

Quantile Regression and Beyond in Statistical Analysis of the Data

Lead Guest Editor: Rahim Alhamzawi

Guest Editors: Keming Yu and Himel Mallick





Quantile Regression and Beyond in Statistical Analysis of the Data

Journal of Probability and Statistics

Quantile Regression and Beyond in Statistical Analysis of the Data

Lead Guest Editor: Rahim Alhamzawi

Guest Editors: Keming Yu and Himel Mallick



Copyright © 2019 Hindawi. All rights reserved.

This is a special issue published in “Journal of Probability and Statistics.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

Elio Chiodo, Italy

Hyungjun Cho, Republic of Korea

Alessandro De Gregorio, Italy

Junbin Gao, Australia

Luis A. Gil-Alana, Spain

Heinz Holling, Germany

Yaozhong Hu, Canada

Anna Karczewska, Poland

Dejian Lai, USA

Chin-Shang Li, USA

Chunsheng Ma, USA

Marek T. Malinowski, Poland

Edwin Ortega, Brazil

Zacharias Psaradakis, UK

Ramón M. Rodríguez-Dagnino, Mexico

Jose M. Sarabia, Spain

Steve Su, Australia

Dongchu Sun, USA


Man Lai Tang, Hong Kong

Aera Thavaneswaran, Canada

Rongling Wu, USA

Contents

Quantile Regression and Beyond in Statistical Analysis of Data

Rahim Alhamzawi , Keming Yu, and Himel Mallick 

Editorial (1 page), Article ID 2635306, Volume 2019 (2019)

Conditional Analysis for Mixed Covariates, with Application to Feed Intake of Lactating Sows

S. Y. Park , C. Li , S. M. Mendoza Benavides, E. van Heugten, and A. M. Staicu

Research Article (14 pages), Article ID 3743762, Volume 2019 (2019)

Fully Bayesian Estimation of Simultaneous Regression Quantiles under Asymmetric Laplace Distribution Specification

Josephine Merhi Bleik 

Research Article (12 pages), Article ID 8610723, Volume 2019 (2019)

New Link Functions for Distribution-Specific Quantile Regression Based on Vector Generalized Linear and Additive Models

V. F. Miranda-Soberanis  and T. W. Yee 

Research Article (11 pages), Article ID 3493628, Volume 2019 (2019)

Group Identification and Variable Selection in Quantile Regression

Ali Alkenani  and Basim Shlaibah Msallam

Research Article (7 pages), Article ID 8504174, Volume 2019 (2019)

Retirement Consumption Puzzle in Malaysia: Evidence from Bayesian Quantile Regression Model

Ros Idayuwati Alaudin, Noriszura Ismail , and Zaidi Isa

Research Article (8 pages), Article ID 2723069, Volume 2019 (2019)

Editorial

Quantile Regression and Beyond in Statistical Analysis of Data

Rahim Alhamzawi ¹, Keming Yu,² and Himel Mallick ³

¹Department of Statistics, University of Al-Qadisiyah, Al-Qadisiyah, Iraq

²Brunel University, Uxbridge, UK

³Harvard University, Boston, USA

Correspondence should be addressed to Rahim Alhamzawi; ralhamzawi@yahoo.com

Received 11 June 2019; Accepted 11 June 2019; Published 22 July 2019

Copyright © 2019 Rahim Alhamzawi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Regression is used to quantify the relationship between response variables and some covariates of interest. Standard mean regression has been one of the most applied statistical methods for many decades. It aims to estimate the conditional expectation of the response variable given the covariates. However, quantile regression is desired if conditional quantile functions such as median regression are of interest. Quantile regression has emerged as a useful supplement to standard mean regression. Also, unlike mean regression, quantile regression is robust to outliers in observations and makes very minimal assumptions on the error distribution and thus is able to accommodate nonnormal errors. The value of “going beyond the standard mean regression” has been illustrated in many scientific subjects including economics, ecology, education, finance, survival analysis, microarray study, growth charts, and so on. In addition, inference on quantiles can accommodate transformation of the outcome of the interest without the problems encountered in standard mean regression. Overall, quantile regression offers a more complete statistical model than standard mean regression and now has widespread applications.

There has been a great deal of recent interest in Bayesian approaches to quantile regression models and the applications of these models. In these approaches, uncertain parameters are assigned prior distributions based on expert judgment and updated using observations through the Bayes formula to obtain posterior probability distributions. In this special issue on “Quantile regression and beyond in statistical analysis of data,” we have invited a few papers that address such issues. The first paper of this special issue addresses a fully Bayesian approach that estimates multiple quantile

levels simultaneously in one step by using the asymmetric Laplace distribution for the errors, which can be viewed as a mixture of an exponential and a scaled normal distribution. This method enables characterizing the likelihood function by all quantile levels of interest using the relation between two distinct quantile levels. The second paper presents a new link function for distribution-specific quantile regression based on vector generalized linear and additive models to directly model specified quantile levels. The third paper presents a novel modeling approach to study the effect of predictors of various types on the conditional distribution of the response variable. The fourth paper introduces the regularized quantile regression method using pairwise absolute clustering and sparsity penalty, extending from mean regression to quantile regression setting. The final paper of this special issue uses Bayesian quantile regression for studying the retirement consumption puzzle, which is defined as the drop in consumption upon retirement, using the cross-sectional data of the Malaysian Household Expenditure Survey 2009/2010.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Rahim Alhamzawi
Keming Yu
Himel Mallick

Research Article

Conditional Analysis for Mixed Covariates, with Application to Feed Intake of Lactating Sows

S. Y. Park ¹, C. Li ², S. M. Mendoza Benavides,³ E. van Heugten,³ and A. M. Staicu²

¹Eli Lilly and Company, Indianapolis, IN 46285, USA

²Department of Statistics, North Carolina State University, Raleigh, NC 27695, USA

³Department of Animal Science, North Carolina State University, Raleigh, NC 27695, USA

Correspondence should be addressed to S. Y. Park; parksoyoung3859@gmail.com and C. Li; cli9@ncsu.edu

Received 5 December 2018; Revised 28 April 2019; Accepted 28 May 2019; Published 16 July 2019

Guest Editor: Rahim Alhamzawi

Copyright © 2019 S. Y. Park et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a novel modeling framework to study the effect of covariates of various types on the conditional distribution of the response. The methodology accommodates flexible model structure, allows for joint estimation of the quantiles at all levels, and provides a computationally efficient estimation algorithm. Extensive numerical investigation confirms good performance of the proposed method. The methodology is motivated by and applied to a lactating sow study, where the primary interest is to understand how the dynamic change of minute-by-minute temperature in the farrowing rooms within a day (functional covariate) is associated with low quantiles of feed intake of lactating sows, while accounting for other sow-specific information (vector covariate).

1. Introduction

Many modern applications routinely collect data on study participants comprising scalar responses and covariates of various types, vector, function, and image, and the main question of interest is to examine how the covariates affect the response. For example, in our motivating experimental study, the goal is to analyze how the minute-by-minute daily temperature and humidity of the farrowing rooms, where sows are placed after giving birth for nursing, affect their feed intake during a lactation period. The covariates consist of temperature profile, humidity, and sow age, where the response is a total amount of daily feed intake of sows. A popular approach in these cases is to use a nonparametric framework and assume that the mixed covariates solely affect the mean response; see Cardot et al. [1], James [2], Ramsay and Silverman [3, 4], Ferraty and Vieu [5, 6], Goldsmith et al. [7], McLean et al. [8], and others. However, for our application, while it is important to study the mean feed intake, animal scientists are often more concerned with the left tail of the feed intake distribution. This is because low feed intake of lactating sows could lead to many serious issues, including decrease in milk production and negative

impact on the sows reproductive system; see, for reference, Quiniou and Noblet [9], Renaudeau and Noblet [10], St-Pierre et al. [11], among others. In this paper we focus on regression models that study the effects of covariates on the entire distribution of response. Our contribution is the development of a modeling framework that accommodates a comprehensive study of various types (vector and functional) of covariates on a scalar response.

Quantile regression models the effect of scalar/vector covariates beyond the mean response; it provides a more comprehensive study of the covariates on the response and has attracted great interest [12, 13]. For prespecified quantile levels, quantile regression models the conditional quantiles of the response as a function of the observed covariates; this approach has been extended more recently to ensure noncrossing of quantile functions [14]. Quantile regression has been also extended to handle functional covariates. Cardot et al. [15] discussed quantile regression models by employing a smoothing spline modeling based approach. Kato [16] considered the same problem and used a functional principal component (fPC) based approach. Both papers mainly discussed the case of having a single functional covariate and it is not clear how to extend them to the

case where there are multiple functional covariates or mixed covariates (vector and functional).

More recently, Tang and Cheng [17], Lu et al. [18], and Yu et al. [19] studied quantile regression when the covariates are of mixed types and introduced the partial functional linear quantile regression modeling framework. The first two publications used fPC basis while the last one considered partial quantile regression (PQR) basis. These approaches all are suitable when the interest is studying the effect of covariate at a particular quantile level and do not handle the study of covariate effects at simultaneous quantile levels due to the well-known crossing-issue.

Ferraty et al. [20] and Chen and Müller [21] considered a different perspective and studied the effect of a functional predictor on the quantiles of the response by modeling the conditional distribution of the response directly. However their approach is limited to one functional predictor. In this paper we fill this gap and propose a unifying modeling framework and estimation technique that allows studying the effect of mixed type covariates (i.e., scalar, vector, and functional) on the conditional distribution of a scalar response in a computationally efficient manner.

Let $Q_{Y|X(\cdot)}(\tau)$ denote the τ th conditional quantile of Y given a functional covariate $X(\cdot)$, and let $F_{Y|X(\cdot)}(y)$ denote the conditional distribution of Y given $X(\cdot)$. We model the conditional distribution using a generalized function-on-function regression framework, i.e., $F_{Y|X(\cdot)}(y) = E\{\mathbb{1}(Y \leq y) \mid X(\cdot)\} = g^{-1}\{\int X(t)\beta(t, y)dt\}$, where $\mathbb{1}(\cdot)$ is an indicator function and g is the logit link function, and we study the conditional quantiles by exploiting the relationship between $Q_{Y|X(\cdot)}(\tau)$ and $F_{Y|X(\cdot)}(y)$ through $Q_{Y|X(\cdot)}(\tau) = \inf\{y : F_{Y|X(\cdot)}(y) \geq \tau\}$ for $0 < \tau < 1$. The advantage and contribution of our proposed method mainly come from the following reasons: (1) our modeling approach is spline-based, and as a result it can easily accommodate smooth effects of scalar variables as well as of functional covariates and (2) our estimation approach is based on a single step function-on-function (or function-on-scalar) penalized regression, which enables efficient implementation by exploiting off-the-shelf software and leads to competitive computations.

The remainder of the paper is structured as follows. Section 2 discusses the details of the proposed method and Section 3 describes the estimation procedure and extensions. Section 4 performs a thorough simulation study evaluating the performance of the proposed method and its competitors. We apply the proposed method to analyze the sow data in Section 5. We conclude the paper with a discussion in Section 6.

2. Methodology

2.1. Statistical Framework. Let i be index subjects, j index repeated measurements, n the number of subjects, and m_i the number of observations for subject i . Suppose we observe $\{Y_i, \mathbf{X}_{i1}, X_{i2}, \{t_{ij}, X_{i3}(t_{ij})\}_{1 \leq j \leq m_i}\}_{1 \leq i \leq n}$, where Y_i is the response, which is a scalar random variable, \mathbf{X}_{i1} is a p -dimensional

vector of nuisance covariates, X_{i2} is a scalar covariate, and $X_{i3}(\cdot)$ is a functional covariate, which is assumed to be square-integrable on a closed domain \mathcal{T} .

We propose the following model for the conditional distribution of Y_i given \mathbf{X}_{i1} , X_{i2} , and $X_{i3}(\cdot)$:

$$\begin{aligned} F_{Y_i|\mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)}(y) &= E\{\mathbb{1}(Y_i \leq y) \mid \mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)\} \\ &= g^{-1}\left\{\beta_0(y) + \mathbf{X}_{i1}^T \boldsymbol{\beta}_1 + X_{i2} \beta_2(y) \right. \\ &\quad \left. + \int X_{i3}(t) \beta_3(t, y) dt\right\}, \end{aligned} \quad (1)$$

where $F(\cdot)$ denotes the conditional distribution function as before, $g(\cdot)$ is a known, monotone link function, namely, the logit link function defined as $g(x) = \log\{x/(1-x)\}$ for arbitrary scalar $x \in [0, 1]$, $\beta_0(\cdot)$ is an unknown and smooth functional intercept, $\boldsymbol{\beta}_1$ is a p -dimensional parameter capturing the linear additive effect of the covariate vector \mathbf{X}_{i1} , $\beta_2(\cdot)$ is an unknown and smooth function, and $\beta_3(\cdot, \cdot)$ is an unknown and smooth bivariate function. Here, the effect of the nuisance covariates \mathbf{X}_{i1} is $\boldsymbol{\beta}_1$; it is assumed to be constant over y while the smooth intercept $\beta_0(y)$ is y -variant. The effect of X_{i2} is $\beta_2(y)$, which varies smoothly over y ; $\beta_3(\cdot, y)$ quantifies y -variant linear effects of the covariate $X_{i3}(\cdot)$. If the parameter function $\beta_2(\cdot)$ is zero then the covariate X_{i2} has no effect on the distribution of the response Y_i , which is equivalent to X_{i2} having no effect on any quantile level of Y_i . Similarly, it is easy to see that a null effect, say $\beta_3(\cdot, \cdot) \equiv 0$, is equivalent to the case that the functional covariate $X_{i3}(\cdot)$ has no effect on any quantile level of the response. Chen and Müller [21] (CM, henceforth) considered a similar model; however, their approach is restrictive to a single functional covariate. We discuss the differences between their method and ours in Section 2.2.

To explain our ideas, we consider the case that the functional covariates are observed without noise on a fine, regular, and common grid of sampling points, i.e., $t_{ij} = t_j$ with $j = 1, \dots, m$ for all i . Bear in mind, this assumption is made for illustration only, and our framework can be extended to more general cases, including settings where the functional covariate is observed with noise and at irregular sampling points; see Section 3.3.

2.2. Modeling of the Covariate Effects. We model $\beta_0(y)$ and $\beta_2(y)$ by using prespecified, truncated univariate basis. Let $\{B_{0,d_0}(\cdot) : d_0 = 1, \dots, \kappa_0\}$ and $\{B_{1,d_1}(\cdot) : d_1 = 1, \dots, \kappa_1\}$ be two bases of dimensions κ_0 and κ_1 , respectively. $\beta_0(y) \approx \sum_{d_0=1}^{\kappa_0} B_{0,d_0}(y) \theta_{0,d_0}$ and $\beta_2(y) \approx \sum_{d_1=1}^{\kappa_1} B_{1,d_1}(y) \theta_{1,d_1}$, where θ_{0,d_0} 's and θ_{1,d_1} 's are unknown basis coefficients. We represent $\beta_3(t, y)$ using the tensor product of two univariate bases functions, $\{B_{2,d_2}^t(t) : d_2 = 1, \dots, \kappa_{2,t}\}$ and $\{B_{2,d_2}^y(y) : d_2 = 1, \dots, \kappa_{2,y}\}$, where $\kappa_{2,t}$ and $\kappa_{2,y}$ are the bases dimensions; $\beta_3(t, y) \approx \sum_{d_2=1}^{\kappa_{2,t}} \sum_{d_2'=1}^{\kappa_{2,y}} B_{2,d_2}^t(t) B_{2,d_2'}^y(y) \theta_{2,d_2,d_2}'$, where θ_{2,d_2,d_2}' 's are unknown basis coefficients. In practice, the integration term $\int X_{i3}(t) \beta_3(t, y) dt$ is approximated by Riemann integration $\int X_{i3}(t) \beta_3(t, y) dt \approx \sum_{j=1}^m X_{i3}(t_j) \beta_3(t_j, y) (t_{j+1} - t_j)$ but other numerical approximation scheme can be also used.

Define $Z_i(y) = \mathbb{1}(Y_i \leq y)$ for $y \in \mathbb{R}$. In practice for each y in a fine grid, we view $Z_i(y)$ as a binary-valued random functional variable. It follows that model (1) can be written equivalently as a generalized function-on-function regression model through relating the ‘‘artificial’’ binary functional response $Z_i(y)$ and the mixed covariates $\mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)$. This model can be fitted by using, for example, the ideas of Scheipl et al. [22] which we briefly summarize next.

Model (1) can be represented as the following generalized additive model:

$$\begin{aligned} E[Z_i(y) \mid \mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)] &= g^{-1}\{\eta_i(y)\}; \\ \eta_i(y) &= \sum_{d_0=1}^{\kappa_0} B_{0,d_0}(y) \theta_{0,d_0} + \mathbf{X}_{i1}^T \boldsymbol{\beta}_1 \\ &+ X_{i2} \sum_{d_1=1}^{\kappa_1} B_{1,d_1}(y) \theta_{1,d_1} + \sum_{j=1}^m (t_{j+1} - t_j) X_{i3}(t_j) \\ &\cdot \sum_{d_2=1}^{\kappa_{2,t}} \sum_{d_2^y=1}^{\kappa_{2,y}} B_{2,d_2}^t(t_j) B_{2,d_2^y}^y(y) \theta_{2,d_2,d_2^y}. \end{aligned} \quad (2)$$

For convenience, we use the notation $B_{X_{i2},d_1}(y) = X_{i2} B_{1,d_1}(y)$, $\bar{X}_{i3}(t_j) = (t_{j+1} - t_j) X_{i3}(t_j)$. We let $\mathbf{B}_0(y) = \{B_{0,1}(y), \dots, B_{0,\kappa_0}(y)\}^T$, $\mathbf{B}_{i1}(y) = \{B_{X_{i2},1}(y), \dots, B_{X_{i2},\kappa_1}(y)\}^T$, $\mathbf{B}_{2,t}(t) = \{B_{2,1}^t(t), \dots, B_{2,\kappa_{2,t}}^t(t)\}^T$, $\mathbf{B}_{2,y}(y) = \{B_{2,1}^y(y), \dots, B_{2,\kappa_{2,y}}^y(y)\}^T$, $\boldsymbol{\theta}_0 = \{\theta_{0,1}, \dots, \theta_{0,\kappa_0}\}^T$, $\boldsymbol{\theta}_1 = \{\theta_{1,1}, \dots, \theta_{1,\kappa_1}\}^T$, and $\boldsymbol{\theta}_2 = [\theta_{2,d_2,d_2^y}]_{1 \leq d_2 \leq \kappa_{2,t}, 1 \leq d_2^y \leq \kappa_{2,y}}$ is a coefficient matrix. Then $\beta_3(t, y) = \{\mathbf{B}_{2,t}(t) \otimes \mathbf{B}_{2,y}(y)\}^T \boldsymbol{\theta}_2$, where $\boldsymbol{\theta}_2$ is the vectorization of $\boldsymbol{\Theta}_2$. We let $\bar{\mathbf{X}}_{i3}(\mathbf{t}) = \{\bar{X}_{i3}(t_1), \dots, \bar{X}_{i3}(t_m)\}^T \in \mathbb{R}^m$, $\mathbf{B}_2(\mathbf{t}, y) = \{\mathbf{B}_{2,t}(t_1) \otimes \mathbf{B}_{2,y}(y), \dots, \mathbf{B}_{2,t}(t_m) \otimes \mathbf{B}_{2,y}(y)\}^T \in \mathbb{R}^{m \times \kappa_{2,t} \times \kappa_{2,y}}$, and $\mathbf{B}_{i2}(y) = \mathbf{B}_2(\mathbf{t}, y)^T \bar{\mathbf{X}}_{i3}(\mathbf{t})$. Now model (2) can be written as

$$\begin{aligned} E[Z_i(y) \mid \mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)] &= g^{-1}\{\mathbf{X}_{i1}^T \boldsymbol{\beta}_1 + \mathbf{B}_0(y)^T \boldsymbol{\theta}_0 \\ &+ \mathbf{B}_{i1}(y)^T \boldsymbol{\theta}_1 + \mathbf{B}_{i2}(y)^T \boldsymbol{\theta}_2\}. \end{aligned} \quad (3)$$

The general idea is to set the bases dimensions $\kappa_0, \kappa_1, \kappa_{2,t}$, and $\kappa_{2,y}$ to be sufficiently large to capture the complexity of the coefficient functions and control the smoothness of the estimator through some roughness penalties. This approach of using roughness penalties has been widely used; see, for example, Eilers and Marx [23]; Ruppert [24]; Wood [25, 26] among many others.

It is important to emphasize that, even in the case of a single functional covariate, our methodology differs from [21] in two directions: (1) our proposed method is based on modeling the unknown smooth coefficient functions using prespecified basis function expansion and using penalties to control their roughness. In contrast, CM uses data-driven basis and chooses the number of basis functions through the percentage of explained variance (PVE) of the functional predictors. This key difference allows our method to accommodate covariates of different types as well as

nonlinear effects. (2) Our estimation approach is based on a single step penalized function-on-function regression while CM uses pointwise estimation based on functional principal component bases and thus requires fitting multiple generalized regressions. This nice feature leads to an computational advantage.

