeyoung ROH[2], Jangwon LEE[2], and Taeryon CHOI[2*] ⓘ

[1]*Graduate School of Social Sciences, Chiba University, Chiba, 263-8522, Japan*
[2]*Department of Statistics, Korea University, Seoul, 02841, Republic of Korea*

Abstract: We propose a flexible Bayesian semiparametric quantile regression model based on Dirichlet process mixtures of generalized asymmetric Laplace distributions for fitting curves with shape restrictions. The generalized asymmetric Laplace distribution exhibits more flexible tail behaviour than the frequently used asymmetric Laplace distribution in Bayesian quantile regression. In addition, nonparametric mixing over the shape and scale parameters with the Dirichlet process mixture extends its flexibility and improves the goodness of fit. By assuming the derivatives of the regression functions to be the squares of the Gaussian processes, our approach ensures that the resulting functions have shape restrictions such as monotonicity, convexity and concavity. The introduction of shape restrictions prevents overfitting and helps obtain smoother and more stable estimates of the quantile curves, especially in the tail quantiles for small and moderate sample sizes. Furthermore, the proposed shape-restricted quantile semiparametric regression model deals with sparse estimation for regression coefficients using the horseshoe+ prior distribution, and it is extended to cases with group-specific curve estimation and censored data. The usefulness of the proposed models is demonstrated using simulated datasets and real applications. *The Canadian Journal of Statistics* 49: 698–730; 2021 © 2020 Statistical Society of Canada

Résumé: Les auteurs proposent un modèle semi-paramétrique bayésien de régression quantile basé sur un mélange de distributions de Laplace asymétriques généralisées pondéré par un processus de Dirichlet pour l'ajustement de courbes avec des contraintes de forme. La distribution de Laplace asymétrique généralisée exhibe une plus grande flexibilité codale que la distribution de Laplace asymétrique fréquemment employée pour la régression quantile bayésienne. De plus, la pondération non paramétrique des paramètres de forme et d'échelle avec un processus de Dirichlet améliore la flexibilité et l'adéquation. En supposant que les dérivées des fonctions de régression sont le carré de processus gaussiens, les auteurs s'assurent que les fonctions résultantes respectent des restrictions de forme telles que la monotonicité, la convexité ou la concavité. L'introduction des contraintes de forme prévient le surapprentissage et contribue à donner des estimés des courbes quantiles plus lisses et plus stables, notamment dans les quantiles des queues pour des tailles d'échantillon petites ou modérées. De plus, le modèle proposé peut gérer l'estimation des coefficients de régression éparses en utilisant la loi a priori en fer à cheval+. Les auteurs généralisent leur modèle aux cas d'estimation de courbes spécifiques aux groupes et aux données censurées. Ils démontrent son utilité avec des données simulées et réelles. *La revue canadienne de statistique* 49: 698–730; 2021 © 2020 Société statistique du Canada

---

* *Author to whom correspondence may be addressed.*
E-mail: trchoi@korea.ac.kr

1. INTRODUCTION

Investigating the relationship between quantiles of the conditional distribution of a response variable and a set of covariates offers a key insight into the structure of the data. As an important approach to quantile inference, quantile regression has received substantial attention and has gained prominence as an important alternative to conventional mean regression since the seminal work of Koenker & Bassett (1978). There now exists a large body of the literature on the theory and methods of quantile regression under the frequentist approach. See, for example, Koenker (2005) and Davino, Furno & Vistocco (2014) for an overview.

Recently, interest in the Bayesian approach to quantile regression models has been growing. For example, Yu & Moyeed (2001) and Kozumi & Kobayashi (2011) considered Bayesian quantile regression models for continuous responses using the asymmetric Laplace (AL) distribution. The models are estimated using the Markov chain Monte Carlo (MCMC) method. Additionally, Yu & Stander (2007), Kobayashi & Kozumi (2012) and Kobayashi (2017) considered models for censored responses. These studies only considered regression models in which the conditional quantile function is modelled through a linear function of the covariates. As in conventional regression analysis, however, linear models under the quantile regression framework often fail to describe the complex relationship between the conditional distribution and covariates adequately. In addition, a set of linear quantile regression estimates obtained for multiple quantile levels may violate the monotonicity of the quantile function (Koenker, 2005).

Bayesian quantile regression models that include nonlinear functions of covariates have also been actively studied. The unknown nonlinear function can be modelled by placing a suitable prior distribution, such as the Gauss Markov random field (Yue & Rue, 2011) or Gaussian processes (Jo, Roh & Choi, 2016) or by using a linear combination of some basis spline functions (e.g., Chen & Yu, 2009; Thompson et al., 2010; Waldmann et al., 2013; Hu, Zhao & Lian, 2015). Although Bayesian nonlinear quantile regression provides a flexible method for studying the effect of covariates on the response distribution, the resulting estimates often exhibit large fluctuations and less interpretability due to overfitting, especially for lower and higher quantiles. While these works have aimed to estimate the conditional quantile at one quantile level at a time, some recent studies have simultaneously estimated the conditional function as a function of the quantile level. See, for example, Reich & Smith (2013), Yang & Tokdar (2017) and Das & Ghosal (2018).

Compared with the aforementioned works, this article considers a flexible shape-restricted Bayesian nonlinear quantile regression model. Its contribution toward improving the stability and interpretability of the estimates is three-fold. First, following the work of Lenk & Choi (2017), we incorporate shape restrictions into the Bayesian nonlinear quantile regression model proposed by Jo, Roh & Choi (2016), which is based on the spectral representation of the Gaussian process. Specifically, the derivatives of the functions are assumed to be the squares of the Gaussian processes such that the resulting functions are monotonic, monotonic convex, or concave. By introducing a unique point at which the derivatives change signs, it is also possible to express U-shaped functions. In addition, we consider a Bayesian hierarchical approach to estimate group-specific quantile curves across observations in different subpopulations, which typically arise in cross-sectional or longitudinal studies.

Second, we adopt the generalized asymmetric Laplace (GAL) distribution, which was recently developed by Yan & Kottas (2017) and Rahman & Karnawat (2019), and propose the Dirichlet process mixture of the GAL distribution (GALDP) over the shape and scale parameters. Although the frequently used AL distribution provides posterior consistency (Sriram, Ramamoorthi & Ghosh, 2013), concerning the data distribution, it exhibits severe limitations in its tail behaviour. In addition, the location of its mode is constrained to be equal to zero, which is the quantile at the level of interest. The GAL distribution also allows for flexibility in the distributional shape, and its mode is not necessarily equal to zero. Further, the shape of the GAL distribution is controlled

by two parameters, one of which is fixed to the quantile level of interest in quantile regression. However, fixing the quantile at the level of interest to zero can still limit tail behaviour, even though it mitigates the limitations of the AL distribution. Hence, we consider the more flexible distribution modelled by the GALDP. By nonparametrically mixing over the shape and scale parameters of the GAL distribution, we gain further flexibility and accuracy regarding the nonlinear effect on the conditional quantile function of interest. Hence, the proposed Dirichlet process mixture approach presented in this article improves the estimation of the quantile level of interest and better explains tail behaviour than those based on either parametric or nonparametric extensions of the AL distribution considered in the literature (e.g., Kottas & Krnjajić, 2009; Reich, Bondell & Wang, 2010; Taddy & Kottas, 2010; Wichitaksorn, Choy & Gerlach, 2014; Jo, Roh & Choi, 2016).

Third, we introduce the horseshoe+ (HS+) prior distribution proposed by Bhadra et al. (2017) for the coefficient vector in the linear component, which is known to possess the super-efficient shrinkage property. While various shrinkage priors such as lasso (Kozumi and Kobayashi, 2011), elastic net (Li, Xi & Lin, 2010), and adaptive lasso (Alhamzawi, Yu & Benoit et al., 2012) have been applied to Bayesian quantile regression, this article is the first to employ the HS+ prior for sparse estimation in the context of Bayesian quantile semiparametric regression. Additionally, the present article extends the proposed models to cases with group-specific curve estimation and censored data to meet the demand for quantile regression analysis for these types of data.

The rest of this article is organized as follows. Section 2 discusses the quantile semiparametric regression model to motivate our methodology. We start by briefly reviewing the Bayesian nonlinear regression model with spectral representation and quantile regression with the AL and GAL distributions. Then, we propose the GALDP for quantile regression. The proposed methodology is combined with shape restrictions on the nonparametric function and the sparse estimation with the HS+ prior. Section 3 presents posterior inference, including the detailed MCMC steps for posterior computation and model selection procedures for shape selection and the choice of error distribution, together with possible extensions to group-specific estimation and the censored quantile regression model. The simulation studies in various settings are presented in Section 4. The proposed models are also illustrated using real applications in Section 5. Finally, Section 6 concludes.

2. METHODOLOGY

2.1. Bayesian Spectral Analysis Quantile Regression

Consider the partial linear additive quantile regression model given by

$$y_i = g(x_i) + w_i'\beta + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where y_i is the response variable, g is an unknown function of the observed scalar covariate $x_i \in [0, 1]$ with the mean-centring condition $\int_0^1 g(x)dx = 0$, $w_i = (1, w_{1i}, \dots, w_{qi})'$ is the vector of covariates that includes the intercept but does not include functions of x_i, and $\beta = (\beta_0, \beta_1, \dots, \beta_q)'$ is the coefficient vector for w_i. The mean-centring condition is introduced to identify the intercept. This condition is also required to reduce the posterior uncertainty of g and the posterior correlation between the intercept and g, preventing both of them from being poorly estimated (see, Lenk & Choi, 2017 for further discussion). It is possible to include more than one unknown function in an additive manner, $\sum_{k=1}^K g_k(x_i)$ in Equation (1). The pth quantile of the error term ϵ_i is equal to zero, and thus the pth conditional quantile of y_i given the covariates (w, x) is expressed as

$$Q_{y|w,x}(p) = g(x) + w'\beta.$$

In the Bayesian quantile regression literature, it is frequently assumed that the ϵ_i independently and identically follows the AL distribution with probability density function given by

$$f_{AL}(\epsilon|\sigma,p) = \frac{p(1-p)}{\sigma} \exp\left\{-\gamma_p\left(\frac{\epsilon}{\sigma}\right)\right\},$$

where $p \in (0,1)$, $\sigma > 0$ is the scale parameter, $\gamma_p(u)$ is the asymmetric check loss given by $\gamma_p(u) = u(p - I(u < 0))$, and $I(\cdot)$ is the indicator function (Yu & Zhang, 2005). The AL distribution allows the following location scale mixture of the normal distribution,

$$\epsilon = Av + \sqrt{\sigma Bv}u, \quad v \sim Exp(\sigma), \quad u \sim N(0,1), \tag{2}$$

where $Exp(\psi)$ denotes the exponential distribution with mean ψ, $A = (1 - 2p)/(p - p^2)$, and $B = 2/(p - p^2)$ (Kozumi & Kobayashi, 2011). For a given value of p, the mode and the pth quantile of the AL distribution coincide at zero. Posterior consistency for Bayesian quantile regression based on the AL likelihood is provided by Sriram, Ramamoorthi & Ghosh (2013). However, in the context of quantile regression where p is fixed by users, the AL distribution has a severe limitation as an error distribution because its shape is solely determined by p.

The nonlinear function g in Equation (1) is assumed to be a second-order Gaussian process Z with mean function equal to zero and covariance function $Cov(s,t) = E[Z(s)Z(t)]$ for $s,t \in [0,1]$. Subsequently, the Gaussian process $g(\equiv Z)$ is expressed as an infinite series expansion with the Karhunen–Loéve representation in terms of the cosine basis functions given by

$$g(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \quad \theta_j = \int_0^1 g(x)\varphi_j(x)dx, \quad 0 \le x \le 1, \tag{3}$$

$$\varphi_0(x) = 1, \quad \varphi_j(x) = \sqrt{2}\cos(\pi j x), \quad j \ge 1.$$

All piecewise continuous functions on $[0,1]$ are expressed by the representation in Equation (3) and the infinite sum is truncated at a sufficiently large J, with $g_J(x) = \sum_{j=0}^{J} \theta_j \varphi_j(x)$ in practice (see, e.g., Lenk, 1999; Choi, Lee & Roy, 2009; Lenk & Choi, 2017). The Bayesian approach to exploiting the cosine series to represent a Gaussian process is known as the Bayesian spectral analysis model (Lenk & Choi, 2017; Jo et al., 2019), and the corresponding quantile regression model is referred to as the Bayesian spectral analysis quantile regression (BSAQ) hereafter.

BSAQ under the AL distribution is a flexible approach to studying the nonlinear effect of a covariate on the conditional distribution of a response. However, nonlinear quantile regression estimates can be unstable, especially for lower and upper quantiles and with small to moderate sample sizes. Furthermore, although the AL distribution provides a valid pseudo likelihood for quantile regression, it has a severe limitation in detecting tail behaviour.

The GAL distribution developed by Yan & Kottas (2017) and Rahman & Karnawat (2019) is a flexible extension of the AL distribution in that the mode of the distribution does not necessarily correspond to the location and thus, the specified quantile. The flexibility of the GAL-based quantile regression model is further enhanced by introducing a nonparametric mixture over the parameters of the GAL distribution.