3. Estimation

3.1. Estimation via Penalized Log-Likelihood. Let $\{y_\ell : \ell = 1, \dots, L\}$ be a set of equally spaced points in the range of the response variable, Y_i 's. Conditioning on $\{\mathbf{X}_{i1}, X_{i2}, X_{i3}(\cdot)\}$, we model $Z_i(y_\ell)$ as independently distributed Bernoulli variables with mean $\mu_i(y_\ell)$, where $g\{\mu_i(y_\ell)\} = \eta_i(y_\ell)$. The coefficients $\boldsymbol{\beta}_1, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ are estimated by minimizing the penalized log-likelihood criterion:

$$\begin{aligned} -2\mathcal{L}(\boldsymbol{\beta}_1, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \mid \{Z_i(y_\ell) : \forall i, \ell\}) &+ \lambda_0 P_0(\boldsymbol{\theta}_0) \\ &+ \lambda_1 P_1(\boldsymbol{\theta}_1) + \lambda_{2,t} P_{2,t}(\boldsymbol{\theta}_2) + \lambda_{2,y} P_{2,y}(\boldsymbol{\theta}_2), \end{aligned} \quad (4)$$

where \mathcal{L} is the log-likelihood function of data $\{Z_i(y_\ell) : \ell = 1, \dots, L\}_{1 \leq i \leq n}$, $\lambda_0, \lambda_1, \lambda_{2,t}$, and $\lambda_{2,y}$ are smoothing parameters, which control the balance between the model fit and its complexity, and $P_0(\cdot), P_1(\cdot), P_{2,t}(\cdot)$, and $P_{2,y}(\cdot)$ are all penalties.

There are several choices to define the penalty matrix in nonparametric regression; see Eilers and Marx [23] and Wood [26]. We use quadratic penalties which penalize the size of the curvature of the estimated coefficient functions. Let $P_0(\boldsymbol{\theta}_0) = \int \{\partial^2 \beta_0(y) / \partial y^2\}^2 dy = \boldsymbol{\theta}_0^T \mathbf{D}_0 \boldsymbol{\theta}_0$, where \mathbf{D}_0 is of dimension $\kappa_0 \times \kappa_0$ with its (s, s') element equal to $\int \{\partial^2 B_{0,s}(y) / \partial y^2\} \{\partial^2 B_{0,s'}(y) / \partial y^2\} dy$. Similarly, $P_1(\boldsymbol{\theta}_1) = \int \{\partial^2 \beta_2(y) / \partial y^2\}^2 dy = \boldsymbol{\theta}_1^T \mathbf{D}_1 \boldsymbol{\theta}_1$, where \mathbf{D}_1 is of dimension $\kappa_1 \times \kappa_1$ with its (s, s') element equal to $\int \{\partial^2 B_{1,s}(y) / \partial y^2\} \{\partial^2 B_{1,s'}(y) / \partial y^2\} dy$. As $\beta_3(\cdot, \cdot)$ is a bivariate function, the choice of penalty penalizing the size of curvature in each direction, respectively: $P_{2,t}(\boldsymbol{\theta}_2) = \int \int \{\partial^2 \beta_3(y, t) / \partial t^2\}^2 dy dt = \boldsymbol{\theta}_2^T \mathbf{D}_{2,t} \boldsymbol{\theta}_2$, where $\mathbf{D}_{2,t} = \mathbf{P}_{2,t} \otimes \mathbf{I}_{\kappa_{2,y}}$ is of dimension $\kappa_{2,y} \kappa_{2,t} \times \kappa_{2,y} \kappa_{2,t}$ with the (s, s') element of $\mathbf{P}_{2,t}$ equal to $\int \{\partial^2 B_{2,s}^t(t) / \partial t^2\} \{\partial^2 B_{2,s'}^t(t) / \partial t^2\} dt$ for some orthonormal spline bases. Similarly, $P_{2,y}(\boldsymbol{\theta}_2) = \int \int \{\partial^2 \beta_3(y, t) / \partial y^2\}^2 dy dt = \boldsymbol{\theta}_2^T \mathbf{D}_{2,y} \boldsymbol{\theta}_2$, where $\mathbf{D}_{2,y} = \mathbf{I}_{\kappa_{2,t}} \otimes \mathbf{P}_{2,y}$ is of dimension $\kappa_{2,y} \kappa_{2,t} \times \kappa_{2,y} \kappa_{2,t}$ with the (s, s') element of $\mathbf{P}_{2,y}$ equal to $\int \{\partial^2 B_{2,s}^y(y) / \partial y^2\} \{\partial^2 B_{2,s'}^y(y) / \partial y^2\} dy$.

Criterion (4) can be viewed as a penalized quasilielihood (PQL) of the corresponding generalized linear mixed model

$$\begin{aligned} Z_i(y_\ell) \mid \boldsymbol{\beta}_1, \boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 &\sim \text{Bernoulli}(\mu_i(y_\ell)), \\ \ell &= 1, \dots, L; \end{aligned}$$

$$\boldsymbol{\theta}_0 \sim N(\mathbf{0}, \lambda_0^{-1} \mathbf{D}_0^-); \quad (5)$$

$$\boldsymbol{\theta}_1 \sim N(\mathbf{0}, \lambda_1^{-1} \mathbf{D}_1^-);$$

$$\boldsymbol{\theta}_2 \sim N(\mathbf{0}, \mathbf{D}_2^-),$$

where \mathbf{D}_0^- is the generalized inverse matrix of \mathbf{D}_0 ; \mathbf{D}_1^- and $\mathbf{D}_2^- = (\lambda_{2,t} \mathbf{D}_{2,t} + \lambda_{2,y} \mathbf{D}_{2,y})^-$ are defined similarly. Wood [26]

discusses an alternative way to deal with the rank-deficient matrices in the context of restricted maximum likelihood (REML) estimation. Here we do not account for the dependence over y ; see Scheipl et al. [22] for a general formulation. See also Ivanescu et al. [27] who use the mixed model representation of a similar regression model to (5), but with a Gaussian functional response. The smoothing parameters are estimated using REML.

3.2. Extension to Nonlinear Model. One advantage of the proposed framework is that it can be easily extended to allow for more flexible effects, i.e., extending the ideas to accommodate multiple covariates, scalar or functional, and varied types of effects. In particular, the smooth effect $X_{i2}\beta_2(y)$ can be replaced by $h_1(X_{i2}, y)$ and $\int X_{i3}(t)\beta_3(t, y)dt$ by $\int h_2\{X_{i3}(t), t, y\}dt$, where $h_1(\cdot, \cdot)$ and $h_2(\cdot, \cdot, \cdot)$ are unknown bivariate and trivariate smooth functions, respectively; see Scheipl et al. [22] and Kim et al. [28]. These changes require little additional computational effort. The modeling and estimation follow roughly similar ideas as Scheipl et al. [22]. We consider the nonlinear model in the simulation study for the case of having a scalar covariate only, i.e., $h_1(X_i, y)$, and the corresponding results are presented in Section S1.1 of the Supplementary Materials. The results show excellent prediction performance compared to the competitive nonlinear quantile regression method, namely, constrained B-spline smoothing (COBS) [29].

3.3. Extension to Sparse and Noisy Functional Covariates. In practice the functional covariates are often observed at irregular times across the units and the measurements are possibly corrupted by noises. In such case, one needs to first smooth and denoise the trajectories before fitting. When the sampling design of the functional covariate is dense, the common approach is to smooth each trajectory using splines or local polynomial smoothing, as proposed in Ramsay and Silverman [3] and Zhang and Chen [30]. When the design is sparse, the smoothing can be done by pooling all the subjects and following the PACE method proposed in Yao et al. [31]. As recovering the trajectories has been extensively discussed in the literatures, we do not review the procedures here. Instead, we discuss some available computing resources that can be used to fit these methods. In our numerical study, we used `fPCA.sc` function in the `refund` R package [32] for recovering the latent trajectories, irrespective of a sampling design (dense or sparse). Alternatively, one can use `fPCA.face` [33] in `refund` for regular dense design and `face.sparse` [34] in the R package `face` [35] for irregular sparse design. Once the latent trajectories are estimated, they can be used in the fitting criterion (4).

3.4. Estimation of Conditional Quantile. Let $\hat{\beta}_1, \hat{\theta}_0, \hat{\theta}_1$, and $\hat{\theta}_2$ be parameter estimates in (5). It follows that the estimated distribution function $\hat{F}_{Y_i|X_{i1}, X_{i2}, X_{i3}(\cdot)}(y)$ can be obtained by plugging in the estimated coefficients. The τ th conditional quantile is estimated by inverting the estimated distribution, i.e., $\hat{Q}_{Y_i|X_{i1}, X_{i2}, X_{i3}(\cdot)}(\tau) = \inf\{y : \hat{F}_{Y_i|X_{i1}, X_{i2}, X_{i3}(\cdot)}(y) \geq \tau\}$. The

estimated distribution function is not a monotonic function yet. In practice we suggest to first apply a monotonicity method as described in Section 3.5 and then estimate the conditional quantiles by inverting the resulting estimated distribution.

3.5. Monotonization and Implementation. While a conditional quantile function is nondecreasing, the resulting estimated quantiles may not be. Two approaches are widely used: one is to monotonize the estimated conditional distribution function and the other is to monotonize the estimated conditional quantile function. We choose the former as $\hat{F}_{Y_i|X_{i1}, X_{i2}, X_{i3}(\cdot)}(y)$ is readily available at dense grid points y_ℓ 's. We use an isotonic regression model [36] for monotonization, which imposes an order restriction; this is done by using the R function `isoreg`. Other monotonization approaches include Chernozhukov et al. [37], which was employed in Kato [16].

Our approach is implemented by first creating an artificial binary response and then fitting a penalized function-on-function regression model and using the logit link function. Fitting models in (4) can be done by extending the ideas of Ivanescu et al. [27] for Gaussian functional response; the extension of the model to the non-Gaussian functional response has recently been studied and implemented by Scheipl et al. [22] as the `pffr` function in `refund` package [32].

4. Simulation Study

4.1. Simulation Setting. In this section we evaluate the empirical performance of the proposed method. We present results for the case when we have both functional and scalar covariates; additional results when there is only a single scalar or a single functional covariate are discussed in the Supplementary Materials, Section S1.

Suppose the observed data for the i th subject are $[Y_i, X_{1i}, \{(W_{i1}, t_{i1}), \dots, (W_{im_i}, t_{im_i})\}]$, $t_{ij} \in [0, 10]$, where $X_{1i} \stackrel{i.i.d.}{\sim} \text{Unif}(-16, 16)$, $W_{ij} = X_{2i}(t_{ij}) + \epsilon_{ij}$ for $1 \leq i \leq n, 1 \leq j \leq m_i$. Let $X_{2i}(t_{ij}) = \mu(t_{ij}) + \sum_{k=1}^4 \xi_{ik} \phi_k(t_{ij}) + \epsilon_{ij}$, where $\mu(t) = t + \sin(t)$, $\phi_k(t) = \cos\{(k+1)\pi t/10\}/\sqrt{5}$ for odd values of k , $\phi_k(t) = \sin\{k\pi t/10\}/\sqrt{5}$ for even values of k , $\xi_{ik} \stackrel{i.i.d.}{\sim} N(0, \lambda_k)$, $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \{16, 9, 7.56, 5.06\}$, and $\epsilon_{ij} \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$. We assume three cases for generating response Y_i :

- (i) **Gaussian:** $Y_i | X_{1i}, X_{2i}(\cdot) \sim N(2 \int X_{2i}(t)\beta(t)dt + 2X_{1i}, 5^2)$; this corresponds to the quantile regression model $Q_{Y_i|X_{1i}, X_{2i}(\cdot)}(\tau) = 2 \int X_{2i}(t)\beta(t)dt + 2X_{1i} + 5\Phi^{-1}(\tau)$, where $\Phi(\cdot)$ is the distribution function of the standard normal;
- (ii) **Mixture of Gaussians:** $Y_i | X_{1i}, X_{2i}(\cdot) \sim 0.5N(\int X_{2i}(t)\beta(t)dt + X_{1i}, 1^2) + 0.5N(3 \int X_{2i}(t)\beta(t)dt + 3X_{1i}, 4^2)$, where the true quantiles can be approximated numerically by using `qnorMix` function in the R package `norMix`;
- (iii) **Gaussian with heterogeneous error:** $Y_i | X_{1i}, X_{2i}(\cdot) \sim N(2 \int X_{2i}(t)\beta(t)dt + 2X_{1i}, 5^2 \int X_{2i}(t)^2 dt)$

$$\sum_{k=1}^4 \lambda_k); \text{ the true quantiles are given by } Q_{Y|X_1, X_2(\cdot)}(\tau) = 2 \int X_{2i}(t)\beta(t)dt + 2X_{1i} + 5\sqrt{\int X_{2i}(t)^2 dt / \sum_{k=1}^4 \lambda_k} \Phi^{-1}(\tau).$$

Let $\beta(t) = \sum_{k=1}^4 \beta_k \phi_k(t)$, where $\beta_k = 1$ for $k = 1, \dots, 4$.

For each case, we use different combinations of signal to noise ratio (SNR), sample size, and sampling designs to generate 500 simulated datasets. We define SNR as $\sqrt{\sum_{k=1}^4 \lambda_k} / \sigma_\epsilon$, and we consider five levels of noise: $\text{SNR} = \{150, 10, 5, 2, 1\}$. Two levels of sample size are $n = 100$ and $n = 1000$. Two sampling designs are considered: (i) *sparse design*, where $\{t_{ij} : j = 1, \dots, m\}$ are $m = 15$ randomly selected points from a set of 30 equispaced grids in $[0, 10]$; and (ii) *dense design*, where the sampling points $\{t_{ij} = t_j : j = 1, \dots, m\}$ are $m = 30$ equispaced time points in $[0, 10]$.

The performance is evaluated on a test set of 100 subjects, for which we have $\{X_{1i^*}, (W_{i^*j}, t_{i^*j}), j = 1, \dots, m\}$ available, in terms of mean absolute error (MAE) for quantile levels $\tau = 0.05, 0.1, 0.25$, and 0.5 :

$$\text{MAE}(\tau) = \frac{1}{100} \sum_{i^*=1}^{100} |\widehat{Q}_{i^*}(\tau) - Q_{i^*}(\tau)|. \quad (6)$$

4.2. Competing Methods. We denote the proposed method by Joint QR to emphasize the single step estimation approach. We compare our method with two alternative approaches: (1) a variant of our proposed approach using pointwise estimation, denoted by Pointwise QR. This approach consists of fitting multiple regression models with binomial link function as implemented by the penalized functional regression `pfr`, developed by Goldsmith et al. [7], of the `rEfund` package for generalized scalar responses; (2) a modified version of the CM method, denoted by Mod CM, that we developed to account for additional scalar covariates and which fits multiple generalized linear models with scalar covariates and fPC scores as predictors; (3) a linear quantile regression approach using the quantile loss function and the partial quantile regression bases for functional covariates, proposed by Yu et al. [19] and denoted by PQR. Notice that although the formulation of the first two methods implicitly accounts for a varying effect of the covariates on the response distribution, they do not ensure that this effect is smooth. The third approach can only estimate a specific quantile rather than the entire conditional distribution. Note that all the competing methods are monotone for a fair comparison.

The R function `pfr` can incorporate both scalar/vector and functional predictors by adopting a mixed effects model framework. The functional covariates are presmoothed by fPC analysis [31]. Throughout the simulation study we fix PVE as 0.95 for fPC analysis to determine the number of principal components and use REML to select the smoothing parameters for our proposed methods. Other basis settings are set to their default values. We use 100 equally distanced points between the minimum and maximum of the observed Y_i 's to set the grid $\{y_\ell : \ell = 1, \dots, L\}$ for the conditional distribution function.

4.3. Simulation Results. Tables 1 and 2 show the accuracy of the quantile prediction for the two cases (normal and mixture) when the functional covariate is observed sparsely and the sample size is $n = 100$ (Table 1) and $n = 1000$ (Table 2). Table 3 presents the prediction accuracy for the case of heteroskedasticity with sparsely observed functional covariates. The results based on dense sampling design show similar patterns and thus are relegated to the supplement; see Section S1.3. The comparison of running times is presented in Table 4.

For the case when the response is Gaussian, Tables 1 and 2 suggest that the Joint QR typically outperforms its competitors especially for lower quantile levels ($\tau = 0.05$ and $\tau = 0.1$). For very small noise level ($\text{SNR} = 150$), PQR performs the best, followed closely by the proposed Joint QR. The variant Pointwise QR, which has a poorer performance, is generally better than the modified CM approach. As expected, as the sample size increases ($n = 1000$), all the accuracy results improve; the proposed Joint QR continues to yield most accurate quantiles for the low quantile levels. For mixture of Gaussians, the results are somewhat similar. The accuracy of the quantile estimators with the Pointwise QR improves greatly; in fact the Joint QR and Pointwise QR outperform the other approaches for quantile levels $\tau = 0.05, 0.1, 0.25$ irrespective of the SNR. Finally, Table 3 shows that the results for Gaussian with heterogeneous error are close to those for the case of Gaussian. Again, the proposed method has competitive performance in terms of prediction accuracy.

Table 4 compares the three methods that involve estimating the conditional distribution in terms of the running time required for fitting. The times are reported based on a computer with a 2.3 GHz CPU and 8 GB of RAM. Not surprisingly by fitting the model a single time, Joint QR is the fastest, in some cases being order of magnitude faster than the rest. Pointwise QR can be up to twice as fast as Mod CM.

For completeness, we also compare our proposed method to the appropriate competitive methods for the cases (1) when there is a single scalar covariate and (2) when there is a single functional covariate. In the Supplementary Materials, Section S1.1 discusses the former case and compares Joint QR and Pointwise QR with the linear quantile regression (LQR) [12], implemented by `rq` function in the R package `quantreg`, and the constrained B-splines nonparametric regression quantiles (COBS), implemented by the `cobs` function in the R package `COBS` [29], in an extensive simulation experiment that involves both linear quantile settings and nonlinear quantile settings. Overall the results show that the proposed methods have similar behavior as LQR; see Table S1. Furthermore we consider the proposed method and its variant with nonlinear modeling of the conditional distribution as discussed in Section 3.2, which we denote with Joint QR (NL) for joint fitting and Pointwise QR (NL) for pointwise fitting. Nonlinear versions of the proposed methods have an excellent MAE performance, which is comparable to or better than that of the COBS method.

Finally, Section S1.2 in the Supplementary Materials discusses the simulation study for the case of having a single functional covariate and compares the proposed methods

TABLE 1: Average MAE (standard error in parentheses) of the predicted τ -level quantile for the case of having a scalar covariate and a sparsely observed functional covariate. Sample size $n = 100$.

Distribution	SNR	Method	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$
Normal	150	Joint QR	3.67 (0.03)	3.53 (0.03)	3.30 (0.02)	3.17 (0.02)
		Pointwise QR	4.96 (0.03)	4.61 (0.03)	4.22 (0.02)	4.18 (0.02)
		Mod CM	6.04 (0.03)	5.81 (0.03)	5.55 (0.03)	5.14 (0.03)
		PQR	3.17 (0.04)	2.71 (0.03)	2.31 (0.02)	2.16 (0.02)
Normal	10	Joint QR	6.32 (0.03)	6.00 (0.03)	5.76 (0.02)	5.73 (0.02)
		Pointwise QR	7.44 (0.04)	6.85 (0.03)	6.39 (0.03)	6.28 (0.03)
		Mod CM	8.20 (0.04)	8.10 (0.04)	8.04 (0.04)	8.01 (0.04)
		PQR	6.82 (0.05)	6.11 (0.04)	5.34 (0.03)	5.09 (0.02)
Normal	5	Joint QR	7.84 (0.04)	7.34 (0.03)	6.93 (0.03)	6.84 (0.03)
		Pointwise QR	8.91 (0.04)	8.12 (0.04)	7.45 (0.03)	7.26 (0.03)
		Mod CM	9.34 (0.04)	9.23 (0.04)	9.14 (0.05)	9.06 (0.05)
		PQR	8.68 (0.06)	7.81 (0.05)	6.74 (0.03)	6.34 (0.02)
Normal	2	Joint QR	10.05 (0.05)	9.22 (0.04)	8.47 (0.03)	8.28 (0.03)
		Pointwise QR	10.91 (0.06)	9.86 (0.04)	8.87 (0.04)	8.54 (0.03)
		Mod CM	10.85 (0.05)	10.55 (0.05)	10.34 (0.06)	10.21 (0.06)
		PQR	11.21 (0.08)	10.03 (0.06)	8.56 (0.04)	7.96 (0.03)
Normal	1	Joint QR	11.50 (0.06)	10.41 (0.05)	9.40 (0.04)	9.11 (0.03)
		Pointwise QR	12.12 (0.06)	10.88 (0.05)	9.70 (0.04)	9.30 (0.03)
		Mod CM	11.95 (0.06)	11.46 (0.06)	11.07 (0.06)	11.05 (0.07)
		PQR	12.82 (0.08)	11.38 (0.06)	9.60 (0.04)	8.86 (0.03)
Mixture	150	Joint QR	6.92 (0.06)	6.23 (0.06)	6.16 (0.06)	4.81 (0.06)
		Pointwise QR	8.10 (0.08)	6.80 (0.06)	6.66 (0.06)	5.25 (0.06)
		Mod CM	9.18 (0.07)	8.99 (0.07)	8.93 (0.07)	7.90 (0.07)
		PQR	8.43 (0.06)	7.18 (0.04)	6.22 (0.04)	5.48 (0.14)
Mixture	10	Joint QR	9.02 (0.06)	7.95 (0.05)	7.85 (0.05)	6.19 (0.06)
		Pointwise QR	10.11 (0.07)	8.52 (0.05)	7.95 (0.05)	6.42 (0.06)
		Mod CM	11.33 (0.07)	10.95 (0.07)	10.79 (0.07)	9.80 (0.08)
		PQR	10.72 (0.08)	8.99 (0.05)	7.63 (0.04)	5.33 (0.09)
Mixture	5	Joint QR	10.18 (0.06)	8.91 (0.05)	8.53 (0.05)	6.82 (0.05)
		Pointwise QR	11.18 (0.07)	9.38 (0.05)	8.58 (0.05)	6.98 (0.05)
		Mod CM	12.19 (0.07)	11.75 (0.07)	11.52 (0.07)	10.47 (0.08)
		PQR	12.12 (0.09)	10.12 (0.06)	8.40 (0.04)	5.73 (0.07)
Mixture	2	Joint QR	11.93 (0.07)	10.26 (0.05)	9.46 (0.05)	7.61 (0.05)
		Pointwise QR	12.68 (0.08)	10.63 (0.06)	9.51 (0.05)	7.70 (0.05)
		Mod CM	13.33 (0.08)	12.60 (0.08)	12.27 (0.08)	11.08 (0.10)
		PQR	14.16 (0.10)	11.72 (0.06)	9.55 (0.04)	6.41 (0.05)
Mixture	1	Joint QR	13.17 (0.08)	11.19 (0.05)	10.06 (0.05)	8.04 (0.05)
		Pointwise QR	13.71 (0.09)	11.44 (0.06)	10.10 (0.05)	8.13 (0.05)
		Mod CM	14.16 (0.08)	13.22 (0.08)	12.71 (0.09)	11.44 (0.11)
		PQR	15.44 (0.11)	12.84 (0.07)	10.23 (0.04)	6.89 (0.04)

with CM in terms of MAE as well as computation time; see results displayed in Tables S2 and S3. The results show that the proposed Joint QR is comparable to CM in terms of the prediction accuracy and has less computation time. In our simulation study we also consider the joint fitting of the model by treating the binary response as normal and use `pffr` [27] with Gaussian link, denoted by Joint QR (G).

5. Sow Data Application

Our motivating application is an experimental study carried out at a commercial farm in Oklahoma from July 21, 2013, to August 19, 2013 [38]. The study comprises 480 lactating sows of different parities (i.e., number of previous pregnancies, which serves as a surrogate for age and body weight) that

were observed during their first 21 lactation days; their feed intake was recorded daily as the difference between the feed offer and the feed refusal. In addition the study contains information on the temperature and humidity of the farrowing rooms, each recorded at five minute intervals. The final dataset we used for the analysis consists of 475 sows after five sows with unreliable measurements were removed by the experimenters.

The experiment was conducted to gain better insights into the way that the ambient temperature and humidity of the farrowing room affect the feed intake of lactating sows. Previous studies seem to suggest a reduction in the sow's feed intake due to heat stress: above 29°C sows decrease feed intake by 0.5 kg per additional degree in temperature [9]. Studying the effect of heat stress on lactating sows is a very

TABLE 2: Average MAE (standard error in parentheses) of the predicted τ -level quantile for the case of having a scalar covariate and a sparsely observed functional covariate. Sample size $n = 1000$.

Distribution	SNR	Method	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$
Normal	150	Joint QR	1.68 (0.01)	1.65 (0.01)	1.62 (0.01)	1.60 (0.01)
		Pointwise QR	1.94 (0.01)	1.92 (0.01)	1.88 (0.01)	1.81 (0.01)
		Mod CM	1.93 (0.01)	1.88 (0.01)	1.87 (0.01)	1.87 (0.01)
		PQR	1.72 (0.01)	1.61 (0.01)	1.51 (0.01)	1.48 (0.01)
Normal	10	Joint QR	5.45 (0.02)	4.97 (0.02)	4.66 (0.02)	4.65 (0.02)
		Pointwise QR	5.64 (0.02)	5.13 (0.02)	4.78 (0.02)	4.75 (0.02)
		Mod CM	5.69 (0.02)	5.21 (0.02)	4.87 (0.02)	4.85 (0.02)
		PQR	5.85 (0.02)	5.37 (0.02)	4.81 (0.02)	4.60 (0.02)
Normal	5	Joint QR	7.34 (0.03)	6.54 (0.02)	5.94 (0.02)	5.85 (0.02)
		Pointwise QR	7.53 (0.03)	6.69 (0.02)	6.04 (0.02)	5.94 (0.02)
		Mod CM	7.53 (0.03)	6.77 (0.02)	6.18 (0.02)	6.06 (0.02)
		PQR	7.84 (0.03)	7.05 (0.02)	6.16 (0.02)	5.81 (0.02)
Normal	2	Joint QR	9.97 (0.03)	8.70 (0.03)	7.62 (0.02)	7.38 (0.02)
		Pointwise QR	10.15 (0.03)	8.85 (0.03)	7.71 (0.02)	7.45 (0.02)
		Mod CM	10.12 (0.03)	8.93 (0.03)	7.87 (0.03)	7.60 (0.03)
		PQR	10.55 (0.04)	9.34 (0.03)	7.91 (0.03)	7.34 (0.02)
Normal	1	Joint QR	11.69 (0.04)	10.10 (0.03)	8.68 (0.03)	8.32 (0.03)
		Pointwise QR	11.88 (0.04)	10.25 (0.04)	8.77 (0.03)	8.39 (0.03)
		Mod CM	11.85 (0.04)	10.37 (0.04)	8.96 (0.03)	8.57 (0.03)
		PQR	12.29 (0.04)	10.77 (0.04)	9.02 (0.03)	8.29 (0.03)
Mixture	150	Joint QR	4.56 (0.02)	4.44 (0.02)	4.33 (0.03)	3.68 (0.03)
		Pointwise QR	4.34 (0.03)	4.24 (0.02)	4.20 (0.03)	3.66 (0.04)
		Mod CM	4.68 (0.03)	4.59 (0.02)	4.38 (0.03)	4.01 (0.03)
		PQR	7.84 (0.03)	6.45 (0.02)	5.29 (0.02)	3.19 (0.01)
Mixture	10	Joint QR	7.56 (0.03)	6.79 (0.02)	6.14 (0.03)	5.02 (0.03)
		Pointwise QR	7.49 (0.03)	6.61 (0.02)	5.88 (0.03)	5.02 (0.03)
		Mod CM	7.67 (0.04)	6.97 (0.03)	6.17 (0.03)	5.52 (0.04)
		PQR	10.11 (0.04)	8.29 (0.03)	6.71 (0.02)	3.62 (0.02)
Mixture	5	Joint QR	9.20 (0.04)	7.13 (0.03)	6.94 (0.03)	5.6 (0.03)
		Pointwise QR	9.17 (0.04)	7.01 (0.03)	6.71 (0.03)	5.58 (0.03)
		Mod CM	9.32 (0.04)	7.37 (0.03)	7.08 (0.03)	6.18 (0.04)
		PQR	11.46 (0.04)	9.36 (0.03)	7.49 (0.03)	4.32 (0.02)
Mixture	2	Joint QR	11.57 (0.05)	9.02 (0.03)	8.05 (0.03)	6.35 (0.03)
		Pointwise QR	11.58 (0.05)	8.98 (0.04)	7.90 (0.03)	6.31 (0.03)
		Mod CM	11.69 (0.05)	9.38 (0.04)	8.38 (0.04)	7.01 (0.04)
		PQR	13.49 (0.05)	10.95 (0.04)	8.61 (0.03)	5.32 (0.02)
Mixture	1	Joint QR	13.18 (0.05)	10.31 (0.04)	8.79 (0.04)	6.79 (0.03)
		Pointwise QR	13.21 (0.05)	10.28 (0.04)	8.69 (0.04)	6.73 (0.03)
		Mod CM	13.26 (0.05)	10.67 (0.04)	9.21 (0.04)	7.47 (0.04)
		PQR	14.93 (0.06)	12.05 (0.04)	9.35 (0.04)	5.96 (0.02)

important scientific question because of a couple of reasons. First, the reduction of feed intake of the lactating sows is associated with a decrease in both their bodyweight (BW) and milk production, as well as the weight gain of their litter [10, 39, 40]. Sows with poor feed intake during lactation continue the subsequent reproductive period with negative energy balance [41], which leads to preventing the onset of a new reproductive cycle. Second, heat stress reduces farrowing rate (number of sows that deliver a new litter) and number

of piglets born [42]; the reduction in reproduction due to seasonality is estimated to cost 300 million dollars per year for the swine industry [11]. Economic losses are estimated to increase [43] because high temperatures are likely to occur more frequently due to global warming [44].

Our primary goal is to understand the thermal needs of the lactating sows for proper feeding behavior during the lactation time. We are interested in how the interplay between the temperature and humidity of the farrowing room

TABLE 3: Average MAE (standard error in parentheses) of the predicted τ -level quantile for the case of heteroskedasticity with a scalar covariate and a sparsely observed functional covariate.

Sample size	SNR	Method	$\tau = 0.05$	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$
100	150	Joint QR	4.41 (0.03)	3.99 (0.02)	3.48 (0.02)	3.25 (0.02)
		Pointwise QR	5.40 (0.03)	4.84 (0.03)	4.26 (0.02)	4.21 (0.03)
		Mod CM	6.00 (0.03)	5.71 (0.03)	5.42 (0.03)	5.39 (0.03)
		PQR	4.38 (0.05)	3.43 (0.03)	2.52 (0.02)	2.14 (0.02)
100	10	Joint QR	6.93 (0.03)	6.40 (0.03)	5.89 (0.02)	5.76 (0.02)
		Pointwise QR	7.94 (0.04)	7.20 (0.03)	6.48 (0.03)	6.30 (0.03)
		Mod CM	8.41 (0.04)	8.22 (0.04)	8.13 (0.04)	8.05 (0.04)
		PQR	7.76 (0.06)	6.57 (0.04)	5.47 (0.03)	5.08 (0.02)
100	5	Joint QR	8.44 (0.04)	7.72 (0.03)	7.05 (0.03)	6.86 (0.03)
		Pointwise QR	9.44 (0.05)	8.50 (0.04)	7.57 (0.03)	7.26 (0.03)
		Mod CM	9.65 (0.05)	9.40 (0.05)	9.36 (0.05)	9.26 (0.05)
		PQR	9.51 (0.07)	8.25 (0.05)	6.87 (0.03)	6.33 (0.02)
100	2	Joint QR	10.68 (0.05)	9.63 (0.04)	8.60 (0.03)	8.30 (0.03)
		Pointwise QR	11.55 (0.06)	10.29 (0.05)	9.00 (0.04)	8.54 (0.03)
		Mod CM	11.27 (0.05)	10.80 (0.05)	10.43 (0.05)	10.40 (0.06)
		PQR	12.07 (0.08)	10.53 (0.06)	8.66 (0.04)	7.92 (0.03)
100	1	Joint QR	12.23 (0.06)	10.91 (0.05)	9.54 (0.04)	9.11 (0.03)
		Pointwise QR	12.87 (0.07)	11.39 (0.05)	9.85 (0.04)	9.29 (0.03)
		Mod CM	12.43 (0.06)	11.80 (0.06)	11.19 (0.06)	11.09 (0.07)
		PQR	13.70 (0.09)	11.91 (0.06)	9.70 (0.04)	8.84 (0.03)
1000	150	Joint QR	2.87 (0.01)	2.42 (0.01)	1.89 (0.01)	1.65 (0.01)
		Pointwise QR	3.06 (0.01)	2.60 (0.01)	2.07 (0.01)	1.86 (0.01)
		Mod CM	3.13 (0.01)	2.65 (0.01)	2.10 (0.01)	1.91 (0.01)
		PQR	3.10 (0.01)	2.47 (0.01)	1.78 (0.01)	1.48 (0.01)
1000	10	Joint QR	6.21 (0.02)	5.45 (0.02)	4.79 (0.02)	4.66 (0.02)
		Pointwise QR	6.38 (0.02)	5.59 (0.02)	4.90 (0.02)	4.77 (0.02)
		Mod CM	6.46 (0.02)	5.70 (0.02)	5.01 (0.02)	4.86 (0.02)
		PQR	6.68 (0.03)	5.82 (0.02)	4.92 (0.02)	4.60 (0.02)
1000	5	Joint QR	8.08 (0.03)	7.01 (0.02)	6.07 (0.02)	5.85 (0.02)
		Pointwise QR	8.27 (0.03)	7.15 (0.02)	6.16 (0.02)	5.94 (0.02)
		Mod CM	8.30 (0.03)	7.26 (0.02)	6.31 (0.02)	6.08 (0.02)
		PQR	8.64 (0.03)	7.50 (0.03)	6.27 (0.02)	5.81 (0.02)
1000	2	Joint QR	10.76 (0.03)	9.21 (0.03)	7.76 (0.02)	7.38 (0.02)
		Pointwise QR	10.95 (0.04)	9.34 (0.03)	7.85 (0.03)	7.45 (0.02)
		Mod CM	10.92 (0.03)	9.45 (0.03)	8.01 (0.03)	7.60 (0.03)
		PQR	11.38 (0.04)	9.79 (0.03)	8.03 (0.03)	7.34 (0.02)
1000	1	Joint QR	12.54 (0.04)	10.65 (0.03)	8.83 (0.03)	8.32 (0.03)
		Pointwise QR	12.74 (0.04)	10.81 (0.04)	8.92 (0.03)	8.39 (0.03)
		Mod CM	12.70 (0.04)	10.91 (0.04)	9.10 (0.03)	8.56 (0.03)
		PQR	13.16 (0.05)	11.26 (0.04)	9.13 (0.03)	8.29 (0.03)

TABLE 4: Average computing time (in seconds) of the three approaches that involve estimating the conditional distribution for the case of having a scalar covariate and a densely observed functional covariate.

Distribution	Method	$n = 100$	$n = 1000$
Normal	Joint QR	12	133
	Pointwise QR	148	271
	Mod CM	278	511
Mixture	Joint QR	13	154
	Pointwise QR	151	296
	Mod CM	327	532

affects the feed intake demeanor of lactating sows of different parities. We focus on three specific time points during the lactation period—beginning (lactation day 4), middle (day 11), and end (day 18)—and the analyses are done separately for each time point. We consider two types of responses that are meant to assess the feed intake behavior using the current and the previous lactation days. The first one quantifies the absolute change in the feed intake over two consecutive days and the second one quantifies the relative change and takes into account the usual sow's feed intake. We define them formally after introducing some notation.

Let FI_{ij} be the j th measurement of the feed intake observed for the i th sow and denote by the lactation day LD_{ij} when FI_{ij} is measured; here $j = 1, \dots, n_i$. Most sows are observed every day within the first 21 lactation days and thus have $n_i = 21$. First define the absolute change in the feed intake between two consecutive days as $\Delta_{i(j+1)}^{(1)} = FI_{i(j+1)} - FI_{ij}$ for j that satisfies $LD_{i(j+1)} - LD_{ij} = 1$. For instance, $\Delta_{i(j+1)}^{(1)} = 0$ means there was no change in feed intake of sow i between the current day and the previous day, while $\Delta_{i(j+1)}^{(1)} < 0$ means that the feed intake consumed by the i th sow in the current day is smaller than the feed intake consumed in the previous day. However, the same amount of change in the feed intake may reflect some stress level for a sow who typically eats a lot and a more serious stress level for a sow that usually has a lower appetite. For this, we define the relative change in the feed intake by $\Delta_{i(j+1)}^{(2)} = (FI_{i(j+1)} - FI_{ij}) / \{(LD_{i(j+1)} - LD_{ij}) \cdot TA_i\}$, where TA_i is the trimmed average of feed intake of i th sow calculated as the average feed intake after removing the lowest 20% and highest 20% of the feed intake measurements $\{FI_{i1}, \dots, FI_{in_i}\}$ taken for the corresponding sow. Here TA_i is surrogate for the usual amount of feed intake of the i th sow. Trimmed average is used instead of the common average, to remove outliers of very low feed intakes in first few lactating days. For example, consider the situation of two sows: sow i that typically consumes 10lb food per day and sow i' that consumes 5lb food per day. A reduction of 5lb in the feed intake over two consecutive days corresponds to $\Delta_{i(j+1)}^{(2)} = -50\%$ for the i th sow and $\Delta_{i'(j+1)}^{(2)} = -100\%$ for the i' th sow. Clearly both sows are stressed (negative value) but the second sow is stressed more, as its absolute relative change is larger; in view of this we refer to the second response as the *stress index*. Due to the construction of the two types of responses, the data size varies for lactation days 4 ($j = 3$), 11 ($j = 10$), and 18 ($j = 17$); for the first response, $\Delta_{i(j+1)}^{(1)}$, we have sample sizes of 233, 350, and 278, whereas for $\Delta_{i(j+1)}^{(2)}$ the sample sizes are 362, 373, and 336 for the respective lactation days.

In this analysis we center the attention on the effect of the ambient temperature and humidity on the *1st quartile* of the proxy stress measures and gain more understanding of the food consumption of sows that are most susceptible to heat stress. While the association between the feed intake of lactating sows and the ambient conditions of the farrowing room has been an active research area for some time, accounting for the temperature daily profile has not been considered yet hitherto. Figure 1 displays the temperature

and humidity daily profiles recorded at a frequency of 5-minute window intervals for three different days. Preliminary investigation reveals that temperature is negatively correlated with humidity at each time; this phenomenon is caused because the farm uses cool cell panels and fans to control the ambient temperature. Furthermore, it appears that there is a strong pointwise correlation between temperature and humidity. In view of these observations, in our analysis we consider the daily average of humidity. Exploratory analysis of the feed intake behavior of the sows suggests similarities for the sows with parity greater than older sows (ones who are at their third pregnancy or higher); thus we use a parity indicator instead of the actual parity of the sow. The parity indicator P_i is defined as one, if the i th sow has parity one and zero otherwise.

For the analysis we smooth daily temperature measurements of each sow using univariate smoother with 15 cubic regression bases and quadratic penalty; REML is used to estimate smoothing parameter. The smoothed temperature curve for sow i 's j th repeated measure is denoted by $T_{ij}(t)$, $t \in [0, 24)$, and the corresponding daily average humidity is denoted by AH_{ij} . Both temperature and average humidity are centered before being used in the analysis.

For convenience we denote the response with Δ_{ij} by removing the superscript. In this application for fixed j , Δ_{ij} corresponds to Y_i in Section 2, P_i and AH_{ij} correspond to scalar covariates X_{i2} , and $T_{ij}(t)$ and $AH_{ij} \cdot T_{ij}(t)$ correspond to functional covariates $X_{i3}(\cdot)$. We first estimate the conditional distribution of Δ_{ij} given temperature $T_{ij}(t)$, average humidity AH_{ij} , parity P_i , and interaction $AH_{ij} \cdot T_{ij}(t)$. Specifically for each of lactation days of interest ($j = 3, 10$ and 17) we create a set of 100 equispaced grid of points between the fifth smallest and fifth largest values of Δ_{ij} 's and denote the grids with $\mathcal{D} = \{d_\ell : \ell = 1, \dots, 100\}$. Then we create artificial binary responses, $\{\mathbb{1}(\Delta_{ij} \leq d_\ell) : \ell = 1, \dots, 100\}$, and fit the following model for $F_{ij}(d_\ell) = E[\mathbb{1}(\Delta_{ij} \leq d_\ell) | T_{ij}(t), AH_{ij}, P_i]$:

$$\begin{aligned} E \left[\mathbb{1}(\Delta_{ij} \leq d_\ell) | T_{ij}(t), AH_{ij}, P_i \right] &= g^{-1} \left\{ \beta_0(d_\ell) \right. \\ &+ \beta_1(d_\ell) P_i + \beta_2(d_\ell) AH_{ij} + \int \beta_3(d_\ell, t) T_{ij}(t) dt \quad (7) \\ &\left. + AH_{ij} \int \beta_4(d_\ell, t) T_{ij}(t) dt \right\}, \end{aligned}$$

where $\beta_0(\cdot)$ is a smooth intercept, $\beta_1(\cdot)$ quantifies the smooth effect of young sows, $\beta_2(\cdot)$ describes the effect of the humidity, and $\beta_3(\cdot, t)$ and $\beta_4(\cdot, t)$ quantify the effect of the temperature at time t as well as the interaction between the temperature at time t and average humidity. We model $\beta_0(\cdot)$ using 20 univariate basis functions, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ using five univariate basis functions and $\beta_3(\cdot, \cdot)$ and $\beta_4(\cdot, \cdot)$ using tensor product of two univariate bases functions (total of 25 functions). Throughout the analysis, cubic B-spline bases are used and REML is used for estimating smoothing parameters. The estimated conditional distribution, denoted by $\hat{F}_{ij}(d)$, is monotized by fitting isotonic regression to $\{(d_\ell, \hat{F}_{ij}(d_\ell)) : \ell = 10, \dots, 90\}$; ten smallest and ten largest d_ℓ and the corresponding values of $\hat{F}_{ij}(d_\ell)$ are removed to avoid boundary effects. By abuse of

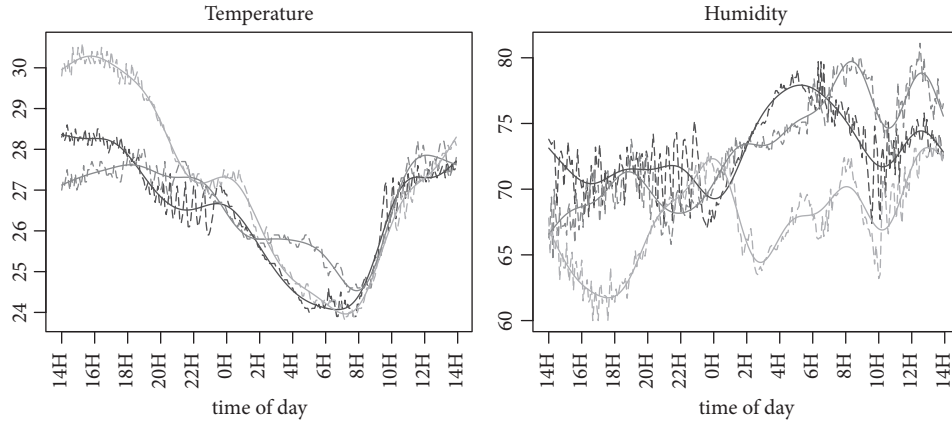


FIGURE 1: Temperature ($^{\circ}\text{C}$) and humidity (%) observed profiles (dashed) for three randomly selected days and the corresponding smoothed ones (solid); the x-axis begins at 14H (2PM).