As an illustrative toy example, 200 observations are generated from Equation (1) with no covariates other than x with $g(x) = 10x^3$ that is monotone increasing, where $x \sim U(0,1)$, and the error distribution is given by the mixture of two normals as $0.7N(0, 0.5^2) + 0.3N(0.5, 2^2)$. The following four Bayesian models are fit to estimate the conditional quantile function: the unrestricted BSAQ under the AL and GAL distributions, the proposed GALDP, as well as the monotone BSAQ under the GAL and GALDP. We also fit the same observations with local
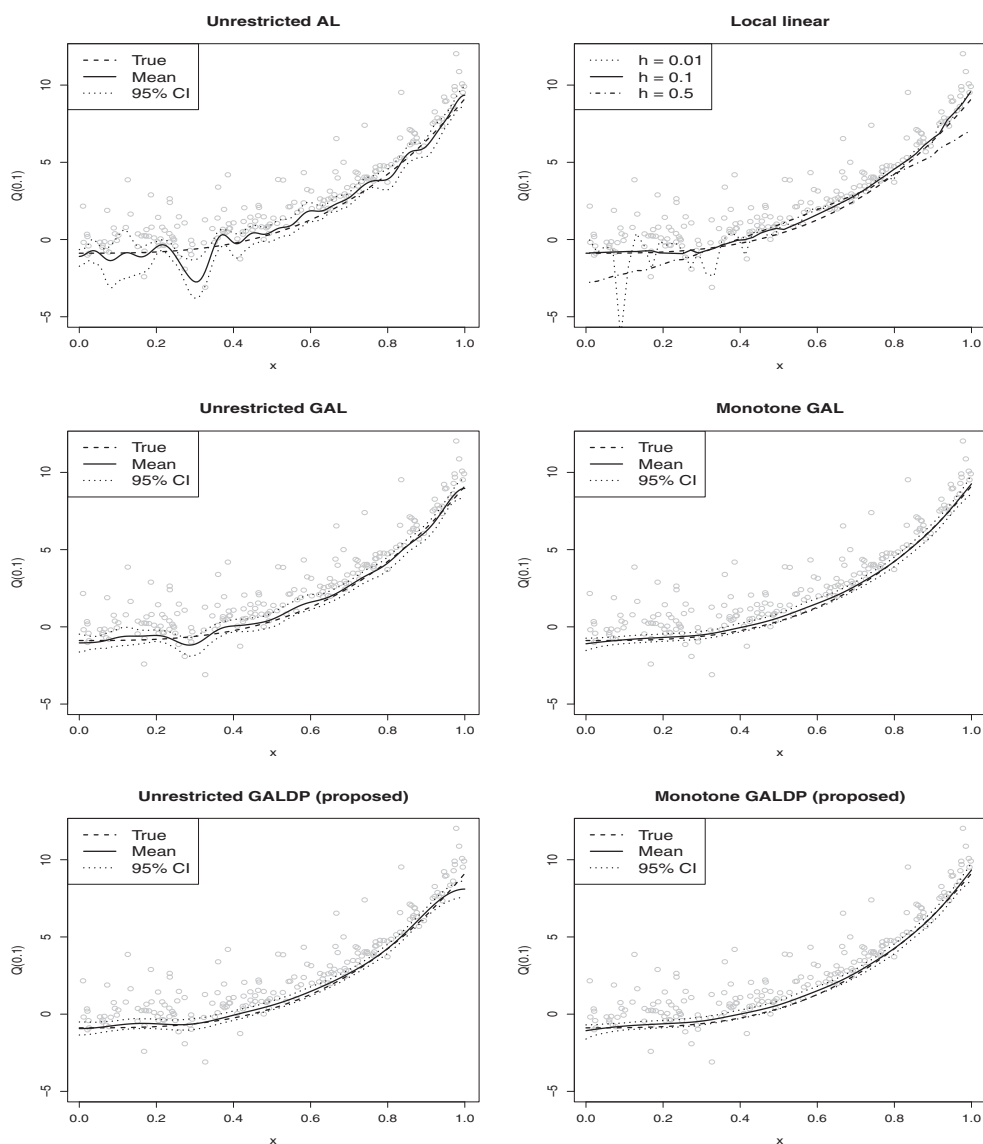
FIGURE 1: An illustrative toy example with a monotone increasing function $g(x) = 10x^3$ generated from the error distribution with a mixture of two normals, with the posterior means and their 95% credible intervals for the 0.1th conditional quantile curve. The models used are the unrestricted BSAQ under the AL and GAL distributions, the GALDP, the monotone BSAQ under the GAL and GALDP distributions, and local linear regression.

linear quantile regression for comparison, using the R package `quantreg` (Koenker, 2019) with three bandwidths, 0.01, 0.1 and 0.5. Figure 1 presents the posterior means and their 95% credible intervals for the 0.1th conditional quantile function based on the aforementioned respective models. Figure 1 shows that the quantile curve estimate based on the unrestricted model under the AL distribution appears wobbly. Moreover, its credible interval widens in the region where the data are relatively sparse and some observations have large magnitudes. Introducing the GAL distribution makes the estimate more stable and smoother, and a further improvement is

observed by employing the GALDP. As shown in the bottom panels of Figure 1, the estimate can be smoothed out further by restricting g to be monotone.

The log pseudo marginal likelihood (LPML, see Section 3.2) of the unrestricted BSAQ under the AL, GAL and GALDP distributions, and of the monotone BSAQ under the GAL and GALDP distributions are -340.2, -287.4, -278.9, -285.010 and -272.5, respectively. The monotone BSAQ under the GALDP distribution results in the largest LPML and is selected as the best model supported by the data. From Figure 1 and the LPML, the use of the GALDP distribution and an appropriate shape restriction leads to an improved stable estimation of the nonlinear effects and better goodness of fit. For local linear quantile regression, monotonicity is not well captured in all cases, depending on the bandwidth, and the choice of bandwidth affects the smoothness of the estimated function, as is the case for any kernel smoothing method.

2.2. GALDP

Yan & Kottas (2017) recently proposed the GAL distribution as a flexible parametric family for Bayesian quantile regression and provided the location-scale mixture of the normal distribution. Rahman & Karnawat (2019) further considered a version of the mixture representation,

$$\epsilon = \alpha s + Av + \sqrt{\sigma B v} u, \tag{4}$$

$$s \sim N^+(0, \sigma^2), \quad v \sim Exp(\sigma), \quad u \sim N(0, 1),$$

where $\alpha \in \mathbb{R}$ is the skewness parameter and $N^+(\mu, \phi^2)$ denotes the normal distribution with mean μ and variance ϕ^2 truncated over $\mathbb{R}^+$. When $\alpha = 0$, Equation (4) reduces to the mixture representation for the AL distribution in Equation (2). By integrating over s and v, the probability density function of the GAL distribution is given by

$$f_{GAL}(\epsilon|\sigma, p, \alpha) = \frac{2p(1-p)}{\sigma} \left(\left[\Phi\left(\frac{\epsilon^*}{\alpha}\right) - \Phi\left(-p_{\alpha_-}\alpha\right) \right] \exp\left\{ -p_{\alpha_-}\epsilon^* + \frac{(p_{\alpha_-}\alpha)^2}{2} \right\} I\left(\frac{\epsilon^*}{\alpha} > 0\right) \right.$$

$$\left. + \Phi\left[p_{\alpha_+}\alpha - \frac{\epsilon^*}{\alpha} I\left(\frac{\epsilon^*}{\alpha} > 0\right) \right] \exp\left\{ -p_{\alpha_+}\frac{\epsilon}{\sigma} + \frac{(p_{\alpha_+}\alpha)^2}{2} \right\} \right), \tag{5}$$

where $\epsilon^* = \epsilon/\sigma$, $p_{\alpha_+} = p - I(\alpha > 0)$, $p_{\alpha_-} = p - I(\alpha < 0)$, $p \in (0, 1)$, $\alpha \in \mathbb{R}$ and $\Phi(\cdot)$ is the distribution function of $N(0, 1)$.

In Equation (5) for $\alpha \neq 0$, the parameter p does not satisfy $\int_{-\infty}^{0} f_{GAL}(\epsilon|\sigma, p, \alpha)d\epsilon = p$. Therefore, Yan & Kottas (2017) considered $\gamma = |\alpha|(I(\alpha > 0) - p)$ such that the resulting density function $f_{GALp_0}(\epsilon|\sigma, p, \gamma)$ satisfies $\int_{-\infty}^{0} f_{GALp_0}(\epsilon|\sigma, p, \gamma)d\epsilon = p_0$ for a specified value of p_0. Thus, the density function of the quantile-fixed GAL distribution is given by

$$f_{GALp_0}(\epsilon|\sigma, \gamma) = \frac{2p(1-p)}{\sigma} \times \left(\left[\Phi\left(-\frac{p_{\gamma_+}\epsilon^*}{|\gamma|} + \frac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma| \right) - \Phi\left(\frac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma| \right) \right] \right.$$

$$\times \exp\left\{ -p_{\gamma_-}\epsilon^* + \frac{\gamma^2}{2}\left(\frac{p_{\gamma_-}}{p_{\gamma_+}}\right)^2 \right\} I\left(\frac{\epsilon^*}{\gamma} > 0\right)$$

$$\left. + \Phi\left[-|\gamma| + \frac{p_{\gamma_+}\epsilon^*}{|\gamma|} I\left(\frac{\epsilon^*}{\gamma} > 0\right) \right] \exp\left\{ -p_{\gamma_+}\epsilon^* \frac{\gamma^2}{2} \right\} \right),$$

where $\epsilon^* = \epsilon/\sigma$, $p = I(\gamma < 0) + [p_0 - I(\gamma < 0)]/r(\gamma)$, $p_{\gamma_+} = p - I(\gamma > 0)$, $p_{\gamma_+} = p - I(\gamma > 0)$, $p_{\gamma_-} = p - I(\gamma < 0)$ and $r(\gamma) = 2\Phi(-|\gamma|)\exp\{\gamma^2/2\}$. The parameter γ has bounded support over the interval (γ_L, γ_U) where γ_L is the negative root of $r(\gamma) = 1 - p_0$ and γ_U is the positive root of $r(\gamma) = p_0$. See Yan & Kottas (2017) for the specific values depending on p_0. The distribution function of the quantile-fixed GAL distribution is given by

$$
\begin{aligned}
F_{GAL_{p_0}}(\epsilon|\sigma,\gamma) = \Bigg(&\left[1 - 2\Phi\left(\epsilon^* \frac{p_{\gamma_+}}{\gamma}\right) + 2p_{\gamma_+}\exp\left\{-\epsilon^* p_{\gamma_-} + \frac{\gamma^2}{2}\left(\frac{p_{\gamma_-}}{p_{\gamma_+}}\right)^2\right\}\right. \\
&\times \left[\Phi\left(-\frac{\epsilon^* p_{\gamma_+}}{|\gamma|} + \frac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma|\right) - \Phi\left(\frac{p_{\gamma_-}}{p_{\gamma_+}}|\gamma|\right)\right] I\left(\frac{\epsilon^*}{\gamma} > 0\right)\bigg] + I(\gamma < 0) \\
&+ 2p_{\gamma_-}\exp\left\{-\epsilon^* p_{\gamma_+} + \frac{\gamma^2}{2}\right\}\Phi\left(-|\gamma| + \epsilon^* \frac{p_{\gamma_+}}{|\gamma|} I\left(\frac{\epsilon^*}{\gamma} > 0\right)\right)\Bigg),
\end{aligned}
$$

as discussed by Rahman & Karnawat (2019) in detail. The mixture representation for the quantile-fixed GAL distribution is given by

$$\epsilon = C|\gamma|s + Av + \sqrt{\sigma Bv}u, \tag{6}$$

$$s \sim N^+(0, \sigma^2), \quad v \sim Exp(\sigma), \quad u \sim N(0, 1),$$

where $C = (I(\gamma > 0) - p)^{-1}$. Unlike the quantile-fixed parametric families proposed by Wichitaksorn, Choy & Gerlach (2014) and Naranjo, Pérez & Martín (2015), since the mode of the quantile-fixed GAL distribution does not necessarily correspond to the p_0th quantile, the quantile-fixed GAL distribution provides a more flexible parametric alternative to the AL distribution.

Nevertheless, the GAL distribution has limitations in that the parametric assumptions might not be supported by the data, and thus, we further explore flexible nonparametric modelling for the error distribution. This nonparametric extension is particularly useful as illustrated by the toy example in Section 2.1, and it yields important advantages in practical applications. Specifically, we develop nonparametric probability models for the error distribution by mixing the GAL distribution over (σ, γ) with Dirichlet process mixtures:

$$f(\epsilon|G) = \int f_{GAL_{p_0}}(\epsilon|\sigma, \gamma)dG(\sigma, \gamma),$$

where G is the mixing distribution supported on $\mathbb{R}^+ \times (\gamma_L, \gamma_U)$. We place the Dirichlet process prior for the mixing distribution $G \sim DP(\nu, G_0)$, where $DP(\nu, G_0)$ denotes the Dirichlet process with the precision parameter $\nu > 0$ and base distribution G_0.

Specifically, for the base measure of (σ, γ), we use $G_0 = Ga(a_{0\sigma}, b_{0\sigma}) \times U(\gamma_L, \gamma_U)$, where $Ga(a, b)$ represents the gamma distribution with the parameters a and b. Provided that the data are appropriately scaled, the values of the hyperparameters are determined in such a way that the model provides a reasonable fit to the data. For the precision parameter of the Dirichlet process, we assume $\nu \sim Ga(a_{0\nu}, b_{0\nu})$ with $a_{0\nu} = b_{0\nu} = 2$ such that both small and large values of ν are allowed. By using the stick-breaking representation of Sethuraman (1994), we can write

$$f(\epsilon|G) = \sum_{k=1}^{\infty} \pi_k f_{GAL_{p_0}}(\epsilon|\sigma_k, \gamma_k), \tag{7}$$

where $\pi_1 = z_1$, $\pi_k = z_k \prod_{l<k}(1 - z_l)$ and $z_k \sim Be(1, \nu)$, the beta distribution with parameters a and b denoted by $Be(a, b)$. Furthermore, the proposed BSAQ under the GALDP is also termed the flexible BSAQ (FBSAQ). We discuss the shape restrictions and sparse estimation for the FBSAQ in the following section.

2.3. Shape Restrictions and Sparse Estimation

Following Lenk & Choi (2017), we first incorporate shape restrictions into the FBSAQ. Specifically, Lenk & Choi (2017) assumed that the derivatives of the regression functions are the squares of Gaussian processes such that the resulting functions are monotonic and monotonic convex or concave. The qth derivative of g is assumed to be positive as the square of a Gaussian process, that is, $g^{(q)} = \delta Z^2(x)$, $\delta = 1$ or -1. For example, when $q = 1$, g is a monotone function given by

$$g(x) = \delta \left[\int_0^x Z^2(s)ds - \int_0^1 \int_0^x Z^2(s)dsdx \right], \tag{8}$$

where the second term is the integration constant that satisfies the mean-centring condition (Lenk & Choi, 2017), and g is nondecreasing when $\delta = 1$ and nonincreasing when $\delta = -1$. Similarly, when $q = 2$, g is either a monotone convex or a monotone concave function given by

$$g(x) = \delta \left[\int_0^x \int_0^s Z^2(t)dtds - \int_0^1 \int_0^x \int_0^s Z^2(t)dtdsdx \right] + \varrho(x - 0.5), \tag{9}$$

where the second and third terms are integration constants that satisfy the mean-centring condition as in Equation (8). In addition, we consider the U-shape constraint for $g(x)$ (Lenk & Choi, 2017) by introducing a decreasing logistic function h between 1 and -1 that forces the first derivative of g to change signs at a unique point w,

$$g'(x) = \delta Z^2(x)h(x), \quad h(x) = \frac{1 - \exp\{\psi(x - \omega)\}}{1 + \exp\{\psi(x - \omega)\}}, \quad \psi > 0, \quad \omega \in (0, 1).$$

Note that we use $\delta = 1$ for the inverted U-shape and $\delta = -1$ for U-shape, which leads to the FBSAQ with a U-shaped restriction (FBSAQ-U) given by

$$g(x) = \delta \left[\int_0^x Z^2(s)h(s)ds - \int_0^1 \int_0^x Z^2(s)h(s)dsdx \right].$$

Based on these integral representations in Equations (8) and (9), the corresponding $g(x)$ is expressed as a quadratic form of $\{\theta_j\}$ in terms of the double sum of infinite series, $g(x) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j \theta_k \tilde{\varphi}_{jk}(x)$. Here, $\tilde{\varphi}_{jk}$ denotes the generic form of a basis function, which is determined by each integral representation and evaluated in either an analytic form (see Appendix) or a numerical integration using the trapezoid rule. Additionally, in practice, the truncated series of the infinite sum at J are used and the resulting $J + 1$ coefficients $\theta_J = (\theta_0, \theta_1, \ldots, \theta_J)'$ are estimated. Further details and discussions can be found in Lenk & Choi (2017).

For the FBSAQ with and without shape restrictions, we employ the prior distributions used by Jo, Roh & Choi (2016) and Lenk & Choi (2017),

$$\theta_0 \sim N^+(0, v_{0\theta_0}^2),$$

$$\theta_j \sim N(0, \tau^2 \exp(-\rho j)), \quad j = 1, \ldots, J,$$

$$\tau^2 \sim IG(a_{0\tau^2}, b_{0\tau^2}), \quad \rho \sim Exp(e_{0\rho}),$$

where N^+ denotes a truncated normal distribution that is positive, and IG denotes an inverse gamma distribution. Note that θ_0 is only used for the FBSAQ with shape restrictions and that, since the spectral coefficients have a sign indeterminacy, we assume $\theta_0 \geq 0$ for identification (Lenk & Choi, 2017). For the remainder of this article, we set hyperparameters for θ_0, τ^2 and ρ, fixed as $v_{0\theta_0}^2 = 100$, $a_{0\tau^2} = 2$, $b_{0\tau^2} = 0.05$ and $e_{0\rho} = 1$ as in Jo et al. (2019) and Lenk & Choi (2017). To ensure the monotonicity of $g(x)$, we assume $\varrho \sim N(0, v_{0\varrho}^2)I(\delta\varrho \geq 0)$ with a default choice, we set $v_{0\varrho}^2 = 100$, similar to the case with $v_{0\theta_0}^2$.

Second, we exploit the shrinkage prior for sparse estimation of the regression coefficient vector $\boldsymbol{\beta}$ in the linear component of the FBSAQ in Equation (1). The class of continuous shrinkage priors has been used for sparse estimation as an alternative to variable selection with the spike-and-slab priors, including the Bayesian lasso and horseshoe priors (Park & Casella, 2008; Carvalho, Polson & Scott, 2010), are easy to implement and can produce results as good as or better than the spike-and-slab prior for handling sparsity and the shrinkage property. In particular, we adopt the HS+ prior considered by Bhadra et al. (2017) for the coefficient vector $\boldsymbol{\beta}$ in the linear component of the FBSAQ in Equation (1). The HS+ prior possesses an efficient shrinkage property (Bhadra et al., 2017) and is given by the hierarchical structure on the regression coefficient of the linear covariates,

$$\beta_j \sim N(0, \lambda_j^2), \quad \lambda_j \sim C^+(0, \varphi\eta_j), \quad \eta_j \sim C^+(0, 1), \quad j = 1, \ldots, q,$$

where $C^+(\mu, \phi)$ denotes the half Cauchy distribution with the location parameter μ and scale parameter ϕ. By integrating over η_j, the density function of λ_j is given by

$$p(\lambda_j|\varphi) = \frac{4}{\pi^2\varphi}\frac{\log(\lambda_j/\varphi)}{(\lambda_j/\varphi)^2 - 1}.$$

Following Bhadra et al. (2017), it is further assumed that $\varphi \sim C^+(0, 1/q)$. For the constant term, $\beta_0 \sim N(0, \lambda_0^2)$ is assumed with $\lambda_0^2 = 100$.

3. POSTERIOR INFERENCE

3.1. Posterior Computation

All the parameters are estimated using the MCMC method for posterior computation, based on the prior distributions specified in the previous section. The MCMC method alternately samples from the full conditional distributions of the parameters and latent variables. Specifically, the main MCMC components for the proposed GALDP implementations involved in the nonparametric mixture are sampled using the slice sampler proposed by Neal (2003), Walker (2007) and Kalli, Griffin & Walker (2011) to avoid computing the infinite sum in the stick-breaking representation of the Dirichlet process mixture. In the slice sampling approach, Equation (7) is augmented with a latent variable t as

$$f(\epsilon, t|G) = \sum_{k=1}^{\infty} I(t < \pi_k) f_{GAL_{p_0}}(\epsilon|\sigma_k, \gamma_k).$$

Given t, the number of components is finite:

$$f(\epsilon|t, G) \propto \sum_{k \in M_t} f_{GAL_{p_0}}(\epsilon|\sigma_k, \gamma_k),$$

where $M_t = \{k : \pi_k > t\}$.

We denote k_i as the index for the component membership for $i = 1, \ldots, n$. In the mixture representation of GAL distributions in Equation (6), A, B and C are also indexed by k or k_i for notational convenience. Moreover, we let k^* denote the minimum integer such that $\sum_{k=1}^{k^*} \pi_k > 1 - \min\{t_1, \ldots, t_n\}$. The mixing variables (σ, γ), $\boldsymbol{\theta}_J$, and ω are sampled after marginalizing the latent variables v_i and s_i following Rahman & Karnawat (2019). Thus, the joint posterior density of all the parameters and latent variables is proportional to

$$\prod_{i=1}^{n} I(t_i < \pi_{k_i}) \left[(\sigma_{k_i} v_i)^{-\frac{1}{2}} \exp\left\{ -\frac{(y_i - g_J(x_i) - \boldsymbol{w}_i' \boldsymbol{\beta} - C_{k_i} |\gamma_{k_i}| s_i - A_{k_i} v_i)^2}{2 B \sigma_{k_i} v_i} \right\} \right.$$
$$\left. \times \sigma_{k_i}^{-1} \exp\left\{ -v_i/\sigma_{k_i} \right\} \times \sigma_{k_i}^{-1} \exp\left\{ -\frac{s_i^2}{2\sigma_{k_i}^2} \right\} I(s_i > 0) \right] \times p(\Theta),$$

where $p(\Theta)$ denotes the prior distribution of all the parameters Θ given in Section 2. The detailed derivation of the full conditional distributions is provided in the Appendix, and the sampling procedures based on them are given below:

- Sampling t_i: Generate t_i from $U(0, \pi_{k_i})$ for $i = 1, \ldots, n$.
- Sampling z_k: Generate z_k from $Be(1 + n_k, n - \sum_{l < k} n_l + v)$ where $n_k = \sum_{i=1} I(k_i = k)$ for $k = 1, \ldots, k^*$.
- Sampling k_i: For $i = 1, \ldots, n$, generate k_i from the categorical distribution with the probabilities proportional to

$$P(k_i = k|\text{Rest}) \propto I(\pi_k > t_i) f_{GAL_{p_0}}(y_i - g_J(x_i) - \boldsymbol{w}_i' \boldsymbol{\beta} | \sigma_k, \gamma_k).$$

- Sampling v: We use the method described by Escobar & West (1995). By introducing $c \sim Be(v + 1, n)$, the full conditional distribution becomes the mixture of two gamma distributions given by

$$\varpi Ga(a_{0v} + n^*, b_{0v} - \log c) + (1 - \varpi) Ga(a_{0v} + n^* - 1, b_{0v} - \log c),$$

where n^* is the number of distinct components and $\varpi/(1 - \varpi) = (a_{0v} + n^* - 1)/(n(b_{0v} - \log c))$.

- Sampling $\boldsymbol{\theta}_J$: The full conditional distribution of $\boldsymbol{\theta}_J = (\theta_1, \ldots, \theta_J)'$ is proportional to

$$p(\boldsymbol{\theta}_J | \text{Rest}) \propto \left[\prod_{i=1}^{n} f_{GAL_{p_0}}\left(y_i - g_J(x_i) - \boldsymbol{w}_i' \boldsymbol{\beta} | \sigma_{k_i}, \gamma_{k_i}\right) \right]$$
$$\times f_N(\theta_0 | 0, v_{\theta_0}^2) I(\theta_0 \geq 0) \left[\prod_{j=1}^{J} f_N(\theta_j | 0, \tau^2 \exp(\rho j)) \right], \tag{10}$$

where g_J is a truncated series based on $\boldsymbol{\theta}_J$ and $f_N(\cdot; \mu, \phi^2)$ denotes the density for $N(\mu, \phi^2)$. Since the full conditional distribution in Equation (10) is not in a familiar form and we use moderately large J, we split $\boldsymbol{\theta}_J$ into blocks and apply the adaptive random walk Metropolis–Hastings (MH) algorithm (Lenk & Choi, 2017) to each block. In other words, given the current value of θ_j in the Markov chain , we use $N^+(\theta_0, s_0^2)$ for $j = 0$ and $N(\theta_j, s_j^2 \exp(-\rho j))$ for $j = 1, \ldots, J$ as proposal distributions. The values of s_j for $j = 0, \ldots, J$

are adapted as the algorithm proceeds in such a manner that the acceptance rate is between 0.2 and 0.3.

- Sampling (σ_k, γ_k): In Yan & Kottas (2017), the scale parameter σ of the GAL distribution was sampled from the generalized inverse Gaussian (GIG) distribution based on the mixture representation of the GAL distribution in Equation (6). However, in our experience, sampling from the GIG distribution with a large negative shape parameter can be computationally unstable. Therefore, we follow Rahman & Karnawat (2019) and sample σ_k and γ_k jointly by integrating out the latent s_i and v_i. The conditional distribution of (σ_k, γ_k) is proportional to

$$p(\sigma_k, \gamma_k | \text{Rest}) \propto \left[\prod_{i:k_i=k} f_{GALp_0} \left(y_i - g_J(x_i) - \mathbf{w}_i' \boldsymbol{\beta} | \sigma_k, \gamma_k \right) \right] p_0(\sigma_k, \gamma_k),$$

where $p(\sigma, \gamma)$ corresponds to the density of the base measure. We sample (σ_k, γ_k) by using the MH algorithm with the two independent normal proposal distributions truncated on the supports of σ and γ.

- Sampling v_i: For $i = 1, \ldots, n$, the full conditional distribution of v_i is proportional to

$$p(v_i | \text{Rest}) \propto f_N(y_i | g_J(x_i) - \mathbf{w}_i' \boldsymbol{\beta} - C_{k_i} |\gamma_{k_i}| s_i - A_{k_i} v_i, B_{k_i} \sigma_{k_i} v_i) \exp \left\{ -v_i / \sigma_{k_i} \right\}$$

$$\propto v_i^{-1/2} \exp \left\{ -\frac{1}{2} (\hat{b}_i^2 v_i^{-1} + \hat{c}_i^2 v_i) \right\}.$$

Therefore, the full conditional distribution is the GIG distribution $\mathcal{GIG}(0.5, \hat{b}_i^2, \hat{c}_i^2)$, where

$$\hat{b}_i^2 = \frac{(y_i - g_J(x_i) - \mathbf{w}_i' \boldsymbol{\beta} - C_{k_i} |\gamma_{k_i}| s_i)^2}{B_{k_i} \sigma_{k_i}}, \quad \hat{c}_i^2 = \frac{A_{k_i}^2}{B_{k_i} \sigma_{k_i}} + \frac{2}{\sigma_{k_i}}.$$

- Sampling s_i: For $i = 1, \ldots, n$, the full conditional distribution of s_i is given by $N^+(\hat{m}_i, \hat{v}_i^2)$, where

$$\hat{v}_i^2 = \left[\frac{(C_{k_i} |\gamma_{k_i}|)^2}{B_{k_i} \sigma_{k_i} v_i} + \frac{1}{\sigma_{k_i}^2} \right]^{-1}, \quad \hat{m}_i = \hat{v}_i^2 \left[\frac{C_{k_i} |\gamma_{k_i}| (y_i - g_J(x_i) - \mathbf{w}_i' \boldsymbol{\beta} - A_{k_i} v_i)}{B_{k_i} \sigma v_i} \right].$$

- Sampling $\boldsymbol{\beta}$: The full conditional distribution of $\boldsymbol{\beta}$ is the normal distribution $N(\hat{\boldsymbol{b}}, \hat{\mathbf{B}})$, where

$$\hat{\mathbf{B}} = \left[\sum_{i=1}^n \frac{\mathbf{w}_i \mathbf{w}_i'}{B_{k_i} \sigma_{k_i} v_i} + \mathbf{B}_0^{-1} \right]^{-1}, \quad \hat{\boldsymbol{b}} = \hat{\mathbf{B}} \left[\sum_{i=1}^n \frac{\mathbf{w}_i (y_i - g_J(x_i) - C_{k_i} |\gamma_{k_i}| s_i - A v_i)}{B_{k_i} \sigma_{k_i} v_i} \right],$$

and $\mathbf{B}_0 = \text{diag}(\lambda_0^2, \lambda_1^2, \ldots, \lambda_q^2)$.

- Sampling λ_j^2: For $j = 1, \ldots, q$, the full conditional distribution of λ_j is proportional to

$$p(\lambda_j | \text{Rest}) \propto f_N(\beta_j; 0, \lambda_j^2) p(\lambda_j | \varphi).$$

Since this is not in a standard form, λ_j is sampled using the random walk MH algorithm.

- **Sampling φ**: The full conditional distribution of φ is given by

$$p(\varphi|\text{Rest}) \propto \left[\prod_{j=1}^{q} p(\lambda_j|\varphi) \right] p(\varphi).$$

The random walk MH algorithm is used to sample from φ.

- Sampling ϱ: The full conditional distribution of ϱ is $N(\hat{a}, \hat{v}_{\varrho}^2)I(\delta\varrho \geq 0)$ where under FBSAQ with convexity

$$\hat{v}_{\varrho}^2 = \left[\sum_{i=1}^{n} \frac{\tilde{x}_i^2}{B_{k_i}\sigma_{k_i}v_i} + \frac{1}{v_{0\varrho}^2} \right]^{-1},$$

$$\hat{a} = \hat{v}_{\varrho}^2 \left[\sum_{i=1}^{n} \frac{\tilde{x}_i(y_i - w_i'\beta - \delta\theta_J'\Phi_J^s(x_i)\theta_J - C_{k_i}|\gamma_{k_i}|s_i - A_{k_i}v_i)}{B_{k_i}\sigma_{k_i}v_i} \right],$$

$\tilde{x}_i = x_i - 0.5$, and $\Phi_J^s(x)$. Here, $s \in \{b, c\}$ is a $(J+1) \times (J+1)$ matrix with the (j, k) element equal to $\varphi_{j,k}^s(x)$ (see Appendix).