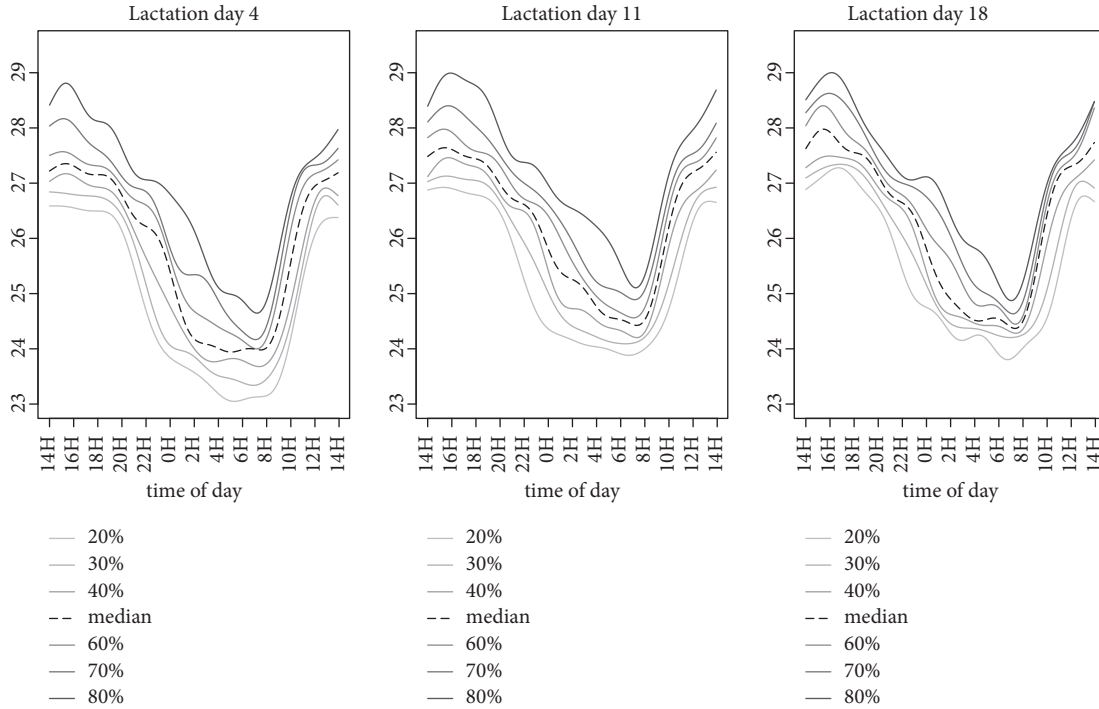


FIGURE 2: Temperature curves with which prediction of quantiles is made. Dashed black line is pointwise average of temperature curves and solid lines are pointwise quartiles; all curves are smoothed.

notation, $\widehat{F}_{ij}(d)$ denotes the resulting monotonized estimated distribution. Finally, we obtain estimated first quartiles, i.e., quantiles at $\tau = 0.25$ level, by inverting $\widehat{F}_{ij}(d)$, namely, $\widehat{Q}(\tau = 0.25 | T_{ij}(t), AH_{ij}, P_i) = \inf\{d : \widehat{F}_{ij}(d) \geq 0.25\}$.

To understand the relationship between the lactating sows feed intake and the thermal condition of the farrowing room, we systematically compare and study the predicted quantiles of two responses at combinations of different values of temperature, humidity, and parity. For each of three lactation days ($j = 3, 10, 17$) we consider three values of average humidity (first quartile, median, and third quartile) and two levels of parity (0 for older sows and 1 for younger sows). Based on the experimenters' interest, for the functional

covariate $T_{ij}(\cdot)$ we consider seven smooth temperature curves given in Figure 2. Each of these curves is obtained by first calculating pointwise quantiles of temperature at five-minute intervals for a specific level and then smoothing it; we considered quantiles levels $\eta = 0.2, 0.3, \dots, \text{and } 0.8$. In short, for each of three lactation days we obtain the first quartile of two responses for 42 different combinations (3 humidity values \times 2 parity levels \times 7 temperature curves) using the proposed method. To avoid extrapolation we ascertain that (i) there are reasonably many observed measurements at each of the combinations and (ii) bottom 25% of the responses are not dominantly from one of the parity group; see distribution of each response by the parity in Figure 3.

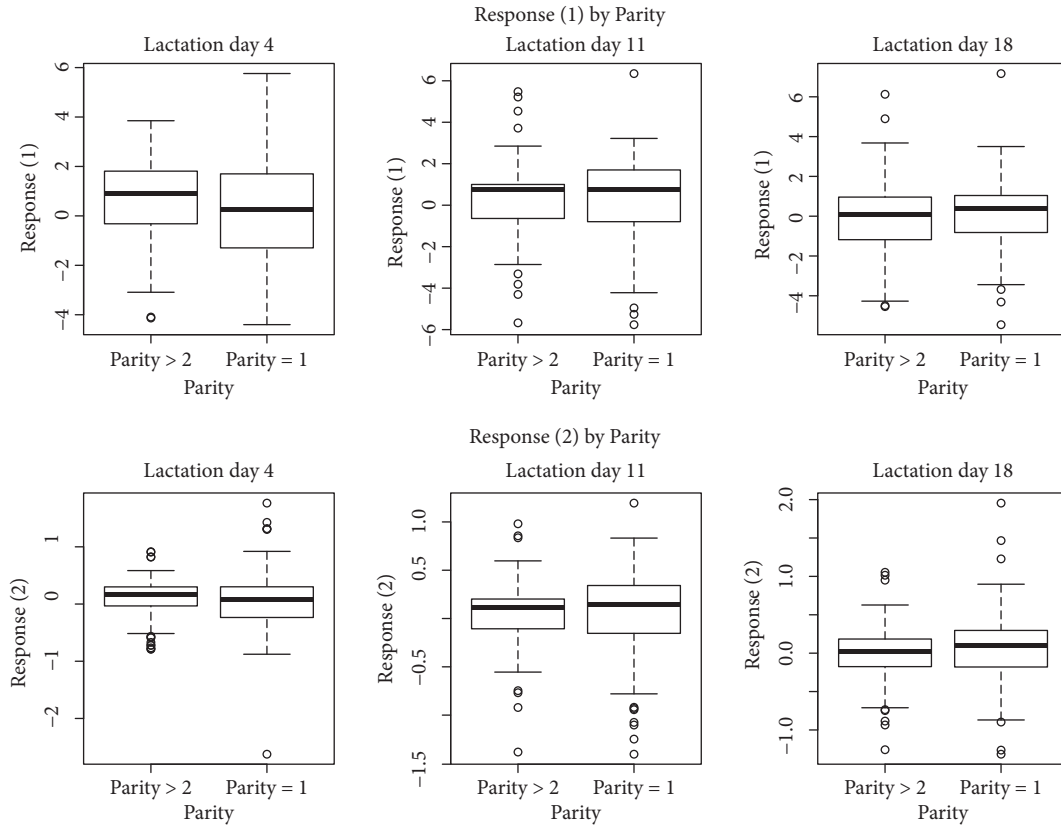


FIGURE 3: Top panels: back-to-back boxplots of the absolute change in feed intake at a specific day by parity. Bottom panels: back-to-back boxplots of the relative change in feed intake at a specific day by parity.

The resulting predicted quantiles are shown in Figure 4. Here we focus our discussion on predicted quantile of $\Delta_{i(j+1)}^{(2)}$ at quantile level $\tau = 0.25$ for lactation day 4 ($j = 3$)—the first plot of the second row in Figure 4. The results suggest that the feed intake of older sows (parity $P_i = 0$; grey lines) is less affected by high temperatures than younger sows (black lines); this finding is in agreement with Bloemhof et al. [42]. We also observe that the effects of humidity and temperature on feed intake change are strongly intertwined. For illustration, we focus on lactation day 4 ($j = 3$) again for younger sows (black lines). For medium humidity (dashed lines) their feed intake stays pretty constant as temperature increases, while for low and high humidity levels (solid and dotted lines, respectively) it changes with an opposite direction. Specifically when temperature increases, the predicted first quartile of $\Delta_{i(j+1)}^{(2)}$ increases for low humidity (solid line) whereas it decreases for high humidity (dotted line). Our results imply that high humidity (dotted line) is related to a negative impact of high temperature on feed intake while low humidity (solid line) alleviates it; and this finding is consistent with a previous study [45]. The analysis result suggests to keep low humidity levels in order to maintain healthy feed intake behavior, when ambient temperature is above 60th percentile; high humidity levels are desirable for cooler ambient temperature.

Interpretation of the other results is similar. While the effects of covariates on feed intake are less apparent toward

the end of lactation period, we still observe similar pattern across all three lactation days. For the 11th day ($j = 10$), the 25th quantile of the feed intake is predicted to decrease when the temperature stays below the 40th percentile, regardless of humidity level and sows age. However, it starts increasing with low humidity while it continues decreasing with high humidity when the temperature rises above the 40th percentile. Similarly, for the 18th day ($j = 17$) when the temperature rises above the 60th percentile, the predicted first quartile increases with low humidity while it decreases with high humidity. The effect of temperature on feed intake seems less obvious for lactation days 11 and 18 than for day 4; while the effect may be due to lactation day, it may also be a result of other factors, such as more fluctuation and variability in temperature curves on day 4 than on other two days (see Figure 2). Overall we conclude that high humidity and temperature affect the sows feed intake behavior negatively and young sows (parity one) are more sensitive to heat stress than older sows (higher parity), especially at the beginning of lactation period.

6. Discussion

The proposed modeling framework opens up a couple of future research directions. A first research avenue is to develop significance tests of null covariate effect. Testing for the null effect of a covariate on the conditional distribution of the response is equivalent to testing that the

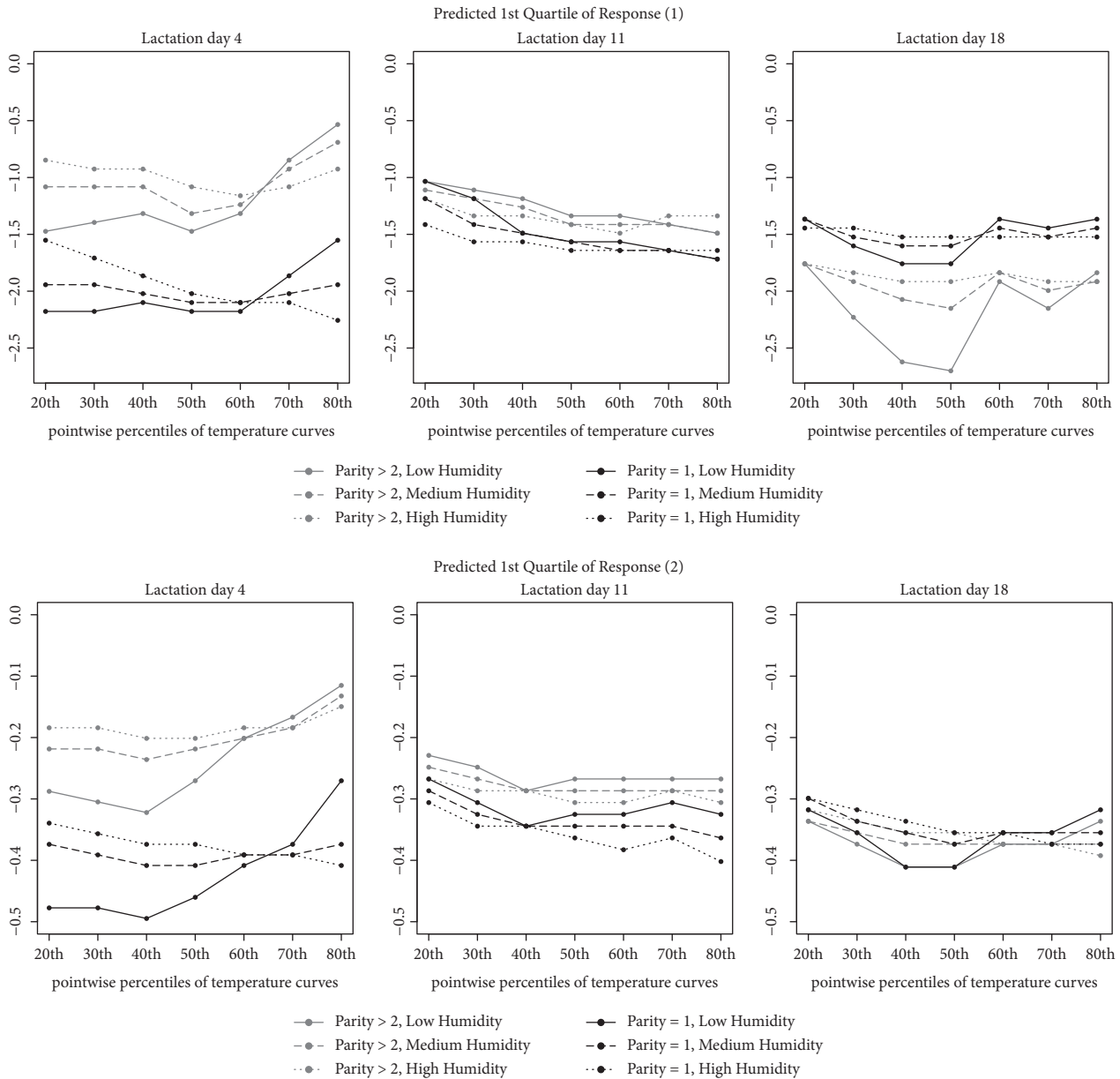


FIGURE 4: Displayed are the predicted quantiles of $\Delta_{i(j+1)}^{(1)}$ and $\Delta_{i(j+1)}^{(2)}$ for different parities, average humidity, and temperature levels. In each of all six panels, black thick lines correspond to the young sows ($P_i = 1$) and grey thin lines correspond to the old sows ($P_i = 0$). Line types indicate different average humidity levels; solid, dashed, and dotted correspond to low, medium, and high average humidity levels (given by the first quartile, median, and the third quartiles of AH_{ij}), respectively. The seven grids in x-axis of each panel correspond to the 7 temperature curves given in the respective panel of Figure 2.

corresponding regression coefficient function is equal to zero in the associated function-on-function mean regression model. Such significance tests have been studied when the functional response is continuous [30, 46]; however their study for binary-valued functional responses is an open problem in functional data literature and only recently has been considered in Chen et al. [47]. Another research avenue is to do variable selection in the setting where there are many scalar covariates and functional covariates. Many current applications collect data with increasing number of mixed covariates and selecting the ones that have an effect on the

conditional distribution of the response is very important. This problem is an active research area in functional mean regression where the response is normal [48, 49]. The proposed modeling framework has the potential to facilitate studying such problem.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interests.

Acknowledgments

The data used originated from work supported in part by the North Carolina Agricultural Foundation, Raleigh, NC. The authors acknowledge Zhen Han for preparing simulations. Staicu's research was supported by National Science Foundation DMS 0454942 and DMS 1454942 and National Institutes of Health grants R01 NS085211, R01 MH086633, and 5P01 CA142538-09.

Supplementary Materials

Section S1 provides additional simulation settings and results for the cases of having either a single scalar covariate or a single functional covariate. Additional results for the case of having a scalar covariate and a densely observed functional covariate are also included. Section S2 presents additional data analysis using the proposed method on the bike sharing dataset [50, 51]. (*Supplementary Materials*)

References

- [1] H. Cardot, F. Ferraty, and P. Sarda, "Functional linear model," *Statistics and Probability Letters*, vol. 45, no. 1, pp. 11–22, 1999.
- [2] G. M. James, "Generalized linear models with functional predictors," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 3, pp. 411–432, 2002.
- [3] J. Ramsay and B. Silverman, *Functional Data Analysis*, Springer, New York, NY, USA, 2005.
- [4] J. Ramsay and B. W. Silverman, *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York, NY, USA, 2002.
- [5] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*, Springer, New York, NY, USA, 2006.
- [6] F. Ferraty and P. Vieu, "Additive prediction and boosting for functional data," *Computational Statistics & Data Analysis*, vol. 53, no. 4, pp. 1400–1413, 2009.
- [7] J. Goldsmith, J. Bobb, C. M. Crainiceanu, B. Caffo, and D. Reich, "Penalized functional regression," *Journal of Computational and Graphical Statistics*, vol. 20, no. 4, pp. 830–851, 2011.
- [8] M. W. McLean, G. Hooker, A.-M. Staicu, F. Scheipl, and D. Ruppert, "Functional generalized additive models," *Journal of Computational and Graphical Statistics*, vol. 23, no. 1, pp. 249–269, 2014.
- [9] N. Quiniou and J. Noblet, "Influence of high ambient temperatures on performance of multiparous lactating sows," *Journal of Animal Science*, vol. 77, no. 8, pp. 2124–2134, 1999.
- [10] D. Renaudeau and J. Noblet, "Effects of exposure to high ambient temperature and dietary protein level on sow milk production and performance of piglets," *Journal of Animal Science*, vol. 79, no. 6, pp. 1540–1548, 2001.
- [11] N. St-Pierre, B. Cobanov, and G. Schnitkey, "Economic losses from heat stress by US livestock industries," *Journal of Dairy Science*, vol. 86, pp. E52–E77, 2003.
- [12] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [13] R. Koenker, *Quantile Regression*, vol. 38, Cambridge University Press, 2015.
- [14] H. D. Bondell, B. J. Reich, and H. Wang, "Noncrossing quantile regression curve estimation," *Biometrika*, vol. 97, no. 4, pp. 825–838, 2010.
- [15] H. Cardot, C. Crambes, and P. Sarda, "Quantile regression when the covariates are functions," *Journal of Nonparametric Statistics*, vol. 17, no. 7, pp. 841–856, 2005.
- [16] K. Kato, "Estimation in functional linear quantile regression," *The Annals of Statistics*, vol. 40, no. 6, pp. 3108–3136, 2012.
- [17] Q. Tang and L. Cheng, "Partial functional linear quantile regression," *Science China Mathematics*, vol. 57, no. 12, pp. 2589–2608, 2014.
- [18] Y. Lu, J. Du, and Z. Sun, "Functional partially linear quantile regression model," *Metrika*, vol. 77, no. 2, pp. 317–332, 2014.
- [19] D. Yu, L. Kong, and I. Mizera, "Partial functional linear quantile regression for neuroimaging data analysis," *Neurocomputing*, vol. 195, pp. 74–87, 2016.
- [20] F. Ferraty, A. Rabhi, and P. Vieu, "Conditional quantiles for dependent functional data with application to the climatic 'el niño' phenomenon," *Sankhya: The Indian Journal of Statistics*, vol. 67, no. 2, pp. 378–398, 2005.
- [21] K. Chen and H.-G. Müller, "Conditional quantile analysis when covariates are functions, with application to growth data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 74, no. 1, pp. 67–89, 2012.
- [22] F. Scheipl, J. Gertheiss, and S. Greven, "Generalized functional additive mixed models," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 1455–1492, 2016.
- [23] P. H. Eilers and B. D. Marx, "Flexible smoothing with b-splines and penalties," *Statistical Science*, pp. 89–102, 1996.
- [24] D. Ruppert, "Selecting the number of knots for penalized splines," *Journal of Computational and Graphical Statistics*, vol. 11, no. 4, pp. 735–757, 2002.
- [25] S. N. Wood, "Thin plate regression splines," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 65, no. 1, pp. 95–114, 2003.
- [26] S. Wood, *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2006.
- [27] A. E. Ivanescu, A.-M. Staicu, F. Scheipl, and S. Greven, "Penalized function-on-function regression," *Computational Statistics*, vol. 30, no. 2, pp. 539–568, 2015.
- [28] J. S. Kim, A.-M. Staicu, A. Maity, R. J. Carroll, and D. Ruppert, "Additive function-on-function regression," *Journal of Computational and Graphical Statistics*, vol. 27, no. 1, pp. 234–244, 2018.
- [29] P. Ng and M. Maechler, "A fast and efficient implementation of qualitatively constrained quantile smoothing splines," *Statistical Modelling. An International Journal*, vol. 7, no. 4, pp. 315–328, 2007.
- [30] J. T. Zhang and J. Chen, "Statistical inferences for functional data," *The Annals of Statistics*, vol. 35, no. 3, pp. 1052–1079, 2007.
- [31] F. Yao, H. Müller, and J. Wang, "Functional data analysis for sparse longitudinal data," *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 577–590, 2005.
- [32] L. Huang, F. Scheipl, J. Goldsmith et al., "R package refund: methodology for regression with functional data (version 0.1-13)," 2015, <https://cran.r-project.org/web/packages/refund/index.html>.
- [33] L. Xiao, V. Zipunnikov, D. Ruppert, and C. Crainiceanu, "Fast covariance estimation for high-dimensional functional data," *Statistics and Computing*, vol. 26, no. 1-2, pp. 409–421, 2016.