- Sampling τ^2: The full conditional distribution of τ^2 is $IG(\hat{a}_\tau^2, \hat{b}_\tau^2)$, where

$$\hat{a}_{\tau^2} = a_{0\tau^2} + \frac{J}{2}, \quad \hat{b}_{\tau^2} = b_{0\tau^2} + \frac{1}{2} \sum_{j=1}^{J} \frac{\theta_j^2}{\exp(-\rho j)}.$$

- Sampling ρ: The full conditional distribution of ρ is proportional to

$$p(\rho|\text{Rest}) \propto \exp\left\{ -\rho\left(\frac{1}{e_{0\rho}} - \sum_{j=1}^{J} \frac{j}{2} \right) - \sum_{j=1}^{J} \frac{\theta_j \exp(\rho j)}{2\tau^2} \right\}.$$

By introducing J auxiliary variables $u_j \sim U(0, \exp\{-c_j \exp(\rho j)\})$ for $j = 1, \ldots, J$ where $c_j = \theta_j^2/2\tau^2$, ρ is sampled using slice sampling. See also Lenk (1999) and Lenk & Choi (2017). The joint distribution of $u_1, \ldots, u_J$ and ρ is proportional to

$$p(\rho, u_1, \ldots, u_J|\text{Rest}) \propto \left[\prod_{j=1}^{J} I(0 < u_j < \exp\{-c_j \exp(\rho j)\}) \right] \exp\{U_J\rho\},$$

where $U_J = \sum_{j=1}^{J} j/2 - 1/e_{0\rho}$. In slice sampling, u_j from $u_j \sim U(0, \exp\{-c_j \exp(\rho j)\})$ given ρ are sampled for $j = 1, \ldots, J$. Given $u_1, \ldots, u_J$, the conditional distribution of ρ is proportional to

$$p(\rho|u_1, \ldots, u_J, \text{Rest}) \propto \exp\left\{ U_J\rho \right\} I(0 < \rho < b),$$

where $b = \min_j \left\{ \log(-\log(u_j)/c_j)/j \right\}$. Then we sample ρ using

$$\rho = b + \log(u + (1 - u)\exp(-U_Jb))/U_J, \quad u \sim U(0, 1).$$

3.2. Shape Selection and the Choice of Error Distribution

Researchers typically believe, either based on theory or the findings of previous analyses, that the quantile curve has shape restrictions such as monotonicity and curvature. These restrictions can

be introduced through the proposed FBSAQ. Although a theoretically known or plausible shape constraint expressed in the prior distribution of the function is assumed, the actual observations often violate functional constraints because of measurement error or other reasons. Thus, when no theory or prior knowledge is available, researchers let the data determine the type of shape restriction in terms of model selection. We explore the feasibility of this approach to confirm the appropriate restrictions.

For example, based on the proposed FBSAQ models, there could be seven candidate models of the FBSAQ: no restriction, two monotone (increasing/decreasing) models, and four monotone convex/concave models. Of these seven models, we are uncertain about which shape is the most appropriate. Then, we select the appropriate shape by testing the adequacy of the shape restrictions in terms of a Bayesian model selection criteria. Specifically, we adopt two measures for model comparison that be easily computed from MCMC output: the LPML and the widely applicable information criterion (WAIC), also known as the Watanabe–Akaike information criterion. Let $\Theta^{(t)}$, for $t = 1, \ldots, T$, denote MCMC output for relevant parameters of the models for which the LPML or WAIC is computed.

The LPML is a predictive measure based on the cross-validated predictive densities that rest on conditional predictive ordinate (CPO) statistics (Geisser & Eddy, 1979). These statistics are computed from

$$\text{CPO}_i = \left[\frac{1}{T} \sum_{t=1}^{T} \frac{1}{f(y_i | \Theta^{(t)})} \right]^{-1}$$

for $i = 1, \ldots, n$ where $f(y_i | \Theta^{(t)})$ corresponds to the likelihood contribution of the ith observation. Then, the LPML is computed as

$$\text{LPML} = \sum_{i=1}^{n} \log \text{CPO}_i.$$

The WAIC provides a Bayesian approach to estimate the expected log predictive density (Watanabe, 2010; Gelman, Hwang & Vehtari, 2014). It is computed as

$$\text{WAIC} = -2 \sum_{i=1}^{n} \log \left[\frac{1}{T} \sum_{t=1}^{T} f(y_i | \Theta^{(t)}) \right] + 2 p_{\text{WAIC}},$$

where p_{WAIC} is the penalty term that adjusts for overfitting, which is computed as

$$p_{\text{WAIC}} = \sum_{i=1}^{n} \frac{1}{T} \sum_{t=1}^{T} \left(\log f(y_i | \Theta^{(t)}) - \frac{1}{T} \sum_{s=1}^{T} \log f(y_i | \Theta^{(s)}) \right)^2.$$

In addition, we consider Bayesian goodness of fit to check the adequacy of the underlying error distribution among the AL, GAL, and GALDP distributions by computing the LPML and WAIC as before. The standard measure for selecting Bayesian models, the marginal likelihood, is difficult to estimate for a semiparametric Bayesian model constructed under a Dirichlet process mixture prior (e.g., Mukhopadhyay & Gelfand, 1997) unless it is given a conjugate base measure as in Basu & Chib (2003), for example. Indeed, the LPML offers a more flexible approach to assessing models based on their predictive densities, and thus, it has frequently been used to assess models with a Dirichlet process mixture in practice (e.g., Jara, Nieto-Barajas & Quintana, 2013; Nieto-Barajas & Contreras-Cristán, 2014; Jo, Roh & Choi, 2016).

3.3. Extensions: Group-Specific Estimation and Censored Data

We consider extending the proposed FBSAQ to deal with two cases, one for group-specific estimation and the other for censored data, to meet the demand for quantile regression analysis in practice.

Group-specific estimation allows us to model nonlinear quantile curves for different subpopulations, which typically arise in cross-sectional, clustered, or longitudinal data. For this purpose, we express the basic model structure of Equation (1) as the group-specific representation

$$y_{i\ell} = g_\ell(x_i) + w'_{i\ell}\boldsymbol{\beta}_\ell + \epsilon_{i\ell}, \quad i = 1,\dots,n_\ell, \ \ell = 1,\dots,L, \tag{11}$$

where $\ell = 1,\dots,L$ denotes the number of groups for the subpopulations. $\boldsymbol{\beta}_\ell$ and $g_\ell(\cdot)$ denote the group-wise linear coefficient vector and nonlinear smooth functions, respectively. Thus, we impose shape restrictions on each group-specific function and represent the function as $g_\ell(x) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\theta_{j\ell}\theta_{k\ell}\tilde{\varphi}_{jk}(x)$. The error distribution $\epsilon_{i\ell}$ is assumed to be independent and modelled with the GALDP as before. The group-wise prior specifications of $\theta_{j\ell}$ and β_ℓ are made in a hierarchical way. Specifically, two-stage hierarchical prior distributions are assigned to g_ℓ and $\boldsymbol{\beta}_\ell$ in the group-wise specification. That is, for the monotonically increasing function g_ℓ, hierarchical smoothing prior distributions are assigned to the coefficients $\theta_{k\ell}$,

$$\theta_{j\ell} \mid \tilde{\theta}_j, \tau_i^2, \gamma_i \sim N\left(\tilde{\theta}_j, \tau_\ell^2 e^{-j\gamma_\ell}\right), \ j = 1,\dots,J, \ \ell = 1,\dots,L,$$

$$\tilde{\theta}_j \sim N\left(0, \omega_1/(j^3 + \omega_1)\right), \ j = 1,\dots,J,$$

$$\theta_{0\ell} \mid \tilde{\theta}_0 \sim N^+\left(\tilde{\theta}_0, v_{\theta_0}^2\right), \ \ell = 1,\dots,L,$$

$$\tilde{\theta}_0 \sim N^+(0,1),$$

$$\tau_\ell^2 \sim IG\left(\frac{r_\tau}{2}, \frac{s_\tau}{2}\right), \ \gamma_\ell \sim Exp\left(\omega_0\right), \ \ell = 1,\dots,L.$$

We can also use the hierarchical normal prior distribution for the linear coefficient vector $\boldsymbol{\beta}_\ell$, with $\boldsymbol{\beta}_\ell \mid \tilde{\boldsymbol{\beta}} \sim N_q\left(\tilde{\boldsymbol{\beta}}, V_\beta\right)$ and $\tilde{\boldsymbol{\beta}} \sim N_q\left(0, V_{\tilde{\beta}}\right)$ for $\ell = 1,\dots,L$. For posterior sampling with the MCMC method, we update the spectral coefficients $\theta_{j\ell}$, $j = 0,1,\dots,J$ for each ℓ using the adaptive random walk MH algorithm (Lenk & Choi, 2017). To update the hierarchical means $\tilde{\theta}_0$ and $\tilde{\theta}_j$, $j = 1,\dots,J$, we sample from

- $\left[\tilde{\theta}_0 \mid \text{others}\right] \sim N\left(m_{n,\tilde{\theta}_0}, v_{n,\tilde{\theta}_0}^2,\right)$, where $v_{n,\tilde{\theta}_0}^{-2} = 1 + Iv_{\theta_0}^{-2}$ and $m_{n,\tilde{\theta}_0} = v_{n,\tilde{\theta}_0}^2 \sum_{\ell=1}^{L}\theta_{0\ell}v_{\theta_0}^{-2}$ and
- $\left[\tilde{\theta}_j \mid \text{others}\right] \sim N\left(m_{n,\tilde{\theta}_j}, v_{n,\tilde{\theta}_j}^2,\right)$, $j = 1,\dots,J$, where $v_{n,\tilde{\theta}_j}^{-2} = (j^3 + \omega_1)/\omega_1 + \sum_{\ell=1}^{L}\tau_\ell^{-2}e^{j\gamma_\ell}$
  and $m_{n,\tilde{\theta}_j} = v_{n,\tilde{\theta}_j}^2 \sum_{\ell=1}^{L}\tau_\ell^{-2}e^{j\gamma_\ell}\theta_{j\ell}$.

To update $\boldsymbol{\beta}_\ell$, $\tilde{\boldsymbol{\beta}}$, and the two hyperparameters τ_ℓ^2 and γ_ℓ, we can easily sample from their full conditional distributions using Gibbs sampling and slice sampling, similar to the previous case without a group-specific specification.

In addition to the extension to the group-specific estimation, we deal with the quantile regression model for censored data (e.g., Portnoy, 2003), often called the *tobit* model, in which the response values of y in a given range are reported as a single value. Suppose that the response variable y_i is observed based on the following mechanism

$$y_i = \max\left\{y_i^*, y_0\right\}, \quad i = 1,\dots,n,$$

where y_0 is a known censoring point and y_i^* is the latent response variable. Without loss of generality, we proceed by assuming $y_0 = 0$. Subsequently, the quantile regression model for the latent response variable is given by

$$y_i^* = g(x_i) + w_i'\beta + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ_i follows the GAL distribution. From the equivalence of the monotone transformation of quantiles, the conditional quantile of the censored observation is given by

$$Q_{y|x,w}(p) = \max\left\{Q_{y^*|x,w}(p), 0\right\} = \max\left\{g(x_i) + w_i'\beta, 0\right\},$$

where $Q_{y^*|x,w}(p)$ is the conditional quantile of the latent variable. The MCMC sampling algorithm for the FBSAQ is also suitable for models with censored responses using the latent variables. See, also Kottas & Krnjajić (2009) for additional applications and discussions. Thus, the proposed FBSAQ models for a censored response can be estimated with a small modification to the algorithm described earlier. Specifically, to sample θ_J, the full conditional distribution in Equation (10) is replaced with

$$p(\theta_J|\text{Rest}) \propto \left[\prod_{i=1}^{n} f_{GALp_0}\left(y_i - g_J(x_i) - w_i'\beta; \sigma_{k_i}, \gamma_{k_i}\right)^{I(y_i > 0)}\right.$$

$$\times F_{GALp_0}(-g_J(x_i) - w_i'\beta; \sigma_{k_i}, \gamma_{k_i})^{I(y_i = 0)}\right]$$

$$\times f_N(\theta_0; 0, v_{\theta_0}^2) I(\theta_0 \geq 0) \left[\prod_{j=1}^{J} f_N(\theta_j; 0, \tau^2 \exp(\gamma j))\right].$$

The target densities for σ_k, γ_k, and ω are modified in the same manner. Furthermore, the additional data augmentation step for y_i^* is included in the MCMC algorithm (e.g., Kozumi & Kobayashi, 2011; Yan & Kottas, 2017). In the data augmentation step, y_i^* is sampled from

$$y_i^*|\text{Rest} \sim y_i I(y_i > 0) + N(g_J(x_i) + w_i'\beta + C_{k_i}|\gamma_{k_i}|s_i + Av_i, B_{k_i}\sigma_{k_i}v_i) I(y_i^* \leq 0, y_i = 0),$$

for $i = 1, \ldots, n$. Then the sampling steps for v_i, s_i, β, and ϱ proceed by replacing all y_i with y_i^*.

4. SIMULATION STUDIES

In this section, we present an empirical analysis using synthetic data with simulation studies and provide numerical evidence to illustrate the performance of the proposed FBSAQ. For this purpose, we test our proposed FBSAQ, the BSAQ under the GALDP and shape restrictions through simulation studies based on synthetic datasets generated from the semiparametric model

$$y_i = g^*(x_i) + w_i'\beta + e_i, \quad i = 1, \ldots, n,$$

where $n = 300$, $x_i = \Phi(w_{q+1,i})$, $(w_{1i}, \ldots, w_{qi}, w_{q+1,i})' \sim N(\mathbf{0}, \mathbf{S})$ with $q = 10$, $\mathbf{S}_{kk} = 1$, $\mathbf{S}_{kl} = 0.6^{|k-l|}$ and $\beta = (0, 2, 2, 2, 2, 0, 0, 0, 0, 0)'$. To check the adequacy of the GALDP for error distributions, the e_is are considered to be independent and identically distributed as (i) $N(0, 1)$, (ii) t_5 and (iii) $0.7N(0, 0.5^2) + 0.3N(0.5, 2^2)$. Furthermore, the following three shape-restricted settings for g are used for the simulation:

TABLE 1: Summary of the average RMSE for $\boldsymbol{\beta}$ for the unrestricted FBSAQ, FBSAQ-M and FBSAQ-MC models with the HS+ and normal priors for $\boldsymbol{\beta}$.

	ϵ_0	p_0	FBSAQ		FBSAQ-M		FBSAQ-MC	
			HS+	Normal	HS+	Normal	HS+	Normal
Setting 1	(i)	0.1	0.064	0.094	0.064	0.092	0.063	0.092
		0.5	0.066	0.095	0.067	0.095	0.067	0.096
		0.9	0.065	0.094	0.063	0.092	0.063	0.092
Setting 2	(ii)	0.1	0.072	0.107	0.071	0.106	0.071	0.106
		0.5	0.067	0.100	0.066	0.099	0.067	0.099
		0.9	0.070	0.106	0.070	0.105	0.071	0.106
Setting 3	(iii)	0.1	0.091	0.133	0.091	0.132	–	–
		0.5	0.096	0.141	0.097	0.141	–	–
		0.9	0.093	0.136	0.094	0.136	–	–

- Setting 1 (increasing convex): $g^*(x) = \exp(6x - 3) + 4$;
- Setting 2 (increasing concave): $g^*(x) = 10\log(10x + 1) - 8$; and
- Setting 3 (increasing): $g^*(x) = 80(x - 0.5)^3 + 6$.