- [34] L. Xiao, C. Li, W. Checkley, and C. Crainiceanu, "Fast covariance estimation for sparse functional data," *Statistics and Computing*, vol. 28, no. 3, pp. 511–522, 2018.
- [35] L. Xiao, C. Li, W. Checkley, and C. Crainiceanu, "R package face: fast covariance estimation for sparse functional data (version 0.1-4)," 2018, <http://cran.r-project.org/web/packages/face/index.html>.
- [36] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, *Statistical inference under order restrictions: The theory and application of isotonic regression*, (No. 04; QA278. 7, B3.). New York: Wiley, 1972.
- [37] V. Chernozhukov, I. Fernandez-Val, and A. Galichon, "Improving point and interval estimators of monotone functions by rearrangement," *Biometrika*, vol. 96, no. 3, pp. 559–575, 2009.
- [38] D. S. Rosero, R. D. Boyd, M. McCulley, J. Odle, and E. van Heugten, "Essential fatty acid supplementation during lactation is required to maximize the subsequent reproductive performance of the modern sow," *Animal Reproduction Science*, vol. 168, pp. 151–163, 2016.
- [39] L. J. Johnston, M. Ellis, G. W. Libal et al., "Effect of room temperature and dietary amino acid concentration on performance of lactating sows. ncr-89 committee on swine management," *Journal of Animal Science*, vol. 77, no. 7, pp. 1638–1644, 1999.
- [40] D. Renaudeau, N. Quiniou, and J. Noblet, "Effects of exposure to high ambient temperature and dietary protein level on performance of multiparous lactating sows," *Journal of Animal Science*, vol. 79, no. 5, pp. 1240–1249, 2001.
- [41] J. L. Black, B. P. Mullan, M. L. Lorsch, and L. R. Giles, "Lactation in the sow during heat stress," *Livestock Production Science*, vol. 35, no. 1-2, pp. 153–170, 1993.
- [42] S. Bloemhof, P. K. Mathur, E. F. Knol, and E. H. van der Waaij, "Effect of daily environmental temperature on farrowing rate and total born in dam line sows," *Journal of Animal Science*, vol. 91, no. 6, pp. 2667–2679, 2013.
- [43] G. C. Nelson, M. W. Rosegrant, J. Koo et al., "Climate change: impact on agriculture and costs of adaptation," *International Food Policy Research Institute*, vol. 21, 2009.
- [44] J. M. Melillo, *Climate Change Impacts in the United States*, Us national climate assessment, Highlights, 2014.
- [45] R. Bergsma and S. Hermes, "Exploring breeding opportunities for reduced thermal sensitivity of feed intake in the lactating sow," *Journal of Animal Science*, vol. 90, no. 1, pp. 85–98, 2012.
- [46] Q. Shen and J. Faraway, "An f test for linear models with functional responses," *Statistica Sinica*, pp. 1239–1257, 2004.
- [47] S. Chen, L. Xiao, and A.-M. Staicu, "Testing for generalized scalar-on-function linear models," <https://arxiv.org/abs/1906.04889>, 2018.
- [48] J. Gertheiss, A. Maity, and A.-M. Staicu, "Variable selection in generalized functional linear models," *Stat*, vol. 2, no. 1, pp. 86–101, 2013.
- [49] Y. Chen, J. Goldsmith, and R. T. Ogden, "Variable selection in function-on-scalar regression," *Stat*, vol. 5, no. 1, pp. 88–101, 2016.
- [50] H. Fanaee-T and J. Gama, "Event labeling combining ensemble detectors and background knowledge," *Progress in Artificial Intelligence*, vol. 2, no. 2-3, pp. 113–127, 2014.
- [51] M. Lichman, "UCI machine learning repository," 2013.

Research Article

Fully Bayesian Estimation of Simultaneous Regression Quantiles under Asymmetric Laplace Distribution Specification

Josephine Merhi Bleik 

Sorbonne Universités, Université de Technologie de Compiègne, LMAC, Compiègne Cedex, France

Correspondence should be addressed to Josephine Merhi Bleik; josephine.merhi-bleik@utc.fr

Received 7 December 2018; Accepted 26 February 2019; Published 2 June 2019

Guest Editor: Rahim Alhamzawi

Copyright © 2019 Josephine Merhi Bleik. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we are interested in estimating several quantiles simultaneously in a regression context via the Bayesian approach. Assuming that the error term has an asymmetric Laplace distribution and using the relation between two distinct quantiles of this distribution, we propose a simple fully Bayesian method that satisfies the noncrossing property of quantiles. For implementation, we use Metropolis-Hastings within Gibbs algorithm to sample unknown parameters from their full conditional distribution. The performance and the competitiveness of the underlying method with other alternatives are shown in simulated examples.

1. Introduction

Quantile regression has received increasing attention from both theoretical and empirical points of view. It is a statistical procedure intended to estimate conditional quantile of a response distribution, given a set of covariates, and to conduct inference about them.

In regression, when we deal with highly skewed data distribution, traditional mean regression may not explore interesting aspects of the relationship between the response variable and the available covariates. Furthermore, in the presence of outliers, the evaluation of the response average becomes much more complicated. However, since a set of quantiles often provides a complete description of the response distribution conditionally to the given covariates, quantile regression offers a practical alternative to the traditional mean regression.

Quantile regression was initially introduced by Koenker and Bassett [1], where inference proceeds by minimizing a check loss function. Based on their theory, many other approaches have been emerged to study the regression quantiles in both frequentist and Bayesian points of view.

In the frequentist framework, different techniques have been developed in the literature to infer regression quantiles. To estimate linear regression quantiles, Koenker and Bassett

[1] used linear programming technique, whereas Gutenbrunner and Jurecková [2] considered the corresponding dual problem whose solution represents the quantile coefficients estimator. From theoretical point of view, asymptotic normality and consistency properties are proven in Koenker and Bassett [1], Zhou et al. [3], and Koenker and Xiao [4]. For nonparametric quantile regression, spline methods are commonly used for quantile regression fitting (see Nychka et al. [5], Koenker and Portnoy [6], and Koenker et al. [7]). Oh et al. [8] proposed, in univariate and bivariate settings, an approximation of the nonlinear optimization problem by a sequence of penalized least squares type nonparametric mean estimation problems. Koenker and Mizera [9] explored the estimation of triograms in bivariate setting. Koenker et al. [10] proposed an additive models estimation method for quantile regression based on selecting the smoothing parameter and constructing confidence bands for nonparametric components, in univariate and bivariate settings. The univariate regression quantile spline is extended to problems with several covariates in He and Ng [11] and Li et al. [12] where, in this latter, regression quantiles are considered in reproducing kernel Hilbert spaces (RKHS).

In the Bayesian framework, different methods have been successfully emerged to study quantile regression. This was pioneered by Yu and Moyeed [13], where the authors specified

the Bayesian quantile regression model and assumed that the error terms are iid according to the asymmetric Laplace distribution (ALD), denoted by $\mathcal{ALD}(\mu, \sigma = 1, p)$ with $\mu \in \mathbb{R}$ being its p -th quantile, σ is the scale parameter, and $p \in [0, 1]$ are the asymmetry parameters. They also showed that the Bayesian approach yields a proper full conditional distribution, even for an improper uniform prior on the quantile coefficients and they used the Metropolis-Hastings algorithm for implementation. Further, Kozumi and Kobayashi [14] and Bernardi et al. [15] considered that the scale parameter σ of the ALD is unknown. They used the location-scale mixture representation of the ALD, which enables developing a Gibbs sampling algorithm for the Bayesian quantile regression model. In the literature, we can also find other methods that extended quantile regression model. For censored data, Kobayashi et al. [16] proposed Tobit quantile regression with endogenous variables and Li and Bradic [17] developed regression adjustment for local adaptive representation of random forests. In continuous dependent variable case and for variable selection and coefficient estimation, Alhamzawi et al. [18] used adaptive LASSO quantile regression and, recently, Sayed-Ahmed [19] applied this approach for small sample sizes, whereas Abbas and Thaher [20] developed Bayesian adaptive Tobit regression. Benoit and Van den Poel [21] proposed Bayesian quantile regression methods for binary response data and Alhamzawi and Ali [22] adapted the quantile regression model to deal with longitudinal ordinal data.

All these methods mentioned above, from both frequentist and Bayesian points of view, studied single regression quantile. However, there are many applications where tackling several quantiles at the same time is needed. For this direction, if one uses methods that deal with single quantile regression to estimate regression quantiles separately, results may not be satisfactory. Indeed, it is known that the separate estimation of a set of quantiles may break the quantile monotonicity property (see He [23]), which means that quantile curves may cross.

Fortunately, many approaches have been proposed to overcome the crossing problem. In the frequentist framework, Liu and Wu [24] and Bondell et al. [25] considered monotonicity constraints for a fixed number of quantiles in both linear and nonparametric cases. Sangnier et al. [26] proposed a nonparametric method that estimates multiple conditional quantiles simultaneously when assuming that the quantile functions spans an RKHS. In this work, they choose an appropriate matrix-valued kernel taking into account the distance between quantiles.

On the other hand, in simultaneous Bayesian framework, Reich et al. [27] proposed a two-stage semiparametric method for linear quantile regression, which is one of the first methods addressing the crossing problem of quantiles. At the first stage, they used Koenker and Bassett's [1] method to estimate quantile coefficients. At the second stage, these coefficients are reestimated with Bernstein polynomial basis function with constraints on their coefficients. Later, Reich [28] proposed a fully Bayesian approach based on interpolating the quantile function on different quantile

levels using a piecewise Gaussian basis function. Tokdar et al. [29] proposed a simultaneous linear quantile regression method for a univariate explanatory variable with a bounded and convex domain. They showed that a linear quantile, $q_\tau(x) = \beta_0(\tau) + x\beta(\tau)$, is monotonically increasing if and only if $\beta_0(\tau)$ and $\beta(\tau)$ are linear combination of two monotonic increasing functions in $\tau \in [0, 1]$. Afterwards, Yang and Tokdar [30] extended this method to the multivariate explanatory case. Another two-stage method is proposed by Rodrigues and Fan [31]. At the first stage, they used the standard Bayesian quantile regression of Yu and Moyeed [13] fitted separately at the different quantile levels. They introduced induced quantiles, at the second stage, using a relation between quantiles of the ALD distribution. At the same stage, they used the Gaussian process regression that monotonicizes quantile functions by borrowing strength from the nearby ones. However, this method is still affected by the outputs of the first stage since different likelihoods are used for initial estimates. For nonparametric quantile regression, a cubic B-spline method is considered by Muggeo et al. [32], solving an L_2 penalisation problem with monotonicity constraint on spline coefficients. Many recent works on simultaneous Bayesian quantile regression can be found in literature such as Das and Ghosal [33], Das and Ghosal [34], Rodrigues et al. [35], and Petrella and Raponi [36].

In this paper, our contribution aims at developing a fully Bayesian approach that estimates multiple quantiles simultaneously in one step. Unlike the other Bayesian methods, our proposed method is not based on misspecification, i.e., the ALD is the real underlying response distribution. We used a relation between quantiles of the ALD distribution and Metropolis-Hastings within Gibbs algorithm for implementation. Our proposed method provides parallel quantile functions estimators and, thus, the noncrossing is a natural result. Our numerical results show, using specific empirical criteria, that the estimated regression quantiles respect the quantile level ordering.

The paper is organised as follows. Section 2 describes the method presenting the model and establishing the relation between quantiles of the ALD distribution used in the estimation procedure. The algorithm associated with the proposed method is given in Section 3, where we introduce the location-scale mixture representation of the ALD. In Section 4, simulated examples are provided to show the performance and the flexibility of the method, and the final section presents our conclusion.

2. Simultaneous Bayesian Estimation of Quantiles

2.1. Model. We consider the quantile regression model

$$Y_i = q_p(\mathbf{X}_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

with $(\epsilon_i)_i \stackrel{iid}{\sim} \mathcal{ALD}(\mu = 0, \sigma, p)$, where μ , σ , and p , respectively, are the location, the scale, and the asymmetry

parameters. This leads to assuming that the response variable Y_i , given the covariable \mathbf{X}_i , follows an ALD distribution:

$$Y_i | (q_p(\mathbf{X}_i), \sigma, p) \sim \mathcal{ALD}(q_p(\mathbf{X}_i), \sigma, p) \quad (2)$$

$$\forall i = 1, \dots, n,$$

whose probability density function (p.d.f.) is given by

$$f(y_i; q_p(\mathbf{x}_i), \sigma, p) = \frac{p(1-p)}{\sigma} \exp \left\{ -\frac{\rho_p(y_i - q_p(\mathbf{x}_i))}{\sigma} \right\}, \quad (3)$$

where $\rho_p(u) = u(p - \mathbb{1}_{u < 0})$ is the check loss function. The explanatory variables \mathbf{X}_i 's are supposed to be iid distributed according to an arbitrary continuous distribution, $\mathbb{P}_{\mathbf{X}}$, whose support is \mathcal{X}^d , $\mathcal{X} \subset \mathbb{R}$, $d \geq 1$. In what follows, the function $q_p(\cdot)$ and the parameters σ , p , and $\mathbb{P}_{\mathbf{X}}$ are supposed to be unknown.

In all subsequent sections, any quantity in bold represents a vector; capital letters denote random variables or vectors, whereas lowercase letters denote observed values of the corresponding random vectors.

2.2. Bayesian Procedure

2.2.1. Likelihood. To infer simultaneously s distinct quantiles with a fully Bayesian approach, say $q_{\tau_1}, \dots, q_{\tau_s}$ with distinct orders τ_1, \dots, τ_s , of the conditional distribution of $Y | \mathbf{X}$, one needs to characterize the likelihood through all these s unknown quantiles.

As explained below, this is done first by partitioning the whole sample into s subsamples, which requires a sufficient number of observations, second by using the relation between any two quantiles of the ALD distribution, and third by rewriting the whole likelihood through s terms, where the j -th term only depends on q_{τ_j} .

- (1) Consider the following partition of the whole sample through s subsamples:

$$(\mathbf{X}_{I_j}, \mathbf{Y}_{I_j}) = \{(\mathbf{X}_i, Y_i)_{i \in I_j}\}, \quad \forall j \in \{1, \dots, s\}, \quad (4)$$

where $I_j = \{(j-1)r + 1, \dots, jr\}$ with $I_j \cap I_k = \emptyset$, $\forall j \neq k$, and $j, k \in \{1, \dots, s\}$; suppose that $r = n/s$ is integer, without loss of generality.

- (2) To characterize the likelihood through all quantiles of interest, we use the relation that links any τ -th quantile to the p -th quantile of the underlying ALD distribution (see Yu and Zhang [37] and Rodrigues and Fan [31]):

$$q_{\tau}(\mathbf{X}) = q_p(\mathbf{X}) + \sigma g(\tau, p), \quad (5)$$

where $g(\tau, p) = (1/(1-p)) \log(\tau/p) \mathbb{1}_{0 < \tau \leq p} - (1/p) \log((1-\tau)/(1-p)) \mathbb{1}_{p < \tau < 1}$.

- (3) From (5), we rewrite the model given by (1) as follows:

$$\mathbf{Y}_{I_j} = \mathbf{q}_{\tau_j, I_j} - \sigma \mathbf{g}_{\tau_j, p} + \boldsymbol{\epsilon}_{I_j}, \quad \forall j = 1, \dots, s, \quad (6)$$

where

$$\begin{aligned} \mathbf{Y}_{I_j} &= (Y_{i_j})_{i_j \in I_j}, \\ \mathbf{q}_{\tau_j, I_j} &= (q_{\tau_j}(\mathbf{X}_{i_j}))_{i_j \in I_j}, \\ \mathbf{g}_{\tau_j, p} &= (g(\tau_j, p) \mathbb{1}_{i_j \in I_j})_{i_j \in I_j}, \\ \boldsymbol{\epsilon}_{I_j} &= (\epsilon_{i_j})_{i_j \in I_j}. \end{aligned} \quad (7)$$

Then, the likelihood associated with the model given by (1) is rewritten as the product of s likelihoods; each one only depends on one subsample:

$$\begin{aligned} L(\mathbf{q}_{\tau_1, I_1}, \dots, \mathbf{q}_{\tau_s, I_s}, \sigma, p; \mathbf{x}_{I_1}, \dots, \mathbf{x}_{I_s}, \mathbf{y}_{I_1}, \dots, \mathbf{y}_{I_s}) \\ = \prod_{j=1}^s \prod_{i_j \in I_j} \frac{p(1-p)}{\sigma} \\ \cdot \exp \left\{ -\frac{\rho_p(y_{i_j} - (q_{\tau_j}(\mathbf{x}_{i_j}) - \sigma g(\tau_j, p)))}{\sigma} \right\}. \end{aligned} \quad (8)$$

2.2.2. Priors. In this section, we specify the prior distributions for all unknown quantities. It is worth mentioning that $\mathbb{P}_{\mathbf{X}}$ is unknown even it is not our primary interest, and therefore it should require a prior; but since any prior on $\mathbb{P}_{\mathbf{X}}$ that is independent of the prior on $((\mathbf{q}_{\tau_j, I_j})_{j=1, \dots, s}, \sigma, p)$ would disappear upon marginalization of the posterior of $(\mathbb{P}_{\mathbf{X}}, (\mathbf{q}_{\tau_j, I_j})_{j=1, \dots, s}, \sigma, p)$ relatively to $\mathbb{P}_{\mathbf{X}}$, we drop it in the sequel. Thus, it suffices to choose a prior distribution for \mathbf{q}_{τ_j, I_j} , for $j = 1, \dots, s$, σ , and p .

- (1) For the quantiles, we distinguish between the linear and nonlinear cases.

(a) *Linear case.* $q_{\tau_j}(\mathbf{X}) = \beta_{\tau_j}^{(0)} + \mathbf{X}^T \boldsymbol{\beta}_{\tau_j}$, $j \in \{1, \dots, s\}$, where \mathbf{X}^T denotes the transpose of \mathbf{X} and $\boldsymbol{\beta}_{\tau_j} \in \mathbb{R}^d$.

Due to (5), one should note that distinct quantiles differ only by the intercept $\beta_{\tau_j}^{(0)}$, so that all quantiles of interest have the same slope ($\boldsymbol{\beta}_{\tau_1} = \dots = \boldsymbol{\beta}_{\tau_s} = \boldsymbol{\beta}_p$); then we consider the unknown vector parameter $\check{\boldsymbol{\beta}} = (\beta_{\tau_1}^{(0)}, \dots, \beta_{\tau_s}^{(0)}, \boldsymbol{\beta}_p)$ of dimension $s + d$ with Gaussian prior,

$$\check{\boldsymbol{\beta}} \sim \mathcal{N}_{s+d}(0, \check{\Sigma}_0), \quad (9)$$

with $\check{\Sigma}_0$ being a positive definite square matrix of dimension $s + d$.

(b) *Nonlinear case.* We define the function h on $[0, 1] \times \mathcal{X}^d$, by $h(\tau, \mathbf{x}) = q_\tau(\mathbf{x})$. We put a Gaussian process prior on h , $h \sim \mathcal{GP}(0, k)$, with a zero-mean function and a covariance function $k: [0, 1]^2 \times \mathcal{X}^{2d} \rightarrow \mathbb{R}$; following Sangnier et al. [26], we choose k to be decomposable and its form is given by

$$k((\tau_i, \mathbf{x}), (\tau_j, \mathbf{x}')) = k_x(\mathbf{x}, \mathbf{x}') \exp(-c(\tau_j - \tau_i)^2), \quad (10)$$

where $k_x(\mathbf{x}, \mathbf{x}') = \exp(-b\|\mathbf{x} - \mathbf{x}'\|^2)$, b and c are positive hyperparameters, and $\|\cdot\|$ is the Euclidean norm on \mathbb{R}^d . Here, $k((\tau_i, \mathbf{x}), (\tau_j, \mathbf{x}'))$, $i \neq j$, encodes the relation between the conditional quantiles $q_{\tau_i}(\mathbf{x})$ and $q_{\tau_j}(\mathbf{x}')$. As explained by Sangnier et al. [26], if $c \rightarrow 0$, the quantile curves are parallel so they do not cross. If $c \rightarrow +\infty$, the quantiles are learned independently and then they may cross. Thus, the choice of c is important to control the crossing.