While these g^* functions are not mean centred, the intercept in the linear term is adjusted such that the nonparametric function satisfies the mean-centring condition.

The following models are fit to the data depending on the shape restrictions. In Setting 1, the FBSAQ, the FBSAQ-M, namely, the monotone FBSAQ with $\delta = 1$, and the FBSAQ-MC, namely, the monotone and convex FBSAQ with $\delta = 1$, are used. In Setting 2, the FBSAQ, FBSAQ-M with $\delta = 1$, and FBSAQ-MC with $\delta = 1$ are used. In Setting 3, the FBSAQ, FBSAQ-M with $\delta = 1$ are used. To demonstrate the performance of the HS+ prior in the FBSAQ, we also consider the standard normal prior $\boldsymbol{\beta} \sim N(\mathbf{0}, 100\mathbf{I})$. The MCMC algorithm is run for 10,000 iterations after the initial 2,000 burn-in period and every fifth draw is retained for posterior inference. For $p_0 = 0.1$, 0.5, and 0.9, we compute the averages of the mean integrated squared error (MISE) and root mean squared error (RMSE) for $\boldsymbol{\beta}$ given by

$$\text{MISE} = \frac{1}{T}\sum_{t=1}^{T}\int(\widehat{g}^{(t)}(x) - g(x))^2 dx, \quad \text{RMSE} = \frac{1}{q}\sum_{j=1}^{q}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(\widehat{\beta}_j^{(t)} - \beta_j)^2},$$

where $g(x)$ denotes the mean-centred $g^*(x)$, $\widehat{g}^{(t)}(x)$ denotes its estimate for the tth replication of the data, and $\widehat{\beta}_j^{(t)}$ denotes the posterior mean of β_j for the tth replication of the data. The performance of the models is compared using the MISE and RMSE based on $T = 100$ replications of the data. In addition, the LPML and WAIC are computed to compare the models.

Table 1 presents the RMSE for $\boldsymbol{\beta}$ with the two priors for $\boldsymbol{\beta}$, averaged over the $T = 100$ replications under Settings 1, 2 and 3, combined with the error distributions (i), (ii) and (iii). As shown in Table 1, the model with the HS+ prior always results in a smaller RMSE than that with the normal prior, demonstrating the advantage of adopting the HS+ prior for the shrinkage property with sparse estimation of regression coefficients in quantile regression. The choice of

TABLE 2: Summary of the average MISE for g for the unrestricted FBSAQ, FBSAQ-M and FBSAQ-MC models with the GALDP or GAL error distribution.

			GALDP			GAL		
	ϵ	p_0	FBSAQ	FBSAQ-M	FBSAQ-MC	FBSAQ	FBSAQ-M	FBSAQ-MC
Setting 1	(i)	0.1	0.085	0.029	0.024	0.084	0.030	0.025
		0.5	0.088	0.031	0.025	0.087	0.031	0.024
		0.9	0.079	0.027	0.022	0.078	0.027	0.022
Setting 2	(ii)	0.1	0.086	0.041	0.031	0.086	0.042	0.032
		0.5	0.087	0.041	0.034	0.087	0.040	0.034
		0.9	0.094	0.043	0.037	0.094	0.043	0.038
Setting 3	(iii)	0.1	0.051	0.021	–	0.056	0.023	–
		0.5	0.043	0.017	–	0.044	0.017	–
		0.9	0.071	0.032	–	0.083	0.040	–

shape restriction does not seem to influence the RMSE for β given the results summarized in Table 1.

Table 2 presents the average MISE for g for the unrestricted FBSAQ, FBSAQ-M, and FBSAQ-MC models with the GALDP and HS+ prior for the settings presented in Table 2. For comparison, we also present the results under the same models with the parametric GAL distribution instead of the GALDP, particularly for Setting 3. In this setting, Table 2 shows that the average MISEs in the GALDP appears to be relatively smaller than those in the GAL, especially for the unrestricted model with $p_0 = 0.1$ and 0.9. In estimating the nonparametric function $g(\cdot)$, the shape-restricted models perform better, resulting in a smaller MISE than in the model without such restrictions.

Table 3 presents averages of the LPML in the case of $p_0 = 0.5$. The LPML is comparable between the GALDP and GAL specifications for the error distributions (i) and (iii). However, we observe a larger LPML under the GALDP than under the GAL in the case of the error distribution (iii). The LPML for the shape-restricted models was greater than that for the unrestricted model. In addition, among the shape-restricted models, those with more compelling shape restrictions have a larger LPML. We further report the number of times the model is selected out of 100 replications of data with the largest LPML and the smallest WAIC under the proposed GALDP specification and HS+ prior. As summarized in parentheses in Table 3, these results agree with the MISE results for the nonparametric function g presented in Table 2; that is, the models with the lowest MISE are chosen the most often based on the LPML. The FBSAQ without restriction is selected no more than 6 times out of 100 in our simulation, with an average LPML below that of the shape-restricted models. Moreover, it is never selected in Settings 1 and 2 with the error distribution (ii). In Setting 1, the FBSAQ-MC is selected about 70% of the time under error distributions (i) and (ii) and about 60% of the times under error distribution (iii). The FBSAQ-M is selected between 30% and 40% of the time. In Setting 2, the FBSAQ-MC is selected about 60% of the time followed by the FBSAQ-M. In Setting 3, FBSAQ-M was selected about 95% of the time. Although not shown here, we have also obtained the same results for WAIC and other quantiles.

The next simulation study we consider is for the censored version with the same simulation settings as before. That is, the latent response variables are generated from the model given by

TABLE 3: Summary of the average LPML and the number of times the model is selected (in parentheses) for the FBSAQ, FBSAQ-M and FBSAQ-MC models with the proposed GALDP and HS+ prior for $p_0 = 0.5$. For comparison, the results for the FBSAQ models with the GAL distribution are also presented.

		GALDP			GAL		
	ϵ	FBSAQ	FBSAQ-M	FBSAQ-MC	FBSAQ	FBSAQ-M	FBSAQ-MC
Setting 1	(i)	-452.4 (4)	−444.8 (28)	−444.5 (68)	−451.6	−444.4	−444.1
	(ii)	−505.1 (0)	−498.7 (27)	−498.0 (73)	−505.1	−498.5	−497.8
	(iii)	−441.6 (0)	−430.8 (38)	−430.0 (62)	−443.9	−435.3	−434.6
Setting 2	(i)	−450.7 (2)	−445.7 (37)	−445.1 (61)	−460.7	−445.4	−444.6
	(ii)	−503.8 (0)	−499.0 (41)	−498.9 (59)	−503.8	−499.0	−498.6
	(iii)	−552.9 (1)	−433.0 (38)	−432.4 (61)	−442.6	−436.1	−435.4
Setting 3	(i)	−450.1 (6)	−445.4 (94)	–	−449.7	−445.0	–
	(ii)	−503.5 (5)	−498.8 (95)	–	−503.4	−498.6	–
	(iii)	−440.0 (4)	−433.2 (96)	–	−442.4	−435.8	–

TABLE 4: Summary of the average MISE for g and RMSE for $\boldsymbol{\beta}$ with $p_0 = 0.5$.

			FBSAQ		FBSAQ-M	
	ϵ	Prior	MISE	RMSE	MISE	RMSE
Setting 3	(iii)	HS+	0.198	0.104	0.116	0.106
		Normal	0.199	0.156	0.111	0.156
			LPML	WAIC	LPML	WAIC
		HS+	−456.6 (0.302)	912.8 (0.296)	−453.9 (0.698)	907.5 (0.704)

Equation (3.3) and the response variables are observed based on $y_i = \max\{y_i^*, 0\}$. The same nonlinear functions, linear component, error distributions, and prior distributions are used with $p_0 = 0.5$. The data are replicated 100 times, with the average censoring rates around 0.2. Table 4 presents the results of the simulations averaged over the $T = 100$ replications, showing the average MISE and RMSE with two different priors for $\boldsymbol{\beta}$ for $p_0 = 0.5$ in Setting 3 combined with the error distribution (iii). It also shows averages of the LPML and WAIC under the HS+ prior and the number of time the model results in the highest LPML and the lowest WAIC, in parentheses, for $p_0 = 0.5$.

As shown in Table 4, the HS+ prior outperforms the normal prior in terms of the RMSE of $\boldsymbol{\beta}$. In addition, the choice of the prior distribution for the linear term does not affect the MISE of the unknown nonlinear function, and the choice of the shape restriction does not affect the RMSE. In terms of the MISE, the shape-restricted model (i.e., FBSAQ-M) also performs better than the model without restrictions (i.e., the FBSAQ). Table 4 also presents averages of the LPML and WAIC under the HS+ prior and the number of times the model results in the highest LPML and the lowest WAIC, in parentheses, for $p_0 = 0.5$. As indicated in the previous simulation and the MISE in Table 4, the FBSAQ-M is favoured over the FBSAQ about 70% of the time, both in terms of LPML and WAIC.
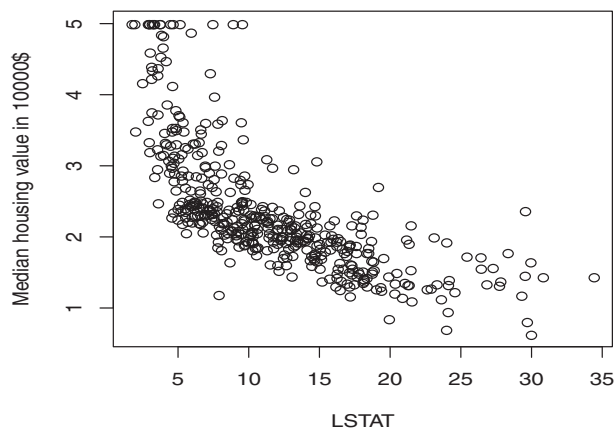
FIGURE 2: Scatter plot of the median housing value in 10,000\$ (response variable) and LSTAT
(covariate with a nonlinear effect) in the Boston housing price data ($n = 506$).

5. REAL DATA APPLICATION

5.1. Boston Housing Data

In this section, to illustrate the HS+ prior in the case of many covariates in the linear component,
we consider the Boston housing data of Harrison & Rubinfeld (1978). The dataset includes
information on the 506 houses in the Boston Standard Metropolitan Statistical Area and the char-
acteristics of their neighbourhoods. The response variable is the median value of owner-occupied
houses in \$10,000, which is assumed to be nonlinearly influenced by the percentage of lower
status individuals in the population (LSTAT). The housing price is expected to monotonically
decrease as the LSTAT proportion increases. Indeed, by looking at the scatter plot of the housing
price and LSTAT in Figure 2, we can observe a nonincreasing and convex relationship between
the variables.

Therefore, we fit the FBSAQ, FBSAQ-M with $\delta = -1$, and FBSAQ-MC with $\delta = -1$. In the
linear component, we include the per capita crime rate by town (CRIM), proportion of residential
land zoned for lots over 25,000 square feet (ZN), proportion of nonretail business acres per town
(INDUS), a dummy variable equal to one if a tract is bounded by the Charles River (CHAS),
nitric oxides concentration (NOX), average number of rooms per dwelling (RM), proportion of
owner-occupied units built before 1940 (AGE), weighted distances to five Boston employment
centres (DIS), index of accessibility to radial highways (RAD), full-value property-tax rate per
\$10,000 (TAX), pupil-teacher ratio by town (PTRATIO) and index on the proportion of blacks
by town (B).

Table 5 presents the LPML and WAIC for the three models. Both the LPML and the WAIC
favour the FBSAQ with shape restrictions over the unrestricted FBSAQ, although there are
subtle differences in the selected model. Based on LPML, we see that the selected model is the
FBSAQ-MC for $p_0 = 0.1$ and 0.5 and the FBSAQ-M for $p_0 = 0.9$, whereas the FBSAQ-MC is
selected for $p_0 = 0, 1$ and the FBSAQ-M for $p_0 = 5$ and $p_0 = 0.9$.

Figure 3 presents posterior means and 95% credible intervals with the parametric residuals
under the unrestricted FBSAQ model and the selected models based on LPML. Figure 3 shows
that the shape-restricted models produced smoother estimates for $g(\text{LSTAT})$ than the unrestricted
FBSAQ model. The credible intervals under the unrestricted FBSAQ model also become wider
than those under the restricted models. Under the selected shape-restricted models, the housing

TABLE 5: The LPML and WAIC for the proposed FBSAQ models fit to the Boston housing data, with and without shape restrictions for $p_0 = 0.1$, 0.5 and 0.9.

p_0		FBSAQ	FBSAQ-M	FBSAQ-MC
0.1	LPML	−145.372	−146.720	−145.074
	WAIC	267.207	266.896	269.161
0.5	LPML	−152.533	−152.825	−152.416
	WAIC	292.286	290.435	290.724
0.9	LPML	−178.887	−171.532	−173.597
	WAIC	343.330	331.249	332.623
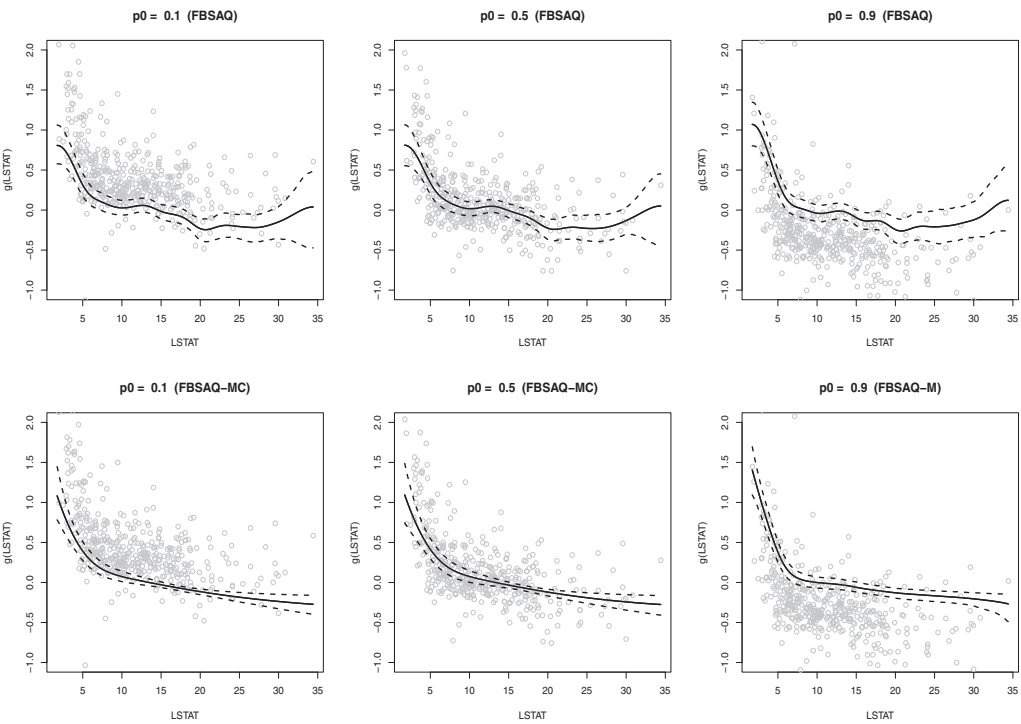
FIGURE 3: The posterior means and 95% credible intervals for $g(\text{LSTAT})$ under the unrestricted FBSAQ (top row) and the selected FBSAQ models with shape restrictions (bottom row) for $p_0 = 0.1$, 0.5 and 0.9 fit to the Boston housing data. The gray circles represent parametric residuals.

price drops sharply in the region of LSTAT < 5, especially for $p_0 = 0.9$, and subsequently shows a gradual decline as LSTAT increases for all the quantiles.

Such a nonlinear effect cannot be captured if a linear quantile regression model is used. Contrary to the shape-restricted models, the estimate under the unrestricted FBSAQ fails to capture the sharp decay in the housing price for LSTAT < 5 and the housing price appears to be slightly increasing for LSTAT > 25.

TABLE 6: The posterior means and 95% credible intervals (in parentheses) for β under the selected FBSAQ models with the given quantiles $p_0 = 0.1, 0.5$ and 0.9 fit to the Boston housing data.

Variable	FBSAQ-MC ($p_0 = 0.1$)	FBSAQ-MC ($p_0 = 0.5$)	FBSAQ-M ($p_0 = 0.9$)
CRIM	−0.090 (−0.117, −0.060)	−0.087 (−0.115, −0.056)	−0.092 (−0.123, −0.059)
ZN	0.001 (−0.001, 0.002)	0.001 (−0.001, 0.002)	0.001 (−0.001, 0.002)
INDUS	−0.000 (−0.005, 0.005)	0.000 (−0.005, 0.006)	−0.001 (−0.008, 0.005)
CHAS	0.053 (−0.015, 0.167)	0.046 (−0.012, 0.170)	0.025 (−0.026, 0.128)
NOX	−0.080 (−0.545, 0.080)	−0.045 (−0.425, 0.095)	−0.020 (−0.288, 0.121)
RM	0.549 (0.455, 0.640)	0.542 (0.443, 0.637)	0.660 (0.580, 0.743)
AGE	−0.003 (−0.005, −0.001)	−0.003 (−0.005, −0.001)	−0.003 (−0.005, −0.001)
DIS	−0.070 (−0.093, −0.045)	−0.065 (−0.090, −0.042)	−0.075 (−0.099, −0.051)
RAD	0.032 (0.021, 0.043)	0.031 (0.019, 0.041)	0.032 (0.020, 0.044)
TAX	−0.001 (−0.001, −0.001)	−0.001 (−0.001, −0.001)	−0.001 (−0.001, −0.000)
PTRATIO	−0.066 (−0.081, −0.051)	−0.063 (−0.078, −0.049)	−0.070 (−0.087, −0.054)
B	0.001 (0.001, 0.001)	0.001 (0.001, 0.002)	0.001 (0.001, 0.002)

Table 6 presents posterior means and 95% credible intervals for the parameters in the linear component. Under the HS+ prior, the credible intervals for ZN, INDUS, CHAS and NOX include zero for all the quantiles. For the rest of the significant covariates, the signs of the coefficients are the same for all the quantiles. Covariates such as AGE, DIS, TAX and PTRATIO are estimated to be decreasing factors for the housing price and their magnitudes of impact do not vary across the quantiles. The positive effect of B is significant and also does not vary across quantiles, but its magnitude is small. The posterior means for RM are 0.549, 0.542 and 0.660 for $p_0 = 0.1, 0.5$ and 0.9, respectively. This finding suggests that the value increased by approximately \$5,500 with every room per dwelling for the lower and mid quantiles and by \$6,600 for the higher quantiles.

5.2. Married Female Labour data

The censored versions of the FBSAQ models discussed in Section 3.3 are illustrated by using the working hour data analyzed by Mroz (1987) for 753 married women in 1975. The response variable is annual working hours in 1,000 hours (HOURS) and the data include 325 responses with zero working hours treated as left censored. Working hours are assumed to be nonlinearly influenced by a woman's years of labour market experience (EXPER). The linear component includes household income not earned by the wife (NWIFEINC), years of education (EDUC), age and number of children under 6 years (KIDSLT6) and over 6 years old (KIDSGE6). Kozumi & Kobayashi (2011) and Yan & Kottas (2017) analyzed the data by using the linear quantile regression model where quadratic term for EXPER was included. We therefore replace the quadratic function of EXPER with a nonparametric function. Since it is natural to assume that working hours would not increase dramatically after a certain point of a woman's career because of age, we fit the following four models: FBSAQ, FBSAQ-M with $\delta = 1$, FBSAQ-MC with $\delta = 1$ and FBSAQ-U with $\delta = 1$. The HS+ prior is used for the parameters in the linear component. Table 7 presents the LPML and WAIC for the five models for $p_0 = 0.1, 0.5$ and 0.9. Both the LPML and the WAIC support the FBSAQ-U for $p_0 = 0.1$ and 0.5 and the FBSAQ-MC for $p_0 = 0.9$.

TABLE 7: The LPML and WAIC for the proposed FBSAQ models fit to the female labour data ($n = 753$), with and without shape restrictions for $p_0 = 0.1$, 0.5 and 0.9.

p_0		FBSAQ	FBSAQ-M	FBSAQ-MC	FBSAQ-U
0.1	LPML	−892.073	−888.304	−887.782	−887.505
	WAIC	1784.048	1776.356	1775.191	1774.552
0.5	LPML	−872.882	−868.555	−869.592	−868.060
	WAIC	1745.318	1735.860	1738.102	1735.131
0.9	LPML	−873.432	−879.247	−871.260	−879.537
	WAIC	1742.911	1749.949	1739.884	1748.515
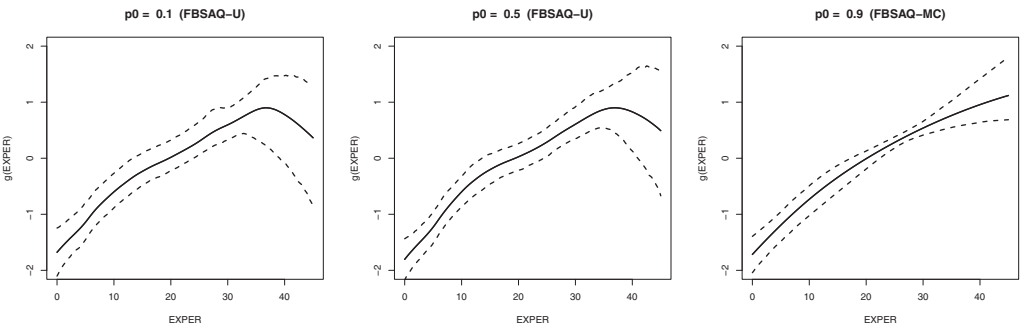
FIGURE 4: The posterior means and 95% credible intervals for g(EXPER) under the FBSAQ-U model for $p_0 = 0.1$ and 0.5 and the FBSAQ-MC model for $p_0 = 0.9$, fit to the female labour data.

Figure 4 presents posterior means and 95% credible intervals under the selected FBSAQ-U and FBSAQ-MC models. For $p_0 = 0.1$ and 0.5, the posterior means of ω for the FBSAQ-U model are approximately 0.8. For the lower and middle quantiles, the conditional quantile of working hours is continuously increasing up to about 36 years of experience in the labour market, after which working hours start to decline, possibly because of aging and retirement. Based on this result, working hours continuously increase within the range of the observed years of work experience for the upper quantiles.

Table 8 presents posterior means and 95% credible intervals for the parameters in the linear term under the selected shape-restricted models for $p_0 = 0.1$, 0.5 and 0.9. Under the HS+ prior, the credible intervals for KIDSGE6 include zero for all the quantiles.

Nonwife household income has a negative effect on working hours with the posterior mean of the coefficient equal to −0.011 for $p_0 = 0.1$, while the credible intervals include zero for $p_0 = 0.5$ and 0.9. As with the previous results, a wife from a household with higher nonwife income tends to work for a shorter time. This effect is slightly greater for $p_0 = 0.1$, for which the posterior mean of the coefficient is −0.011, than for $p_0 = 0.5$ and 0.9, for which the posterior means are −0.005 and −0.003, respectively. The age and presence of small children have negative effects for all quantiles. The posterior means of AGE are −0.057, −0.051 and −0.050 for $p_0 = 0.1$, 0.5 and 0.9, respectively. Hence, the impact of the wife age is similar across the quantiles. Weekly working hours reduce by nearly one hour for every year older a wife is. The presence of small children requires the wife to spend time on childcare at home. The posterior means for KIDSLT6 are −0.921, −0.865 and −0.891 for $p_0 = 0.1$, 0.5 and 0.9, respectively. This effect

TABLE 8: Posterior means and 95% credible intervals for β under the selected FBSAQ models fit to the female labour data.

Variable	FBSAQ-U ($p_0 = 0.1$)	FBSAQ-U ($p_0 = 0.5$)	FBSAQ-MC ($p_0 = 0.9$)
NWIFEINC	−0.011 (−0.020, −0.002)	−0.005 (−0.013, 0.003)	−0.003 (−0.012, 0.004)
EDUC	0.092 (0.047, 0.138)	0.054 (0.013, 0.099)	0.049 (−0.000, 0.098)
AGE	−0.057 (−0.071, −0.042)	−0.051 (−0.064, −0.038)	−0.050 (−0.064, −0.037)
KIDSLT6	−0.921 (−1.177, −0.686)	−0.865 (−1.071, −0.658)	−0.891 (−1.133, −0.676)
KIDSGE6	−0.017 (−0.083, 0.039)	−0.009 (−0.081, 0.059)	−0.004 (−0.068, 0.048)

of the number of small children is, again, slightly greater for $p_0 = 0.1$. For a wife with median working hours, having an additional small child would reduce weekly working hours by about 16.6 hours. A longer year of education leads to longer working hours for all quantiles and its magnitude increases as the quantile level rises.

5.3. Japanese Universities Cost Data

The last real application we consider is to Japanese university cost data to illustrate group-specific curve estimation using the hierarchical Bayesian approach. These data comprise financial and managerial information on universities across Japan. Specifically, they include observations of 80 national universities in Japan between 2004 and 2015. The response variable is the log of education and research-related cost, explained by several covariates that characterize each university, such as total number of students, wages of academic staff, and type of universities and others. The covariate with a nonlinear effect is the log total number of undergraduate students (LNUNDER). It is reasonable to assume that the university cost has a nonlinear relationship and may increase with the number of undergraduate students. All the other covariates beside LNUNDER are assumed to have linear effects, which are the log number of graduate students (LNGRAD), log wages of academic staff (LWAC), log wage of nonacademic staff (LWNAC), year index (YEAR) and its square (YEAR^2), and dummy variables for the type of university (universities without hospitals (NOHOS), schools of science (SCI), schools of arts (ARTS), medical schools (MED) and schools of education (EDU)). The baseline type includes universities with hospitals. The information on costs, wages and number of students is obtained from annual financial and business reports that are publicly available.

We apply the proposed FBSAQ model with three types of shapes for g(LNSTU) and GALDP, the FBSAQ, FBSAQ-M ($\delta = 1$) and FBSAQ-MC ($\delta = 1$). The models are estimated based on every 20th draw of 40,000 MCMC draws after a burn-in period of 10,000 draws. Table 9 presents the LPML and WAIC for the proposed models. For $p_0 = 0.1$ and 0.5, the FBSAQ-M model results in the largest LPML and the smallest WAIC. For $p_0 = 0.9$, the unrestricted FBSAQ model is favoured over the shape-restricted FBSAQ models.

Figure 5 presents posterior means and 95% credible intervals of g(LNUNDER) for $p_0 = 0.1$, 0.5 and 0.9 under the three types of shape restrictions. Figure 5 shows that the estimated function without a shape restriction (FBSAQ) can be unstable, exhibiting a wide credible interval. As we restrict the shape of the function, the estimates become more stable and smoother. Since the estimates under the unrestricted g(LNUNDER) are overall increasing as expected, by forcing g(LNUNDER) to be monotonically increasing, the resulting quantile curve estimate can be stabilized, especially for $p = 0.5$. Furthermore, further restricting g(LNUNDER) to be monotone convex results in a smooth estimate of this function. However, forcing monotone convexity

TABLE 9: LPML and WAIC for FBSAQ, FBSAQ-M and FBSAQ-MC models fit to the Japanese university data.

p_0		FBSAQ	FBSAQ-M	FBSAQ-MC
0.1	LPML	−141.995	−139.797	−145.061
	WAIC	283.548	278.899	289.195
0.5	LPML	−171.290	−161.767	−176.046
	WAIC	340.100	323.219	349.767
0.9	LPML	−142.278	−145.596	−174.857
	WAIC	284.235	288.266	347.045

might oversimplify the underlying structure and we might fail to capture an important aspect of this function. Thus, the nonlinear relationship between university cost and number of students is better identified by imposing monotone restrictions on g(LNUNDER), as also supported by the model selection criteria in Table 9.

Further, we consider group-specific estimation of the quantile curves by applying the group-specific model in Equation (11) to the Japanese university data as

$$y_{it} = g_t(\text{LNUNDER}_i) + w'_{it}\beta_t + \epsilon_{it}, \quad i = 1, \dots, n_t, \ t = 1, \dots, 12,$$

where $t = 1, \dots, 12$ denotes the academic years from 2004 to 2015 as the subpopulations. In the group-specific FBSAQ (GS FBSAQ) representation, $g_t(\cdot)$ is assumed to be either monotonically increasing or unrestricted, and the proposed GALDP model is used for the error distribution. As shown in Table 10, the GS FBSAQ-M is selected as the best model with the largest LPML and the smallest WAIC between two models, as expected.

As an illustration, Figure 6 shows posterior means and 95% credible intervals for g_t (LNUNDER$_i$) for the academic years 2004 and 2015 ($t = 1$ and 12) for $p_0 = 0.5$. Figure 6, under GS FBSAQ-M, the increasing patterns of cost in the number of students exhibit a clear difference between the two academic years. In 2004, the cost increases sharply up to about 1,000 students while cost changes very little between 1,000 and 8,000 students and it increases again above 8,000 students. In 2015, we observe that the cost increases without the flat region and it sharply rises above 8,000 students. Contrary, under the GS FBSAQ model without shape restriction, while the model seem to also suggest that the cost rises as the number of students increases, it fails to capture the striking differences in the increasing pattern of the costs among the academic years.