- (2) $\sigma \sim \mathcal{IG}(a_0, b_0)$ and $a_0 > 0, b_0 > 0$, where \mathcal{IG} denotes the inverse Gamma distribution with positive hyperparameters a_0 and b_0 .
- (3) $p \sim \text{Beta}(\alpha_0, \lambda_0)$ and $\alpha_0, \lambda_0 > 0$.

3. Computations

To compute the full conditional distributions, we shall make use of the location-scale mixture of the ALD distribution (see Kozumi and Kobayashi [14]).

Let ω be an exponential latent variable with parameter $1/\sigma$, denoted by $\mathcal{E}(1/\sigma)$, and let Z be a standard normal variable such that ω and Z are independent. If ϵ has an ALD distribution, $\epsilon \sim \mathcal{ALD}(0, \sigma, p)$; then, it can be represented as a mixture of normal variable:

$$\epsilon = \gamma_p \omega + \delta_p \sqrt{\sigma \omega} z, \quad (11)$$

where $\gamma_p = (1 - 2p)/p(1 - p)$ and $\delta_p^2 = 2/p(1 - p)$.

Exploiting this augmented data structure, the model defined by the system of equations given by (6) admits, conditionally on $\omega_{I_j} = (\omega_i)_{i \in I_j} \stackrel{iid}{\sim} \mathcal{E}(1/\sigma)$, the following Gaussian representation:

$$\mathbf{Y}_{I_j} = \mathbf{q}_{\tau_j, I_j} - \sigma \mathbf{g}_{\tau_j, p} + \gamma_p \omega_{I_j} + \delta_p \sqrt{\sigma \omega_{I_j}} \mathbf{z}_{I_j}, \quad (12)$$

$j = 1, \dots, s,$

where $\mathbf{z}_{I_j} = (z_i)_{i \in I_j} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$.

3.1. Full Conditional Distributions. From the equations given in (12) and the priors defined in Section 2.2.2, we follow Bernardi et al. [15] to derive the full conditional distributions of $\omega_n = (\omega_1, \dots, \omega_n)$ and $(\mathbf{q}_{\tau_j, I_j})_{j=1, \dots, s}$. Yet, the full conditionals of σ and p are not tractable and this is carried out by including a Metropolis-Hasting step to the Gibbs sampler algorithm.

- (1) For all $j = 1, \dots, s$, as well as $i \in I_j$, set $\nu_i = \omega_i^{-1}$, and considering the distribution $\omega_i \sim \mathcal{E}(1/\sigma)$ as a prior on ω_i ,

$$\nu_i | (y_i, q_{\tau_j}(\mathbf{x}_i), \sigma, p) \sim \mathcal{IGauss}(\Psi_i, \phi_p) \quad (13)$$

with

$$\Psi_i = \sqrt{\frac{\gamma_p^2 + 2\delta_p^2}{(y_i - q_{\tau_j}(\mathbf{x}_i) + \sigma g(\tau_j, p))^2}}, \quad (14)$$

$$\phi_p = \frac{\gamma_p^2 + 2\delta_p^2}{\delta_p^2 \sigma},$$

where \mathcal{IGauss} stands for the Inverse Gaussian distribution with $\Psi_i > 0$ and $\phi_p > 0$ as location and shape parameters.

- (2) The conjugate Gaussian prior on quantiles provides Gaussian full conditional distributions in both linear and nonlinear cases.

(a) *Linear case.* Let us introduce some notations:

- (i) $\mathbf{Y}_n = (Y_i)_{1 \leq i \leq n}$,
- (ii) $\check{\mathbf{X}} = (\check{X}_{i,l})_{1 \leq i \leq n, 1 \leq l \leq s+d}$ is the $n \times (s+d)$ design matrix defined by

$$\check{X}_{i,l} = \begin{cases} \mathbb{1}_{i \in I_l} & \text{if } l \in 1, \dots, s, \\ X_{i,l-s} & \text{if } l \in s+1, \dots, s+d, \end{cases} \quad (15)$$

- (iii) $\mathbf{g}_n = (g(\tau_j, p) \mathbb{1}_{i \in I_j})_{1 \leq j \leq s}$.

This allows to rewrite the system of equations (12) into the following format:

$$\mathbf{Y}_n = \check{\mathbf{X}} \check{\boldsymbol{\beta}} - \sigma \mathbf{g}_n + \gamma_p \omega_n + \boldsymbol{\epsilon}_n^*, \quad (16)$$

with $\boldsymbol{\epsilon}_n^* \sim \mathcal{N}(0, \Sigma^*)$ and where $\Sigma^* = \text{diag}(\delta_p^2 \sigma \omega_n)$. With the zero-mean Gaussian prior distribution, the full conditional distribution on $\check{\boldsymbol{\beta}}$ is then Gaussian:

$$\check{\boldsymbol{\beta}} | (\check{\mathbf{x}}, \mathbf{y}_n, \omega_n, \sigma, p) \sim \mathcal{N}_{s+d}(\hat{\boldsymbol{\mu}}_{\check{\boldsymbol{\beta}}}, \hat{\Sigma}_{\check{\boldsymbol{\beta}}}) \quad (17)$$

with

$$\hat{\boldsymbol{\mu}}_{\check{\boldsymbol{\beta}}} = (\check{\mathbf{x}}^\top \Sigma^* \check{\mathbf{x}} + \check{\Sigma}_0^{-1})^{-1} \check{\mathbf{x}}^\top \Sigma^* (\mathbf{y}_n + \sigma \mathbf{g}_n - \gamma_p \omega_n), \quad (18)$$

$$\hat{\Sigma}_{\check{\boldsymbol{\beta}}} = (\check{\mathbf{x}}^\top \Sigma^* \check{\mathbf{x}} + \check{\Sigma}_0^{-1})^{-1}.$$

- (b) *Nonlinear case.* We shall use other extra notations:

$$\mathbf{h}_n = (\mathbf{h}_{\tau_j, I_j})_{j=1, \dots, s} = \left((h(\tau_j, \mathbf{x}_i))_{i \in I_j} \right)_{j=1, \dots, s}, \quad (19)$$

$$\mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) = \left(\left(k(\tau_j, \mathbf{x}_{i_j}) \right)_{i_j \in I_j} \right)_{j=1, \dots, s}.$$

1: **Initialization** $t = 0$: $\sigma^{(0)} = \sigma_0$, $p^{(0)} = p_0$.

2: **Loop** $t = t + 1$:

1. (i) $\sigma_{prop} | \sigma^{(t)} \sim \mathcal{N}_{]0, \infty[}(\sigma^{(t)}, sd_\sigma)$,
- (ii) $\alpha_\sigma = \min \{1, P_\sigma Q_\sigma\}$
- (iii) $u_1 \sim \mathcal{U}_{[0,1]} \implies \sigma^{(t+1)} = \begin{cases} \sigma_{prop} & \text{if } \alpha > u_1 \\ \sigma^{(t)} & \text{otherwise} \end{cases}$

2. (i) $p_{prop} | p^{(t)} \sim \mathcal{N}_{]0,1[}(p^{(t)}, sd_p)$,
- (ii) $\alpha_p = \min \{1, P_p Q_p\}$
- (iii) $u_2 \sim \mathcal{U}_{[0,1]} \implies p^{(t+1)} = \begin{cases} p_{prop} & \text{if } \alpha > u_2 \\ p^{(t)} & \text{otherwise} \end{cases}$

ALGORITHM 1: Metropolis-Hastings step.

This allows to rewrite the system of equations (12) into the following vector format:

$$\mathbf{Y}_n = \mathbf{h}_n - \sigma \mathbf{g}_n + \gamma_p \boldsymbol{\omega}_n + \boldsymbol{\epsilon}_n^*. \quad (20)$$

Combining (20) with the Gaussian process prior on h leads to the following joint distribution of $(\mathbf{Y}_n, \mathbf{h}_n)$ conditional on $(\mathbf{x}, \boldsymbol{\omega}_n, \sigma, p)$:

$$\begin{pmatrix} \mathbf{Y}_n \\ \mathbf{h}_n \end{pmatrix} \sim \mathcal{N}_{2n} \left(\begin{pmatrix} -\sigma \mathbf{g}_n + \gamma_p \boldsymbol{\omega}_n \\ \mathbf{0}_n \end{pmatrix}, \begin{pmatrix} \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) + \Sigma^* & \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) \\ \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) & \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) \end{pmatrix} \right). \quad (21)$$

Finally, classical calculations lead to the desired full conditional distribution:

$$\mathbf{h}_n | (\mathbf{x}, \mathbf{y}_n, \boldsymbol{\omega}_n, \sigma, p) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_n, \hat{\Sigma}_h), \quad (22)$$

where

$$\hat{\boldsymbol{\mu}}_n = \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) (\mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) + \Sigma^*)^{-1} (\mathbf{y}_n - \gamma_p \boldsymbol{\omega}_n + \sigma \mathbf{g}_n), \quad (23)$$

$$\hat{\Sigma}_h = \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) - \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) (\mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}) + \Sigma^*)^{-1} \mathbf{k}_n(\boldsymbol{\tau}, \mathbf{x}).$$

(3) The full conditional distribution, π_σ of $\sigma | \mathbf{x}, \boldsymbol{\omega}_n, p$, is proportional to

$$\frac{1}{(\sqrt{\sigma})^{n+a_0+1}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \mathbf{g}_n - \gamma_p \boldsymbol{\omega}_n)^\top \cdot (\Sigma^*)^{-1} (\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \mathbf{g}_n - \gamma_p \boldsymbol{\omega}_n) \right\}, \quad (24)$$

where $\mathbf{q}_{\boldsymbol{\tau},n} = (\mathbf{q}_{\boldsymbol{\tau},j})_{j=1,\dots,s}$.

(4) The full conditional, π_p of $p | \mathbf{x}, \boldsymbol{\omega}_n, \sigma$, is proportional to

$$p^{\alpha_0} (1-p)^{\lambda_0} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \mathbf{g}_n - \gamma_p \boldsymbol{\omega}_n)^\top \cdot (\Sigma^*)^{-1} (\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \mathbf{g}_n - \gamma_p \boldsymbol{\omega}_n) \right\}. \quad (25)$$

3.2. *Algorithm.* Due to (24) and (25), it is not possible to generate σ and p directly from their full conditional distribution. Therefore, they are simulated by incorporating a random walk Metropolis-Hastings step within Gibbs, as described in Algorithm 1.

For a subset $E \subset \mathbb{R}$, denote by $\mathcal{N}_E(\cdot, \cdot)$ the truncated version on E of the corresponding Gaussian distribution. Note that the choice of the truncated Gaussian distribution is classical.

It is worth to mention that the values of the scale parameters sd_σ and sd_p are calibrated to achieve the equilibrium of the random walk Metropolis-Hastings step quickly; in fact, they are chosen neither too small nor too large so that the acceptance rate becomes practically stable.

The conditional posterior ratios of σ and p are given by

$$P_\sigma = \frac{\pi_\sigma \left(\sigma_{prop} | \mathbf{y}_n, \mathbf{x}_n, \left((\mathbf{q}_{\boldsymbol{\tau},j})_{j=1,\dots,s} \right)^{(t+1)}, \boldsymbol{\omega}_n^{(t+1)}, p^{(t)} \right)}{\pi_\sigma \left(\sigma^{(t)} | \mathbf{y}_n, \mathbf{x}_n, \left((\mathbf{q}_{\boldsymbol{\tau},j})_{j=1,\dots,s} \right)^{(t+1)}, \boldsymbol{\omega}_n^{(t+1)}, p^{(t)} \right)} \quad (26)$$

$$P_p = \frac{\pi_p \left(p_{prop} | \mathbf{y}_n, \mathbf{x}_n, \left((\mathbf{q}_{\boldsymbol{\tau},j})_{j=1,\dots,s} \right)^{(t+1)}, \boldsymbol{\omega}_n^{(t+1)}, p^{(t)} \right)}{\pi_p \left(p^{(t)} | \mathbf{y}_n, \mathbf{x}_n, \left((\mathbf{q}_{\boldsymbol{\tau},j})_{j=1,\dots,s} \right)^{(t+1)}, \boldsymbol{\omega}_n^{(t+1)}, p^{(t)} \right)},$$

where $\pi_\sigma(\cdot | \cdot)$ and $\pi_p(\cdot | \cdot)$, are respectively, given up to a constant in (24) and (25), and the transition probabilities are given by

$$Q_\sigma = \frac{f_{\mathcal{N}_{]0,\infty[}(\sigma^{(t)}, sd_\sigma)}}{f_{\mathcal{N}_{]0,\infty[}(\sigma_{prop}, sd_\sigma)}} = \frac{1 - \phi(-\sigma_{prop}/sd_\sigma)}{1 - \phi(-\sigma^{(t)}/sd_\sigma)},$$

$$Q_p = \frac{f_{\mathcal{N}_{]0,1[}(p^{(t)}, sd_p)}}{f_{\mathcal{N}_{]0,1[}(p_{prop}, sd_p)}} = \frac{\phi((1-p_{prop})/sd_p) - \phi(-p_{prop}/sd_p)}{\phi((1-p^{(t)})/sd_p) - \phi(-p^{(t)}/sd_p)}, \quad (27)$$

where $f_{\mathcal{N}_E(\cdot, \cdot)}$ denotes the pdf of $\mathcal{N}_E(\cdot, \cdot)$ and $\phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

4. Simulation Study

In this section, we study the performance of our method in both linear and nonlinear quantile regressions. For the model given by (1), we shall consider three different designs for the p -th quantile:

- (1) Univariate linear quantile: $q_p(X) = 1 + 2X$, with $X \sim \mathcal{U}_{[-1,1]}$.
- (2) Multivariate linear quantile: $q_p(\mathbf{X}) = 1 + \mathbf{X}^\top \boldsymbol{\beta}_p$, with $\boldsymbol{\beta}_p \in \mathbb{R}^d$, $d = 10$, and either $X_l \sim \mathcal{U}_{[-1,1]}$ or $X_l \sim \mathcal{N}(0, 1)$, $\forall l \in \{1, \dots, d\}$.
- (3) Nonlinear quantile: for $X \sim \mathcal{U}_{[0,1]}$,

$$\begin{aligned}
 q_p(X) &= \cos\left(\frac{5}{2}\pi X \exp\left\{-\frac{3}{2}X\right\}\right) \\
 &+ \left[\frac{1}{4} \exp\{2(X - 0.5)\} - \exp\{-1\}\left(\frac{1}{4} + \frac{X}{2}\right)\right] \\
 &\cdot \mathbb{1}_{(X < 0.5)} + \left[\frac{1}{4} \exp\{-2(X - 0.5)\} \right. \\
 &\left. - \left(\frac{1}{4} + \frac{X}{2}\right) \exp\{-1\} - \left(\frac{1}{2} - X\right)\right] \mathbb{1}_{(X \geq 0.5)}.
 \end{aligned} \tag{28}$$

For all designs, we generate independently 300 observations issued from the model defined in (1). All runs of Metropolis-Hastings within Gibbs algorithm consist in 20000 iterations, one-third of which are burn-in. The prior hyperparameters are chosen as follows: for the inverse Gamma of σ , $a_0 = 1$ and $b_0 = 0.01$; for the Beta prior of p , $\alpha_0 = 2$ and $\lambda_0 = 2$ and $\check{\Sigma}_0$ is set to be the identity matrix for linear quantile and $c = 0.1$ and $b = 5$ for the nonlinear case. The choice of $c = 0.1$ and $b = 5$ is typical; indeed, whatever the value of τ is, $c = 0.1$ minimizes the empirical root mean integrated square error, RMISE; that is, $c = \operatorname{argmin} \sqrt{(1/n) \sum_{i=1}^n (q_\tau(\mathbf{X}_i) - \hat{q}_\tau(\mathbf{X}_i))^2}$, where \hat{q}_τ stands for the posterior mean quantile regression. In order to test the robustness of our procedure with respect to the model parameters, different values of σ and p are considered for the three designs.

The first design is a straightforward example and is carried out just to check the convergence of the algorithm from different tools: \hat{R} of Gelman and Rubin diagnostic, the autocorrelation analysis, and the posterior plots of the various parameters.

Through the second design, we use the crossing loss criterion (see Sangnier et al. [26]) to compare the performance, against crossing, of our method, denoted by “SBQR” with other approaches: the frequentist single quantile method of Koenker and Bassett [1], denoted by “K&B”, the Bayesian single quantile regression method of Yu and Moyeed [13], denoted by “Y&M”, and the simultaneous Bayesian method of Reich and Smith [38], denoted by “R&S”; in addition,

we use the RMISE to evaluate the estimation performance. These other methods are performed using available codes in R Core Team (2017): *rq* function available in *quantreg* package Koenker [39] for “K&B”, *bayesQR* function in *bayesQR* package Benoit et al. [40] for “Y&M”, and *qreg* function in *BSquare* package Smith et al. [41] for “R&S”.

For design 3, we compare our “SBQR” method with both the nonparametric quantile regression method of Muggeo et al. [32], denoted by “M&ST”, and the simultaneous noncrossing method of Rodrigues and Fan [31], denoted by “F&R”. We have implemented “M&ST” with *quantregGrowth* R package (see Muggeo [42]) and “F&R” with an own made code in R. We use features of the RMISE criterion to show how well “SBQR” performs among the other considered methods.

We note that the number of observations, $n = 300$, is chosen carefully to illustrate the intended objectives in all designs.

4.1. Design 1: Univariate Linear Quantile Regression. The iid sample $\epsilon_1, \dots, \epsilon_n$ is generated according to $\mathcal{A}\mathcal{L}\mathcal{D}(0, \sigma = 0.1, p = 0.25)$. We propose to infer three quantiles that are close, namely, the quantiles of order $\tau = 0.2, 0.3$ and 0.4 , respectively.

We fix $sd_\sigma = sd_p = 0.1$. To evaluate the convergence of our algorithm, we use three different seeds and parameter starting values to run three different chains and calculate \hat{R} of Gelman convergence diagnostic. Besides, we use other convergence diagnostics such as the autocorrelation analysis and the posterior plots.

As shown in the top panel of Figures 1, 2, and 3, all posterior distributions shrink at the true parameters value. Furthermore, in the middle panel of Figures 1, 2, and 3, the decrease of the empirical autocorrelation of posterior samples proves that the underlying chains are stationary. The bottom panels of Figures 1, 2, and 3 show that \hat{R} goes to 1 through the iterations, which confirms the convergence of the algorithm.

4.2. Design 2: Multivariate Linear Quantile Regression. The second design is dedicated to the multivariate linear case; hence, we consider the model given by (1) with $\epsilon_1, \dots, \epsilon_n \stackrel{iid}{\sim} \mathcal{A}\mathcal{L}\mathcal{D}(\mu = 0, \sigma = 0.5, p = 0.25)$, $\mathbf{X} \in \mathbb{R}^d$, $d = 10$, $X_k \sim \mathcal{U}_{[-1,1]}$, $k = 1, \dots, d$, and $\boldsymbol{\beta}_p = (1.6, 2.2, 2.8, 3.4, 4, 4.5, 5.1, 5.7, 6.3, 6.9)$.

As commonly known, crossing quantiles is a practical problem that often occurs when there is a large number of covariates. We propose to infer two close quantiles of order $\tau_1 = 0.1$ and $\tau_2 = 0.12$ and to study the crossing throughout the following four methods: “K&B”, “Y&M”, “R&S”, and “SBQR”.

To achieve the desired posterior distribution through MCMC methods, we perform Y&M and R&S with different number of iterations: 1000 for Y&M and 10000 for R&S. For “R&S”, we use the logistic base distribution with 4 basis functions. For “SBQR”, we fix $sd_\sigma = sd_p = 0.01$ and $\check{\Sigma}_0$ to be the identity matrix.