These results contrast with the overall quantile curves in Figure 5. When the basic model in Equation (1) is used, the time effect with different years is not accounted for, and the overall nonlinear relationship between the university cost and number of students can be interpreted as the merging of divergent yearly quantile curves. Thus, a single overall curve can be misleading, as it does not consider the functional interaction effects among academic years. Using the hierarchical Bayesian formulation, we fit their individual curves but combine related information across observations in the 12 academic years by identifying dissimilarity and exploiting possible similarities simultaneously.

Finally, Figure 7 presents posterior means and 95% credible intervals for the selected parameters, LNGRAD, LNWAC, NOHOS and EDU, in the linear component of GS FBSAQ-M for $p_0 = 0.5$ and 0.9 plotted between the years 2004 and 2015 ($t = 1, \dots, 12$). Figure 7 shows
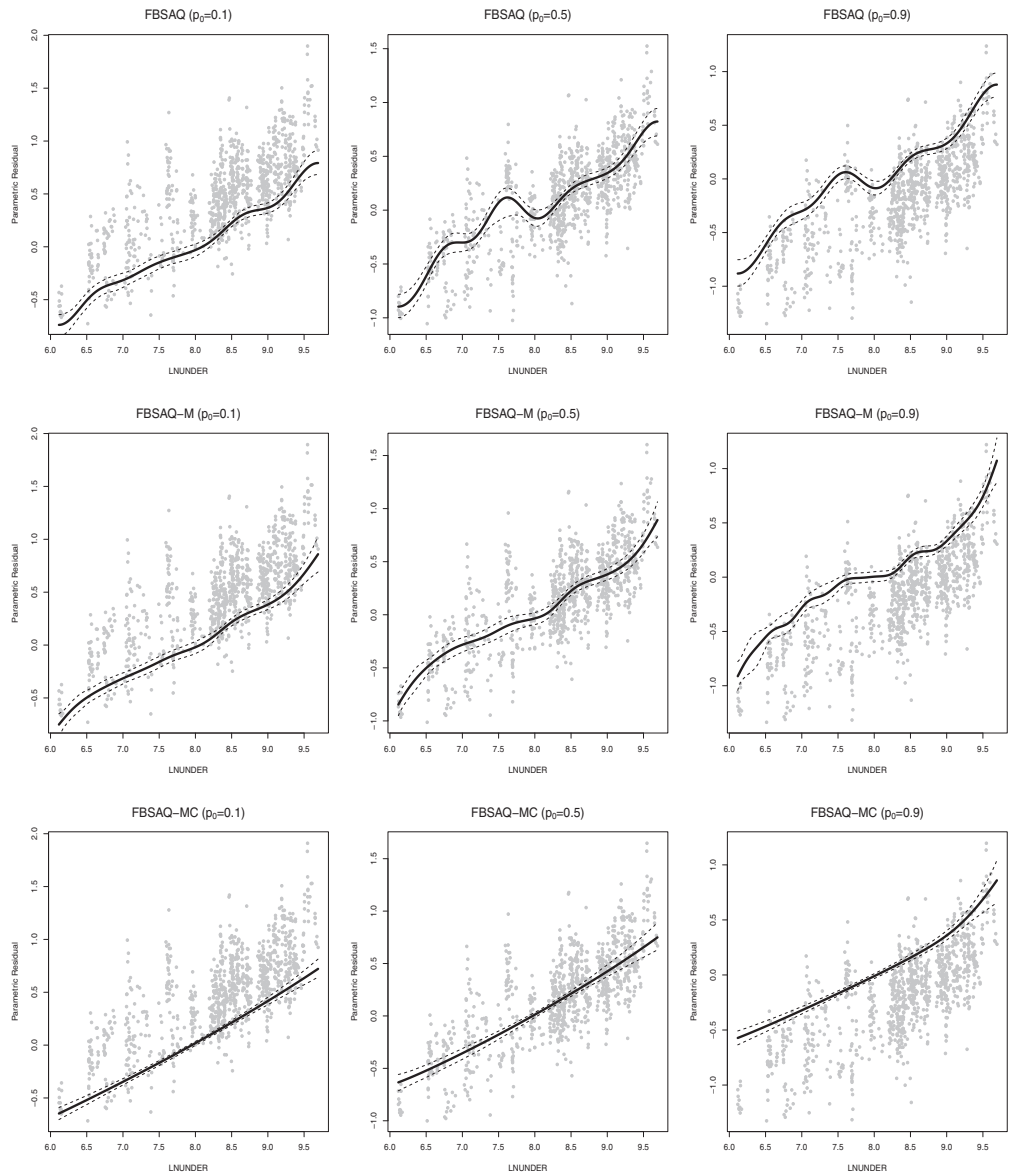
FIGURE 5: Posterior means and 95% credible intervals of g(LNSTU) for conditional quantiles, $p_0 = 0.1, 0.5$ and 0.9 with the FBSAQ (top panels), FBSAQ-M (middle panels) and FBSAQ-MC (bottom panels) models fit to the Japanese university data.

that the changing patterns and locations of the posterior distributions for those parameters are similar between $p_0 = 0.5$ and 0.9, especially for LNGRAD, LNWAC and NOHOS.

In the case of LNGRAD, the posterior distributions for $p_0 = 0.5$ are located around 0.4 throughout the observed period except in the year 2015 where the estimate indicates a drop in the positive effect of number of graduate students on cost. For $p_0 = 0.9$, while Figure 7 shows some fluctuation around the year 2008, we also obtained a drop in 2015 after the flat period around 0.4. The credible intervals for LNWAC include zero up to and including 2012 for

TABLE 10: The LPML and WAIC for the GS FBSAQ and GS FBSAQ-M models fit to the Japanese university data.

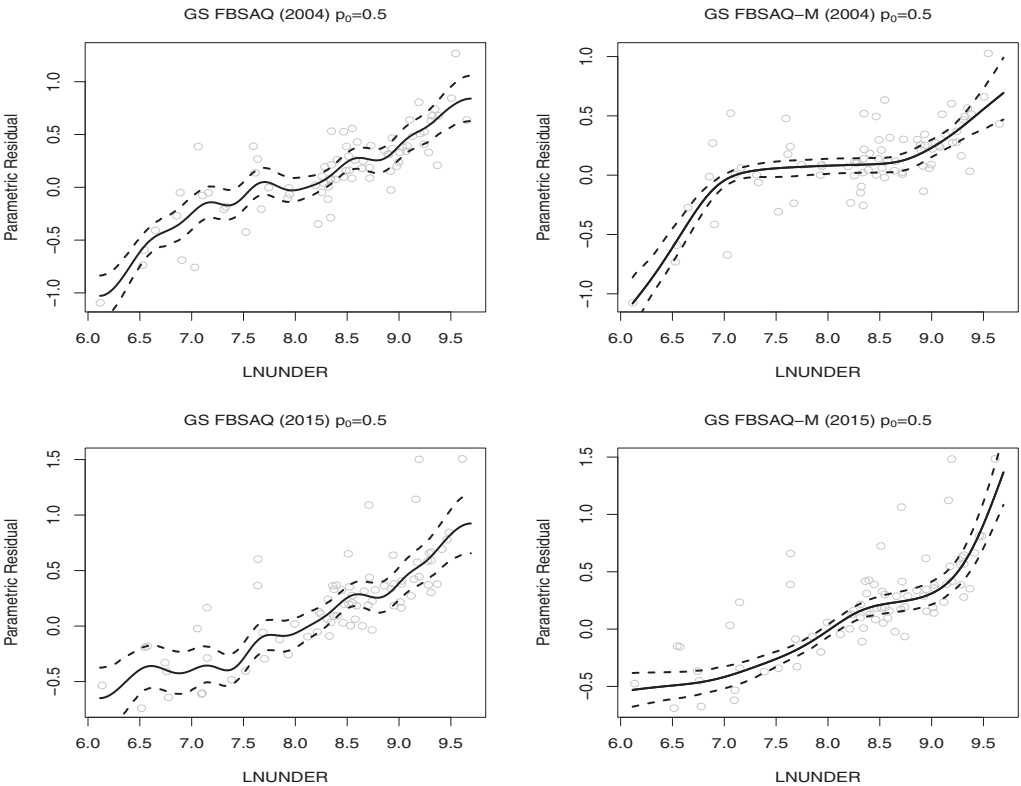|	GS FBSAQ			GS FBSAQ-M		
p_0	0.1	0.5	0.9	0.1	0.5	0.9
LPML	−12.472	−16.366	−456.510	42.588	67.287	−335.462
WAIC	18.595	29.626	713.124	−88.515	−135.357	497.028

FIGURE 6: Posterior means of group-specific g_t(LNUNDER) for the academic year 2004 and 2015 with the GS FBSAQ and GS FBSAQ-M models for $p = 0.5$, fit to the Japanese university data.

both $p_0 = 0.5$ and 0.9. After the year 2013 the credible intervals do not include zero with the positive posterior means for $p_0 = 0.9$ while for $p_0 = 0.5$ the posterior means are also positive but the credible intervals continue to include zero. Figure 7 shows that the NOHOS is credibly negative throughout the observed period for $p_0 = 0.5$. Since the baseline group for the type dummy variables consists of universities with attached hospitals and associated faculties and departments, which would be factors increasing education-and research-related cost, the cost for the universities without one would be naturally smaller when other factors are fixed. For $p_0 = 0.9$, the posterior distributions appear to be slightly shifted to the right in Figure 7 and,
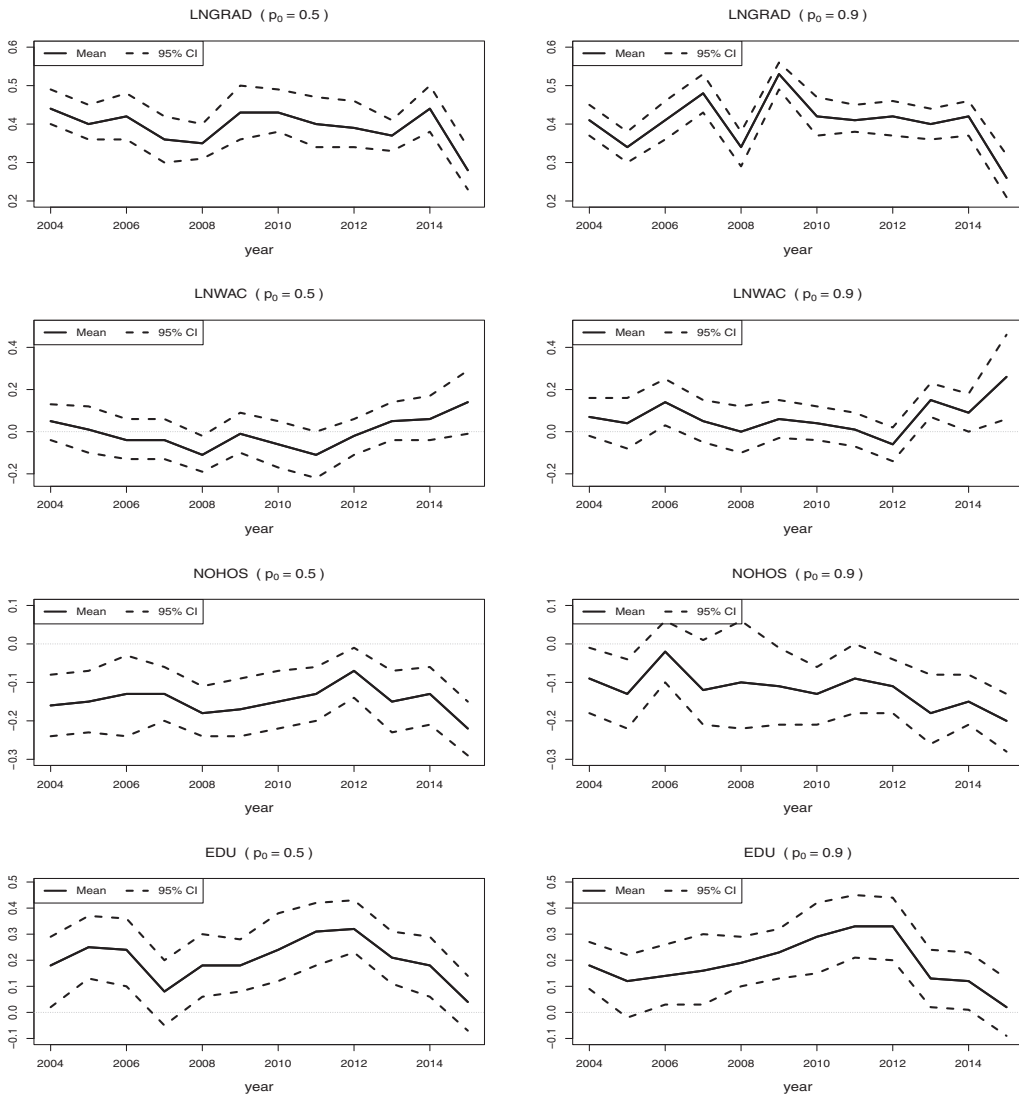
FIGURE 7: Posterior means and 95% credible intervals for the selected parameters in the linear component of the GS FBSAQ-M model fit to the Japanese university data for $p_0 = 0.5$ and 0.9 for $t = 1, \ldots, 12$.

while the posterior means are negative, the credible intervals include zero in some years at the start of the observation period. Lastly, Figure 7 shows that the positive effect of EDU increases toward the year 2012 and declines toward the end of the observation period for both $p_0 = 0.5$ and 0.9. The positive effect of EDU might be counterintuitive at the first glance. The cost for the universities in this group also include that for the attached schools, typically such as elementary and junior high schools. For one of the universities in this group, education-related cost for the attached schools amounts to one-third of total educational cost. Therefore, it is a natural result that the education-related cost of a university in this group is higher than one in the baseline group when other factors are fixed.

6. CONCLUSION

In this article, we proposed a flexible Bayesian nonlinear quantile semiparametric regression model based on the GALDP model with shape restrictions for the nonlinear function and the HS+ prior for the linear component. The numerical examples showed that the use of the nonparametric mixture of GAL distributions with Dirichlet process priors and the introduction of an appropriate shape restriction lead to more flexible and accurate estimates than those based on the parametric GAL distribution and without shape restrictions. The choice of the shape restriction should be based on underlying theory or prior knowledge in context. Otherwise, one should let the data determine the type of shape restriction in terms of model selection. We also demonstrated the efficiency of the HS+ prior in shrinking parameters in the linear component. In addition, the proposed models were extended to cases with censored data and group-specific quantile curve estimation. Using the hierarchical Bayesian formulation allowed us to estimate multiple quantile curves to identify divergent quantile curves for different subpopulations characterized by various features in clustered or cross-sectional data analysis.

Our approach to introducing shape restrictions could be extended to many other important quantile models such as mixed effects models, survival analysis, and functional data analysis for quantiles. These extensions are left for future studies. For example, to explain the Japanese university cost data further, we could consider a model including positive random effects in the framework of stochastic frontier models to study quantile-specific cost frontiers and efficiency.