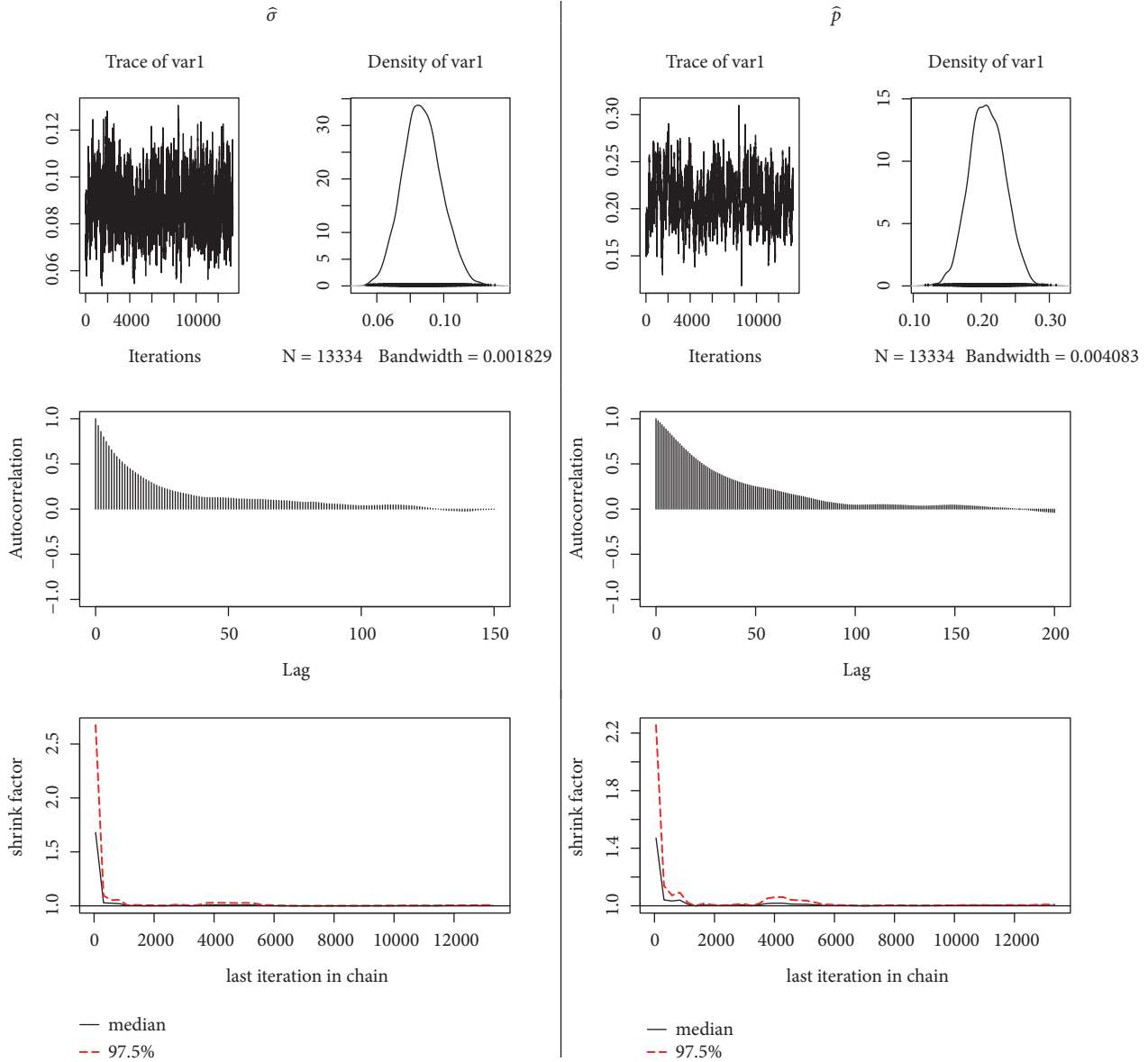


FIGURE 1: Trace and density plot (top), autocorrelation plot (middle), and \hat{R} evolution through iterations (bottom) of $\hat{\sigma}$ (left panel) and \hat{p} (right panel).

To compare the methods, we make use of the crossing loss criterion (see Sangnier et al. [26]) that measures how far $\hat{q}_{0.12}(\mathbf{X})$ goes below $\hat{q}_{0.10}(\mathbf{X})$:

$$crossing\ loss = \frac{1}{n} \sum_{i=1}^n \max(0, \hat{q}_{0.10}(\mathbf{X}_i) - \hat{q}_{0.12}(\mathbf{X}_i)). \quad (29)$$

For a given approach, the less the crossing loss is, the better is the method. As shown in Table 1, the crossing loss, by “K&B”, is significantly of 0.43%, which corresponds to 34 data points of $\hat{q}_{0.1}(\mathbf{X})$ which are above $\hat{q}_{0.12}(\mathbf{X})$. This percentage is considerably weakened when applying the separate Bayesian method “Y&M” (0.38%) but still has crossing quantiles (26 data points of $\hat{q}_{0.1}(\mathbf{X})$ are above $\hat{q}_{0.12}(\mathbf{X})$). However, for simultaneous estimation methods, as our proposed “SBQR” or “R&S” methods, the crossing loss becomes zero; this

means that the simultaneous quantile estimation has the potential to make quantile crossing vanish. Moreover, the estimated quantiles are in the right order. While simultaneous approaches control the monotonicity property of quantiles in a certain sense, separate approaches do not, and they provide more than 8% of violation according to results previously discussed (11.33% of violation by “K&B” and 8.66% of violation by “Y&M”).

To go further, we compute the RMISE for each quantile order and for all methods. The “SBQR” has roughly the smallest RMISE in both quantile levels $\tau = 0.1$ and $\tau = 0.12$. The fact is that “R&S” may oversmooth when estimating the quantiles simultaneously and here, for this linear case, the oversmoothness impacts the flexibility of the method (high RMISE).

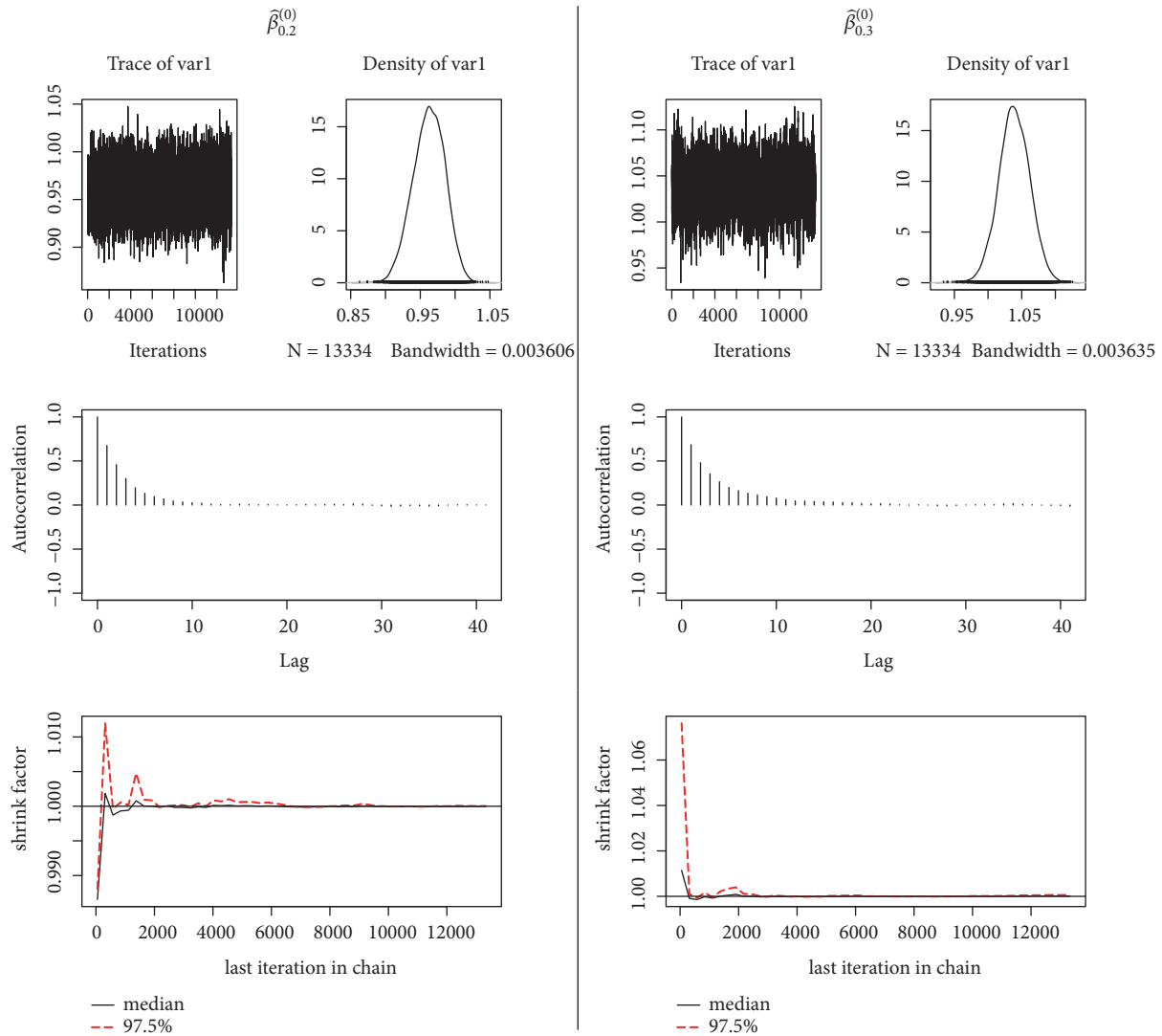


FIGURE 2: Trace and density plot (top), autocorrelation plot (middle), and \hat{R} evolution (bottom) of the intercept $\hat{\beta}_{0.2}^{(0)}$ (left panel) and the slope $\hat{\beta}_{0.3}^{(0)}$ of the 0.2–th and 0.3–th quantile, respectively.

TABLE 1: Table of criteria: crossing loss and RMISE.

Method	nb. crossing	crossing loss	RMISE (0.10)	RMISE (0.12)
“K&B”	34	0.0043	0.3565	0.3218
“Y&M”	26	0.0038	0.3344	0.3362
“R&S”	0	0	0.2449	0.2849
“SBQR”	0	0	0.1989	0.2383

To see if the covariable support has an impact on results, we consider another simulation set in which $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}_d, Id_d)$, that is, $\text{support}(\mathbf{X}) = \mathbb{R}^d$. Table 2 shows that “SBQR” behaves like in the previous example: a zero crossing loss and the smallest RMISE among all the methods.

We also consider another case with a different pair of quantile orders: $\tau_1 = 0.7$ and $\tau_2 = 0.8$. We turn back to the support $[-1, 1]^d$ for \mathbf{X} . Table 3 shows similar results as the ones obtained for $\tau_1 = 0.10$ and $\tau_2 = 0.12$. Thus, “SBQR”

still has the best behavior among the other methods in terms of crossing loss and RMISE.

It should be noted that the estimation of σ and p by “SBQR” is quite good, since their estimated values are near the true ones in the different treated cases.

4.3. Design 3: Nonparametric Quantile Regression. Considering the third design with $X \sim \mathcal{U}_{[0,1]}$ and $\epsilon \sim \mathcal{ALD}(\mu = 0, \sigma = 0.05, p = 0.75)$, we are interested in estimating quantile functions for orders $\tau = 0.10, 0.12, 0.15$, and 0.20 .

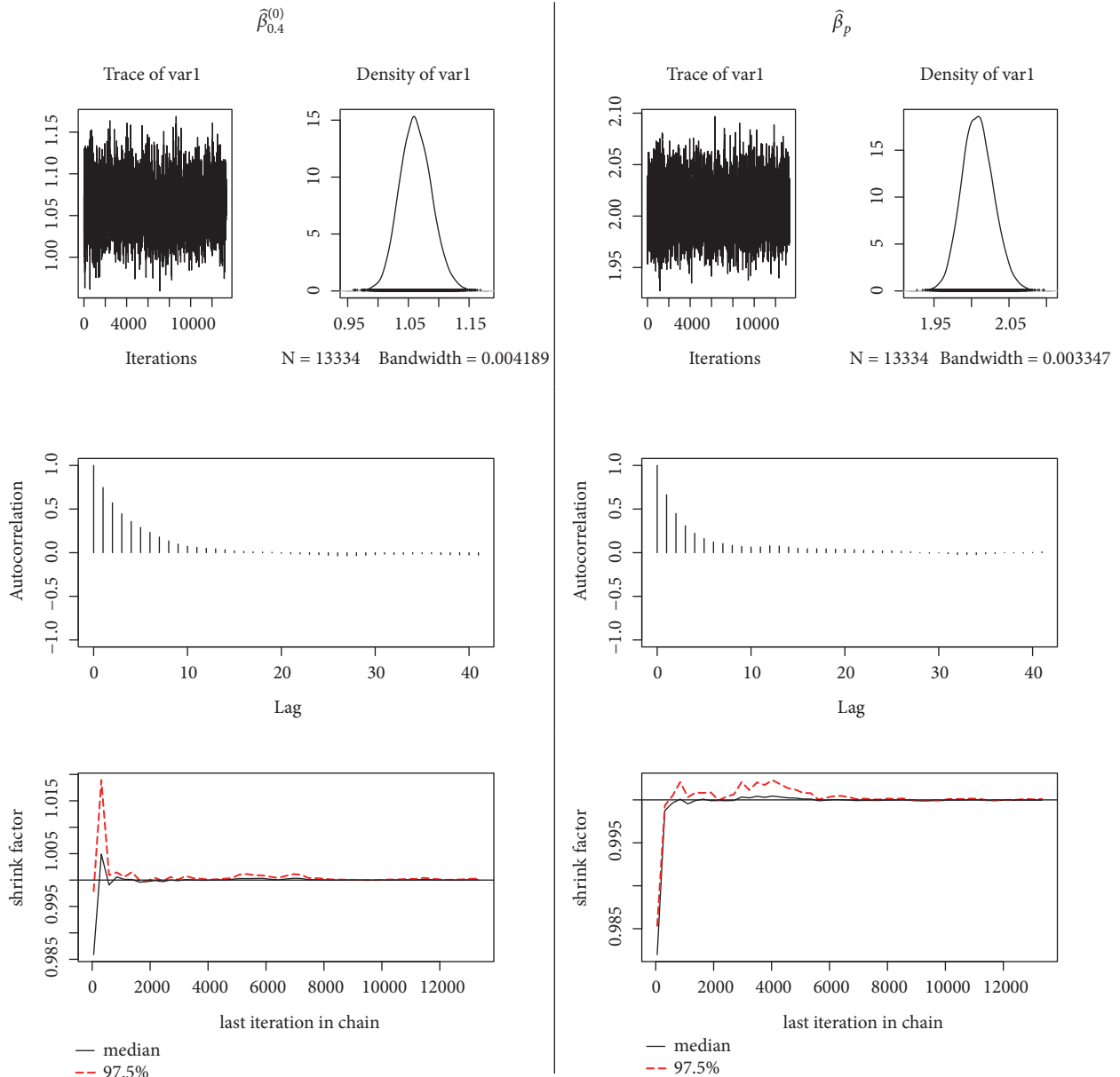


FIGURE 3: Trace and density plot (top), autocorrelation plot (middle), and \hat{R} evolution (bottom) of the intercept $\hat{\beta}_{0.4}^{(0)}$ (left panel) of the 0.4–th quantile and the slope $\hat{\beta}_p$.

TABLE 2: Table of criteria: crossing loss and RMISE for normal covariate case.

Method	nb. crossing	crossing loss	RMISE (0.10)	RMISE (0.12)
“K&B”	38	0.0037	0.2356	0.2151
“Y&M”	5	0.0005	0.4569	0.4391
“R&S”	0	0	0.2865	0.3189
“SBQR”	0	0	0.2218	0.2102

We fix $sd_p = 0.05$ and $sd_\sigma = 0.005$ and we compare our “SBQR” method with two others: the “M&ST” method with three-order cubic B-splines and the “F&R” approach with smoothness parameter value equal to 0.1.

For each value of τ , we evaluate the performance of these methods through the RMISE criterion. Table 4 shows that the RMISE values are significantly smaller for “SBQR” than the ones for “M&ST” and “F&R”. It is worth noting, therefore,

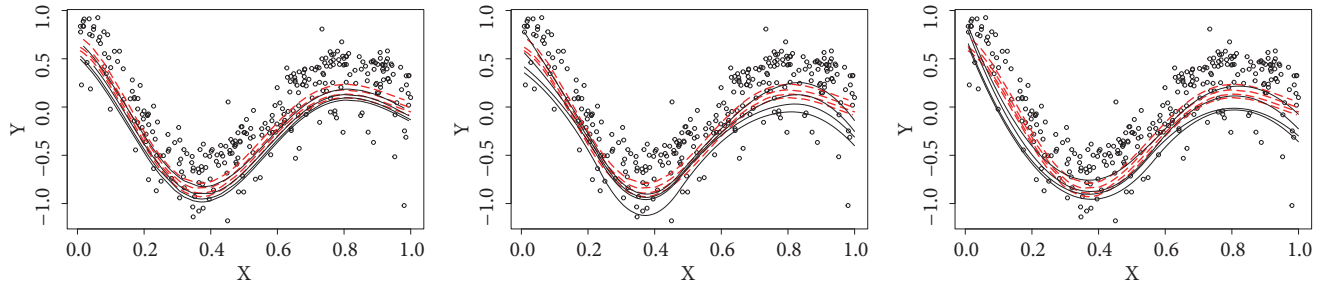


FIGURE 4: Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for SBQR (left panel), F&R (middle panel), and M&ST (right panel) methods.

TABLE 3: Table of criteria: crossing loss and RMISE for uniform covariate case.

Method	nb. crossing	crossing loss	RMISE (0.7)	RMISE (0.8)
“K&B”	6	0.0041	0.4255	0.7177
“Y&M”	1	0.0009	0.403	0.7013
“R&S”	0	0	0.224	0.3791
“SBQR”	0	0	0.2065	0.2708

TABLE 4: RMISE at different quantile levels computed for M&ST, F&R, and SBQR methods.

Method	RMISE (0.10)	RMISE (0.12)	RMISE (0.15)	RMISE (0.20)
“M&ST”	0.1416	0.1397	0.1072	0.0774
“F&R”	0.1520	0.1033	0.0821	0.0664
“SBQR”	0.0692	0.0768	0.0777	0.0660

that our method is significantly better than the two others in quantiles estimation; besides, it provides a reasonable estimation of σ ($\hat{\sigma} = 0.0535$) and p ($\hat{p} = 0.7744$).

Figure 4 gives the quantile curves estimators. While our “SBQR” method (left panel) provides quantile curves estimates that are close to the true ones (red-dashed lines), the 0.1-th quantile curve estimate given by “F&R” method (middle panel) is fairly distant from the true curve, especially when $x \in [0.2, 0.6]$. The same happens for “M&ST” method when $x \in [0.7, 1]$. However, there is no scarred crossing by any of these three methods, since they are tackling simultaneous quantile estimation techniques.

The issue that makes “F&R” method less flexible in simultaneous quantiles fitting is that, in the second stage, the final estimators are quite affected by the first stage output that may be badly estimated. For “M&ST”, the estimators are constructed iteratively, when solving the minimization problem, by adding constraints so that each subsequent quantile function does not cross with the previous one; this may cause overestimation of quantile curves. Our “SBQR” method does not suffer from these mentioned disadvantages, which explains its significant flexibility.

5. Conclusion

Our proposed estimation procedure, “SBQR”, for simultaneous Bayesian quantile regression guarantees the fundamental property of noncrossing. Assuming that the ALD is the underlying data distribution, this method enables characterizing the likelihood function by all quantiles of interest

using the relation between two distinct quantiles. Using the location-scale normal mixture representation of the ALD distribution, we develop a Metropolis-Hastings within Gibbs algorithm to implement our method. Our simulation studies show good results that reflect the good performance of the method and the convergence of the algorithm. Against the crossing problem of estimated quantiles, our method has good performance compared with single quantile estimation methods like Koenker and Bassett [1] and Yu and Moyeed [13] methods. From the RMISE point of view, “SBQR” is very competitive in both linear and nonlinear quantile regression cases.

As future perspectives, we envisage extending this work to deal with any conditional distribution of $Y | X$.

Data Availability

The simulated data used to support the findings of this study are provided by R software and are included within the attached document.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The author would like to thank the Ph.D. supervisor Professor Ghislaine Gayraud, University of Technology of Compiegne,

France, for her support and comments that greatly assisted this paper.