A limitation of the present model would be that it estimates one quantile at a time and there is no built-in mechanism to prevent the estimated quantile function from violating monotonicity. After estimating the quantile function, we could use the adjustment method proposed by Rodriguez & Fan (2017) to correct for monotonicity. Moreover, given important developments in simultaneous quantile regression, such as Das & Ghosal (2018), Yang & Tokdar (2017) and Rodrigues, Dortet-Bernadet & Fan (2019), incorporating shape restrictions into simultaneous quantile regression context would be another interesting direction for future study.

APPENDIX

Shape Restrictions for $g(x)$ and its Basis Expansion

The proposed FBSAQ models deal with a number of shape restrictions for $g(x)$ based on the signs of the first and second derivatives of the function g as the square of a Gaussian process Z, i.e., $g^{(q)}(x) = \delta Z^2(x)$ for $\delta = 1$ or -1 and $q = 1$ or 2:

	$\delta = 1$	$\delta = -1$
$q = 1$	increasing $g(x)$	decreasing $g(x)$
$q = 2$	convex $g(x)$	concave $g(x)$

As shown in Lenk & Choi (2017), the analytical forms for $g(x)$ are available for the monotone and monotone concave/convex restrictions. Under the monotone restriction,

$$\varphi_{00}^{a}(x) = x - 0.5,$$

$$\varphi_{0j}^{a}(x) = \varphi_{j0}^{a} = \frac{2}{\pi j} \sin(\pi j x) - \frac{\sqrt{2}}{(\pi j)^2}[1 - \cos(\pi j)], \quad j \geq 1,$$

$$\varphi_{jj}^{a}(x) = \frac{\sin(2\pi j x)}{2\pi j} + x - 0.5, \quad j \geq 1,$$

$$\varphi_{jk}^a(x) = \frac{\sin(\pi(j+k)x)}{\pi(j+k)} + \frac{\sin(\pi(j-k)x)}{\pi(j-k)}$$

$$-\frac{1-\cos(\pi(j+k))}{[\pi(j+k)]^2} - \frac{1-\cos(\pi(j-k))}{[\pi(j-k)]^2}, \quad j \neq k, \quad j, k \geq 1.$$

By truncating the infinite series at J, the model can be expressed as $g_J(x) = \delta\theta_J'\Phi_J^a(x)\theta_J$ where $\Phi_J^a(x)$ is a $(J+1)\times(J+1)$ matrix with the (j,k)th element equal to $\varphi_{jk}^a(x)$.

The monotone function g with convexity or concavity with the same sign for the first and second derivatives is given by

$$g(x) = \delta \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j\theta_k\varphi_{jk}^b(x) + \varrho(x-0.5),$$

where the second term corresponds to the constant of integration to meet the mean-centring condition and

$$\varphi_{00}^b(x) = \frac{3x^2-1}{6},$$

$$\varphi_{0j}^b(x) = \varphi_{j0}^b = -\frac{\sqrt{2}}{(\pi j)^2}\cos(\pi jx), \quad j \geq 1,$$

$$\varphi_{jj}^b(x) = -\frac{\cos(2\pi jx)}{(2\pi j)^2} + \frac{3x^2-1}{6}, \quad j \geq 1,$$

$$\varphi_{jk}^b(x) = \frac{\cos(\pi(j+k)x}{(\pi(j+k))^2} - \frac{\cos(\pi(j-k)x}{(\pi(j-k))^2}, \quad j \neq k, \quad j, k \geq 1.$$

When the first and second derivatives have opposite signs, the model is given by

$$g(x) = \delta \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j\theta_k\varphi_{jk}^c(x) + \varrho(x-0.5),$$

where $\varphi_{jk}^c(x) = \varphi_{jk}^b(1-x)$. By truncating the infinite series at J, the model can be expressed as $g_J(x) = \delta\theta_J'\Phi_J^s(x)\theta_J + \varrho(x-0.5)$, where $\Phi_J^s(x)$ is a $(J+1)\times(J+1)$ matrix with the (j,k)th element equal to $\varphi_{jk}^s(x)$ and $s \in \{b,c\}$.

Joint Posterior and Full Conditional Distributions

Here the joint posterior and the full conditional distributions in some of the Gibbs steps given the mixture component membership k_i are described. Given the component allocation in the Dirichlet process mixture, recall that the model is given by

$$y_i = g(x_i) + w_i'\beta + \gamma_{k_i}s_i + A_{k_i}v_i + \sqrt{\sigma_{k_i}B_{k_i}v_i}u_i, \quad i = 1, \ldots, m,$$

$$s_i \sim N^+(0, \sigma_{k_i}^2), \quad v_i \sim Exp(\sigma_{k_i}), \quad u_i \sim N(0,1).$$

The joint posterior density is given by Equation (6). With the latent variables s_i and v_i integrated out, the joint posterior is simply proportional to

$$\prod_{i=1}^{n} I(t_i < \pi_{k_i}) \left[f_{GALp_0}\left(y_i - g_J(x_i) - w_i'\beta | \sigma_{k_i}, \gamma_{k_i} \right) \right] p(\Theta),$$

on which the sampling of θ_J, (σ_k, γ_{k_i}) and ω is based.

From Equation (6), the full conditional distribution of v_i is given by

$$p(v_i|\text{Rest}) \propto v_i^{-\frac{1}{2}} \exp\left\{ -\frac{(y_i - g_J(x_i) - w_i'\beta - C_{k_i}|\gamma_{k_i}|s_i - A_{k_i}v_i)^2}{2B_{k_i}\sigma_{k_i}v_i} \right\} \exp\left\{ -v_i/\sigma_{k_i} \right\}$$

$$\propto v_i^{-1/2} \exp\left\{ -\frac{1}{2}\left(\frac{(y_i - g_J(x_i) - w_i'\beta - C_{k_i}|\gamma_{k_i}|s_i)^2}{B_{k_i}\sigma v_i} + \left(\frac{A_{k_i}^2}{B_{k_i}\sigma_{k_i}} + \frac{2}{\sigma_{k_i}} \right)v_i \right) \right\}$$

$$\propto v_i^{-1/2} \exp\left\{ -\frac{1}{2}(\hat{b}_i^2 v_i^{-1} + \hat{c}_i^2 v_i) \right\},$$

which is the kernel of the GIG.

The full conditional distribution of s_i is given by

$$p(s_i|\text{Rest}) \propto \exp\left\{ -\frac{(y_i - g_J(x_i) - w_i'\beta - C_{k_i}|\gamma_{k_i}|s_i - A_{k_i}v_i)^2}{2B_{k_i}\sigma_{k_i}v_i} \right\} \exp\left\{ -\frac{s_i^2}{2\sigma_{k_i}^2} \right\} I(s_i > 0)$$

$$\propto \exp\left\{ \frac{(s_i - \hat{m}_i)^2}{\hat{v}_i^2} \right\} I(s_i > 0).$$

The full conditional distribution of λ_j is given by

$$p(\lambda_j|\text{Rest}) \propto f_N(\beta_j; 0, \lambda_j^2)p(\lambda_j|\varphi)$$

$$\propto \lambda_j^{-1} \exp\left\{ -\frac{\beta_j^2}{2\lambda_j^2} \right\} \frac{\log(\lambda_j/\varphi)}{(\lambda_j/\varphi)^2 - 1}.$$

Since this is not in a standard form, the MH algorithm is used to sample λ_j.

The full conditional distribution of φ is given by

$$p(\varphi|\text{Rest}) \propto \prod_{j=1}^{q} \left[p(\lambda_j|\varphi) \right] p(\varphi) \propto \prod_{j=1}^{q} \left[\frac{1}{\varphi} \frac{\log(\lambda_j/\varphi)}{(\lambda_j/\varphi)^2 - 1} \right] \frac{1}{1 + q^2\varphi^2} I(\varphi > 0).$$

In the case of the censored response model, by augmenting the censored responses the joint posterior is proportional to

$$\prod_{i=1}^{n} \left[I(t_i < \pi_{k_i})(\sigma_{k_i}v_i)^{-\frac{1}{2}} \exp\left\{ -\frac{(y_i^* - g_J(x_i) - w_i'\beta - C_{k_i}|\gamma_{k_i}|s_i - Av_i)^2}{2B_{k_i}\sigma_{k_i}v_i} \right\} \right.$$

$$\times \left(I(y_i^* \le 0, y_i = 0) + I(y_i^* = y_i, y_i > 0) \right)$$

$$\left. \times \sigma_{k_i}^{-1} \exp\left\{ -v_i/\sigma_{k_i} \right\} \sigma^{-1} \exp\left\{ -\frac{s_i^2}{2\sigma_{k_i}^2} \right\} I(s_i > 0) \right] \times p(\Theta),$$

The joint posterior of the parameters after integrating out s_i and v_i is given by

$$\prod_{i=1}^{n} \left[I(t_i < \pi_{k_i}) f_{GALp_0} \left(y_i - g_J(x_i) - \boldsymbol{w}_i' \boldsymbol{\beta}; \sigma_{k_i}, \gamma_{k_i} \right)^{I(y_i>0)} \right. $$
$$\left. \times F_{GALp_0}(-g_J(x_i) - \boldsymbol{w}_i' \boldsymbol{\beta}; \sigma_{k_i}, \gamma_{k_i})^{I(y_i=0)} \right] \times p(\Theta).$$

ACKNOWLEDGEMENTS

BIBLIOGRAPHY

Alhamzawi, R., Yu, K., & Benoit, D. F. (2012). Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, 12, 279–297.

Basu, S. & Chib, S. (2003). Marginal likelihood and Bayes factors for Dirichlet process mixture models. *Journal of the American Statistical Association*, 98, 224–235.

Bhadra, A., Datta, J., Polson, N. G., & Willard, B. (2017). The horseshoe+ estimator of ultra-sparse signals. *Bayesian Analysis*, 12, 1105–1131.

Carvalho, C. M., Polson, N. G., & Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97, 465–480.

Chen, C. & Yu, K. (2009). Automatic Bayesian quantile regression curve fitting. *Statistics and Computing*, 19, 271–281.

Choi, T., Lee, J., & Roy, A. (2009). A note on the Bayes factor in a semiparametric regression model. *Journal of Multivariate Analysis*, 100, 1316–1327.

Das, P. & Ghosal, S. (2018). Bayesian non-parametric simultaneous quantile regression for complete and grid data. *Computational Statistics and Data Analysis*, 127, 172–186.

Davino, C., Furno, M., & Vistocco, D. (2014). *Quantile Regression. Wiley Series in Probability and Statistics*. John Wiley & Sons, Ltd., Chichester.

Escobar, M. D. & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577–588.

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

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24, 997–1016.

Harrison, D. & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102.

Hu, Y., Zhao, K., & Lian, H. (2015). Bayesian quantile regression for partially linear additive models. *Statistics and Computing*, 25, 651–668.

Jara, A., Nieto-Barajas, L. E., & Quintana, F. (2013). A time series model for responses on the unit interval. *Bayesian Analysis*, 8, 723–740.

Jo, S., Choi, T., Park, B., & Lenk, P. J. (2019). bsamGP: An R package for Bayesian spectral analysis models using Gaussian process priors. *Journal of Statistical Software*, 90, 1–41.

Jo, S., Roh, T., & Choi, T. (2016). Bayesian spectral analysis models for quantile regression with Dirichlet process mixtures. *Journal of Nonparametric Statistics*, 28, 177–206.

Kalli, M., Griffin, J. E., & Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, 21, 93–105.

Kobayashi, G. (2017). Bayesian endogenous tobit quantile regression. *Bayesian Analysis*, 12, 161–191.

Kobayashi, G. & Kozumi, H. (2012). Bayesian analysis of quantile regression for censored dynamic panel data. *Computational Statistics*, 27, 359–380.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. (2019). *quantreg: Quantile Regression*. R package version 5.51.

Koenker, R. & Bassett, G. Jr. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

Kottas, A. & Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36, 297–319.

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81, 1565–1578.

Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a Fourier representation. *Journal of the Royal Statistical Society Series B—Statistical Methodology*, 61, 863–879.

Lenk, P. J. & Choi, T. (2017). Bayesian analysis of shape-restricted functions using Gaussian process priors. *Statistica Sinica*, 27, 43–69.

Li, Q., Xi, R., & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5, 533–556.

Mroz, T. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55, 765–799.

Mukhopadhyay, S. & Gelfand, A. E. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association*, 92, 633–639.

Naranjo, L., Pérez, C. J., & Martín, J. (2015). Bayesian analysis of some models that use the asymmetric exponential power distribution. *Statistics and Computing*, 25, 497–514.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31, 705–767.

Nieto-Barajas, L. E. & Contreras-Cristán, A. (2014). A Bayesian nonparametric approach for time series clustering. *Bayesian Analysis*, 9, 147–170.

Park, T. & Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103, 681–686.

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98, 1001–1012.

Rahman, M. A. & Karnawat, S. (2019). Flexible Bayesian quantile regression in ordinal models. *Advances in Econometrics*, 40B, 211–251.

Reich, B. J., Bondell, H. D., & Wang, H. J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, 11, 337–352.

Reich, B. J. & Smith, L. B. (2013). Bayesian quantile regression for censored data. *Biometrics*, 69, 651–660.

Rodrigues, T., Dortet-Bernadet, J. -L., & Fan, Y. (2019). Pyramid quantile regression. *Journal of Computational and Graphical Statistics*, 28, 732–746.

Rodriguez, T. & Fan, Y. (2017). Regression adjustment for noncrossing Bayesian quantile regression. *Journal of Computational and Graphical Statistics*, 26, 275–284.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639–650.

Sriram, K., Ramamoorthi, R. V., & Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Analysis*, 8, 479–504.

Taddy, M. A. & Kottas, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *Journal of Business and Economic Statistics*, 28, 357–369.

Thompson, P., Cai, Y., Moyeed, R., Reeve, D., & Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics & Data Analysis*, 54, 1138–1150.

Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., & Flexeder, C. (2013). Bayesian semiparametric additive quantile regression. *Statistical Modelling*, 13, 223–252.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics. Simulation and Computation*, 36, 45–54.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11, 3571–3594.

Wichitaksorn, N., Choy, S. T. B., & Gerlach, R. (2014). A generalized class of skew distributions and associated robust quantile regression models. *The Canadian Journal of Statistics*, 42, 579–596.

Yan, Y. & Kottas, A. (2017). *A New Family of Error Distributions for Bayesian Quantile Regression*. ArXiv e-prints, arXiv:1701.05666v2.

Yang, P. & Tokdar, S. T. (2017). Joint estimation of quantile planes over arbitrary predictor spaces. *Journal of the American Statistical Association*, 112, 1107–1120.

Yu, K. & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54, 437–447.

Yu, K. & Stander, J. (2007). Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics*, 137, 260–276.

Yu, K. & Zhang, J. (2005). A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics. Theory and Methods*, 34, 1867–1879.

Yue, Y. R. & Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis*, 55, 84–96.