Supplementary Materials

The simulated data used to support the findings of this study are provided by R software and are included within the attached document. (*Supplementary Materials*)

References

- [1] R. Koenker and G. Bassett Jr., “Regression quantiles,” *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [2] C. Gutenbrunner and J. Jurecková, “Regression rank scores and regression quantiles,” *The Annals of Statistics*, vol. 20, no. 1, pp. 305–330, 1992.
- [3] K. Q. Zhou and S. L. Portnoy, “Direct use of regression quantiles to construct confidence sets in linear models,” *The Annals of Statistics*, vol. 24, no. 1, pp. 287–306, 1996.
- [4] R. Koenker and Z. Xiao, “Inference on the quantile regression process,” *Econometrica*, vol. 70, no. 4, pp. 1583–1612, 2002.
- [5] D. Nychka, G. Gray, P. Haaland, D. Martin, and M. O’Connell, “A nonparametric regression approach to syringe grading for quality improvement,” *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1171–1178, 1995.
- [6] R. Koenker and S. Portnoy, *Nonparametric Estimation of Conditional Quantile Functions*, 1992.
- [7] R. Koenker, P. Ng, and S. Portnoy, “Quantile smoothing splines,” *Biometrika*, vol. 81, no. 4, pp. 673–680, 1994.
- [8] H.-S. Oh, T. C. M. Lee, and D. W. Nychka, “Fast nonparametric quantile regression with arbitrary smoothing methods,” *Journal of Computational and Graphical Statistics*, vol. 20, no. 2, pp. 510–526, 2011.
- [9] R. Koenker and I. Mizera, “Penalized triograms: total variation regularization for bivariate smoothing,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 1, pp. 145–163, 2004.
- [10] R. Koenker, “Additive models for quantile regression: model selection and confidence bands,” *Brazilian Journal of Probability and Statistics*, vol. 25, no. 3, pp. 239–262, 2011.
- [11] X. He and P. Ng, “Quantile splines with several covariates,” *Journal of Statistical Planning and Inference*, vol. 75, no. 2, pp. 343–352, 1999.
- [12] Y. Li, Y. Liu, and J. Zhu, “Quantile regression in reproducing kernel Hilbert spaces,” *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 255–268, 2007.
- [13] K. Yu and R. A. Moyeed, “Bayesian quantile regression,” *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.
- [14] H. Kozumi and G. Kobayashi, “Gibbs sampling methods for Bayesian quantile regression,” *Journal of Statistical Computation and Simulation*, vol. 81, no. 11, pp. 1565–1578, 2011.
- [15] M. Bernardi, G. Gayraud, and L. Petrella, “Bayesian tail risk interdependence using quantile regression,” *Bayesian Analysis*, vol. 10, no. 3, pp. 553–603, 2015.
- [16] G. Kobayashi, “Bayesian endogenous Tobit quantile regression,” *Bayesian Analysis*, vol. 12, no. 1, pp. 161–191, 2017.
- [17] A. H. Li and J. Bradic, “Censored quantile regression forests,” 2019, <https://arxiv.org/abs/1902.03327>.
- [18] R. Alhamzawi, K. Yu, and D. F. Benoit, “Bayesian adaptive Lasso quantile regression,” *Statistical Modelling*, vol. 120, no. 3, pp. 279–297, 2012.
- [19] N. Sayed-Ahmed, “Efficiency of bayesian approaches in quantile regression with small sample size,” *Asian Journal of Probability and Statistics*, pp. 1–13, 2018.
- [20] H. K. Abbas and R. J. Thaher, “Bayesian adaptive lasso tobit regression,” *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 110, no. 1, 2019.
- [21] D. F. Benoit and D. Van den Poel, “Binary quantile regression: a bayesian approach based on the asymmetric laplace distribution,” *Journal of Applied Econometrics*, vol. 27, no. 7, pp. 1174–1188, 2012.
- [22] R. Alhamzawi and H. T. Ali, “Bayesian quantile regression for ordinal longitudinal data,” *Journal of Applied Statistics*, vol. 45, no. 5, pp. 815–828, 2018.
- [23] X. He, “Quantile curves without crossing,” *The American Statistician*, vol. 51, no. 2, pp. 186–192, 1997.
- [24] Y. Liu and Y. Wu, “Stepwise multiple quantile regression estimation using non-crossing constraints,” *Statistics and Its Interface*, vol. 2, no. 3, pp. 299–310, 2009.
- [25] H. D. Bondell, B. J. Reich, and H. Wang, “Noncrossing quantile regression curve estimation,” *Biometrika*, vol. 97, no. 4, pp. 825–838, 2010.
- [26] M. Sangnier, O. Fercoq, and F. D’Alché-Buc, “Joint quantile regression in vector-valued RKHSs,” in *Proceedings of the 30th Annual Conference on Neural Information Processing Systems, NIPS 2016*, pp. 3693–3701, Spain, December 2016.
- [27] B. J. Reich, M. Fuentes, and D. B. Dunson, “Bayesian spatial quantile regression,” *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 6–20, 2011.
- [28] B. J. Reich, “Spatiotemporal quantile regression for detecting distributional changes in environmental processes,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, no. 4, pp. 535–553, 2012.
- [29] S. T. Tokdar and J. B. Kadane, “Simultaneous linear quantile regression: a semiparametric Bayesian approach,” *Bayesian Analysis*, vol. 7, no. 1, pp. 51–72, 2012.
- [30] Y. Yang and S. T. Tokdar, “Joint estimation of quantile planes over arbitrary predictor spaces,” *Journal of the American Statistical Association*, vol. 112, no. 519, pp. 1107–1120, 2017.
- [31] T. Rodrigues and Y. Fan, “Regression adjustment for noncrossing Bayesian quantile regression,” *Journal of Computational and Graphical Statistics*, vol. 26, no. 2, pp. 275–284, 2017.
- [32] V. M. R. Muggeo, M. Sciandra, A. Tomasello, and S. Calvo, “Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology,” *Environmental and Ecological Statistics*, vol. 20, no. 4, pp. 519–531, 2013.
- [33] P. Das and S. Ghosal, “Bayesian quantile regression using random B-spline series prior,” *Computational Statistics & Data Analysis*, vol. 109, pp. 121–143, 2017.
- [34] P. Das and S. Ghosal, “Bayesian non-parametric simultaneous quantile regression for complete and grid data,” *Computational Statistics & Data Analysis*, vol. 127, pp. 172–186, 2018.
- [35] T. Rodrigues, J.-L. Dortet-Bernadet, and Y. Fan, “Pyramid quantile regression,” *Journal of Computational and Graphical Statistics*, pp. 1–25, 2019.
- [36] L. Petrella and V. Raponi, “Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress,” *Journal of Multivariate Analysis*, vol. 173, pp. 70–84, 2019.
- [37] K. Yu and J. Zhang, “A three-parameter asymmetric Laplace distribution and its extension,” *Communications in Statistics—Theory and Methods*, vol. 34, no. 9-10, pp. 1867–1879, 2005.

- [38] B. J. Reich and L. B. Smith, “Bayesian quantile regression for censored data,” *Biometrics: Journal of the International Biometric Society*, vol. 69, no. 3, pp. 651–660, 2013.
- [39] R. Koenker, *Quantile Regression in R: A Vignette*, 2015, <https://cran.r-project.org/web/packages/quantreg/vignettes/rq.pdf>.
- [40] D. Benoit, R. Al-Hamzawi, K. Yu, and D. Van den Poel, *bayesqr: Bayesian Quantile Regression*, 2011.
- [41] L. Smith, B. Reich, and M. L. Smith, “Package ‘bsquare’” 2013.
- [42] V. M. R. Muggeo, *Quantreggrowth-Package: Growth Charts via Regression Quantiles*, 2018.

Research Article

New Link Functions for Distribution-Specific Quantile Regression Based on Vector Generalized Linear and Additive Models

V. F. Miranda-Soberanis ¹ and T. W. Yee ²

¹*School of Engineering, Computer & Mathematical Sciences, Auckland University of Technology, New Zealand*

²*Department of Statistics, University of Auckland, New Zealand*

Correspondence should be addressed to V. F. Miranda-Soberanis; victor.miranda@aut.ac.nz

Received 7 December 2018; Accepted 22 January 2019; Published 7 May 2019

Guest Editor: Rahim Alhamzawi

Copyright © 2019 V. F. Miranda-Soberanis and T. W. Yee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the usual quantile regression setting, the distribution of the response given the explanatory variables is unspecified. In this work, the distribution is specified and we introduce new link functions to directly model specified quantiles of seven 1-parameter continuous distributions. Using the vector generalized linear and additive model (VGLM/VGAM) framework, we transform certain prespecified quantiles to become linear or additive predictors. Our parametric quantile regression approach adopts VGLMs/VGAMs because they can handle multiple linear predictors and encompass many distributions beyond the exponential family. Coupled with the ability to fit smoothers, the underlying strong assumption of the distribution can be relaxed so as to offer a semiparametric-type analysis. By allowing multiple linear and additive predictors simultaneously, the quantile crossing problem can be avoided by enforcing parallelism constraint matrices. This article gives details of a software implementation called the VGAMextra package for R. Both the data and recently developed software used in this paper are freely downloadable from the internet.

1. Introduction

1.1. Background. Much of modern regression analysis for estimating conditional quantile functions may be viewed as starting from Koenker and Bassett [1], who offered a systematic strategy for examining how covariates influence the entire response distribution. The fundamental idea is based on the linear specification of the τ th quantile function $Q_y(\tau | \mathbf{x}) = \boldsymbol{\beta}_\tau^T \mathbf{x}$ and finding $\boldsymbol{\beta}_\tau \in \mathbb{R}^p$ that solves the optimization problem

$$\min_{\boldsymbol{\beta}_\tau \in \mathbb{R}^p} \sum \rho_\tau(y_i - \boldsymbol{\beta}_\tau^T \mathbf{x}_i), \quad (1)$$

for independent and identically distributed (i.i.d.) observations from a family of linear quantile regression models, say $y_i = \boldsymbol{\beta}_\tau^T \mathbf{x}_i + \varepsilon_{i,\tau}$, $i = 1, \dots, n$. Equation (1) can be reformulated as a linear programming problem using the piecewise linear

function $\rho_\tau(u) = u \cdot [\tau - I(u < 0)]$ for $\tau \in (0, 1)$. More details can be found in Koenker [2].

In the spirit of quantile regression, the conditional distribution $Y | \mathbf{x}$ is usually unspecified, although it relies on normal-based asymptotic theory that is used for inference, whilst the assumption of homoskedasticity of the error terms $\varepsilon_{i,\tau}$ is dropped. In this paper we use an alternative approach of conditional-quantile regression based on assuming a pre-specified distribution for the response. Parametric quantile regression has some advantages over many nonparametric approaches, including overcoming the quantile crossing problem. Two examples are Noufaily and Jones [3] which is based on the generalized gamma distribution and generalized additive models for location, scale, and shape (GAMLSS; [4]). Further examples are the LMS-BCN method involving the standard normal distribution and a three-parameter Box-Cox transformation [5] and the classical method of

TABLE I: Some VGLM/VGAM link functions. The 4th row is implemented in VGAMextra.

Functions	Links $g_j(\theta_j)$	Domain of θ_j	Link names
<code>loge()</code>	$\log \theta_j$	$(0, \infty)$	Logarithmic
<code>cloglog()</code>	$\log(-\log(1 - \theta_j))$	$(0, 1)$	Complementary log–log
<code>logit()</code>	$\log \frac{\theta_j}{1 - \theta_j}$	$(0, 1)$	Logit
<code>logffMeanlink()</code> [‡]	$\text{logit}(\theta_j) - \text{cloglog}(\theta_j)$	$(0, 1)$	logffMeanlink
<code>rhobit()</code>	$\log((1 + \theta_j)/(1 - \theta_j))$	$(-1, 1)$	rhobit

[‡]This is the VGLM–link for the mean function of the logarithmic distribution.

quantile regression based on the asymmetric Laplace distribution (ALD).

Our approach uses the vector generalized linear and additive model (VGLM/VGAM; [6, 7]) framework. We develop new link functions, \mathcal{F} , for the quantile regression model

$$Y \mid \mathbf{x} \sim \mathcal{F}(\mathbf{x}; \theta), \quad (2)$$

$$\eta_{\tau}(\theta) = \mathcal{F}(Q_y(\tau \mid \mathbf{x}, \theta)), \quad (3)$$

for a vector of quantiles $\boldsymbol{\tau} = (\tau_1, \dots, \tau_L)^T$. Our methodology relies on the prespecification of the distribution \mathcal{F} . We will also show that the quantile crossing problem can be overcome by this modelling framework. Equations (2)–(3) state that the conditional distribution of the response at a given value of \mathbf{x} has a distribution involving a parameter θ and that the transformed quantile of the distribution becomes a linear predictor of the form (5). This can be achieved by defining link functions that connect (3) to (5). The reason for the linear predictors is that generalized linear modelling [8] is a very well-established method for regression modelling. GLMs are estimated by iteratively reweighted least squares (IRLS) and Fisher scoring, and this algorithm is also adopted by VGLMs and VGAMs.

The method presented in this paper differs from conventional quantile regression [1] in that we assume \mathcal{F} is known whereas the conventional case does not but use an empirical method instead to obtain the quantiles ξ_{τ} : the expectation of the check function $\rho_{\tau}(u)$ results in the property $\tau = F(\xi_{\tau})$ which defines the τ -quantile (F is the cumulative distribution function (CDF) of \mathcal{F}). In this paper we consider the \mathcal{F} s listed in Table 2.

1.2. VGLMs and VGAMs. VGLMs/VGAMs provide the engine and overall modelling framework in this work—the VGAM R package described below fits over 150 models and distributions—therefore we only sketch the details here. VGLMs are defined in terms of M linear predictors, $\boldsymbol{\eta} = (\eta_1, \dots, \eta_M)^T$, as any statistical model for which the conditional density of \mathbf{y} given a d -dimensional vector of explanatory variables, $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$ has the form

$$\mathcal{F}(\mathbf{y} \mid \mathbf{x}; \mathbf{B}) = h(\mathbf{y}, \eta_1, \dots, \eta_M; \mathbf{x}), \quad (4)$$

for some known function $h(\cdot)$, with $\mathbf{B} = (\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2 \ \dots \ \boldsymbol{\beta}_M)$, a $d \times M$ matrix of unknown regression coefficients. Ordinarily, $x_1 \equiv 1$ for an intercept.

In general, the η_j of VGLMs may be applied directly to the M parameters, θ_j , of any distribution, transformed if necessary, as the j th linear predictor

$$\eta_j = g_j(\theta_j) = \eta_j(\mathbf{x}) = \boldsymbol{\beta}_j^T \mathbf{x} = \sum_{k=1}^d \beta_{(j)k} x_k, \quad (5)$$

$$j = 1, \dots, M,$$

where g_j is a VGLM–parameter link function, as in Table 1 (see [6] for further choices) and $\beta_{(j)k}$ is the k th element of $\boldsymbol{\beta}_j$. Prior to this work the θ_j were ‘raw’ parameters such as location, scale, and shape parameters; however, in this present work we define them to be quantiles or a very simple function of quantiles.

In matrix form one can write $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x}) =$

$$\begin{pmatrix} \eta_1(\mathbf{x}) \\ \vdots \\ \eta_M(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1^T \mathbf{x} \\ \vdots \\ \boldsymbol{\beta}_M^T \mathbf{x} \end{pmatrix} = \begin{pmatrix} \beta_{(1)1} & \cdots & \beta_{(1)p} \\ \vdots & \ddots & \vdots \\ \beta_{(M)1} & \cdots & \beta_{(M)p} \end{pmatrix} \mathbf{x} \quad (6)$$

$$= \sum_{k=1}^d \boldsymbol{\beta}_{(k)} x_k = \mathbf{B}^T \mathbf{x},$$

where $\boldsymbol{\beta}_{(k)} = (\beta_{(1)k}, \beta_{(2)k}, \dots, \beta_{(M)k})^T$, $k = 1, \dots, d$. Sometimes, for some j , it may be required to model η_j as *intercept-only*, that is, $\eta_j = \beta_{(j)1}$, and $\beta_{(j)k} \equiv 0$ for $k = 2, \dots, d$.

VGAMs are a nonparametric extension of VGLMs, that is, (6) is generalized to

$$\boldsymbol{\eta}(\mathbf{x}) = \boldsymbol{\beta}_{(1)} + \sum_{k=2}^d \mathbf{f}_k(x_k) = \mathbf{H}_1 \boldsymbol{\beta}_{(1)}^* + \sum_{k=2}^d \mathbf{H}_k \mathbf{f}_k^*(x_k) \quad (7)$$

with $\mathbf{f}_k^*(x_k) = (f_{(1)k}^*(x_k), \dots, f_{(\mathcal{G}_k)k}^*(x_k))^T$. Usually the component functions are estimated by splines. Here, $\mathbf{H}_1, \dots, \mathbf{H}_d$ are known full-column rank constraint matrices, and $\boldsymbol{\beta}_{(1)}$ is a vector of unknown intercepts. With no constraints at all, $\mathbf{H}_1 = \dots = \mathbf{H}_d = \mathbf{I}_M$ (the order- M identity matrix). For VGLMs, the \mathbf{f}_k are linear so that, cf. (6),

$$\mathbf{B} = (\mathbf{H}_1 \boldsymbol{\beta}_{(1)}^* \mid \mathbf{H}_2 \boldsymbol{\beta}_{(2)}^* \mid \cdots \mid \mathbf{H}_d \boldsymbol{\beta}_{(d)}^*). \quad (8)$$

The \mathbf{H}_k can enforce a wide range of linear constraints such as parallelism and exchangeability.

TABLE 2: New link functions for the quantiles of some 1-parameter distributions. The selected \mathcal{G} function is also shown.

Distribution	θ	Support of y	Quantile function ξ_τ	Function \mathcal{G}	Quantile link $[\eta(\theta; \tau)]$
Exponential	λ	$(0, \infty)$	$-\frac{1}{\lambda} \log(1 - \tau)$	log-link	$\log \log[(1 - \tau)^{-1/\theta}]$
Benini	s	(y_0, ∞)	$y_0 \exp\left(\sqrt{\frac{-\log(1 - \tau)}{s}}\right)$	log-link	$\log y_0 + \sqrt{\log[(1 - \tau)^{-1/\theta}]}$
Rayleigh	b	$(0, \infty)$	$b\sqrt{-2 \log(1 - \tau)}$	log-link	$\log \theta + \frac{1}{2} \log \log[(1 - \tau)^{-2}]$
Gamma	s	$(0, \infty)$	No closed-form	log-link	$\log \text{qgamma}(\tau, \text{shape} = \theta)$
Maxwell [†]	a	$(0, \infty)$	$\sqrt{\frac{2}{a}} \cdot \text{qgamma}(\tau, 1.5)$	log-link	$\frac{1}{2} \log\left(\frac{2 \text{qgamma}(\tau, 1.5)}{\theta}\right)$
Topp-Leone [‡]	s	$(0, 1)$	$1 - \sqrt{1 - \tau}^{1/s}$	logit	$\text{logit}(1 - \sqrt{1 - \tau}^{1/\theta})$
1-par Normal [‡]	σ	\mathbb{R}	$\mu_0 \pm \sqrt{2 \cdot \theta^2 \kappa(\tau)}$	identity	$\mu_0 \pm \sqrt{2 \cdot \theta^2 \kappa(\tau)}$

[†] $\text{qgamma}()$ is the quantile function of the standard gamma distribution in R.

[‡] $\text{logit}(\theta) = \log(\theta/(1 - \theta))$.

[‡] $\kappa(\tau) = \text{erf}^{-1}(2\tau - 1)$, with $\text{erf}()$ denoting the error function.

1.3. *Estimation.* VGLMs are estimated by maximum likelihood performed by IRLS using the expected information. The VGLM log-likelihood is given by

$$\ell(\boldsymbol{\eta}) = \sum_{i=1}^n w_i \ell_i \{ \eta_1(\mathbf{x}_i), \dots, \eta_M(\mathbf{x}_i) \}, \quad (9)$$

for known fixed positive prior weights w_i , and a Newton-like algorithm for maximizing (9) has the form $\boldsymbol{\beta}^{(a)} = \boldsymbol{\beta}^{(a-1)} + \mathcal{F}(\boldsymbol{\beta}^{(a-1)})^{-1} \mathbf{U}(\boldsymbol{\beta}^{(a-1)})$, where \mathcal{F} is the overall expected information matrix (EIM), \mathbf{U} is the score vector, and a is the iteration number. The vector $\boldsymbol{\beta}^{(a)}$ is obtained as the solution of the generalized least squares problem $\boldsymbol{\beta}^{(a)} = \arg \min_{\boldsymbol{\beta}} \text{RSS}$, where the quantity minimized at each IRLS iteration is the weighted (or residual) sum of squares, $\text{RSS} =$

$$\sum_{i=1}^n w_i \{ \mathbf{z}_i^{(a-1)} - \boldsymbol{\eta}_i^{(a-1)} \}^T \mathbf{W}_i^{(a-1)} \{ \mathbf{z}_i^{(a-1)} - \boldsymbol{\eta}_i^{(a-1)} \}. \quad (10)$$

The $(M \times M)$ \mathbf{W}_i are known as the working weight matrices and they have (j, k) th element given by

$$[\mathbf{W}_i]_{j,k} = -w_i \mathbb{E} \left(\frac{\partial^2 \ell_i}{\partial \eta_j \partial \eta_k} \right). \quad (11)$$

The use of individual EIMs instead of observed information matrices means that Fisher scoring is used rather than the Newton-Raphson algorithm.

VGAMs are also estimated by IRLS, where the difference with respect to VGLMs is that a vector additive model is now fitted to the pseudo-response \mathbf{z}_i with explanatory variables \mathbf{x}_i and working weight matrices \mathbf{W}_i at each IRLS iteration. Two approaches are currently used by VGAM to estimate the component functions \mathbf{f}^* : regression splines and vector smoothing methods with vector backfitting. Rudimentary P-splines [9] are almost operational, albeit this work is not yet complete. Compared to VGLMs, the VGAM log-likelihood

includes a penalty if used with vector smoothing splines. In VGAM the objective function maximized is

$$\ell_i \{ \eta_1(\mathbf{x}_i), \dots, \eta_M(\mathbf{x}_i) \} - \frac{1}{2} \sum_{k=1}^d \sum_{j=1}^{\text{ncol}(\mathbf{H}_k)} \lambda_{(j)k} \int_{a_k}^{b_k} \{ f_{(j)k}^{*''}(t) \}^2 dt. \quad (12)$$

Here, the $\lambda_{(j)k}$ are nonnegative smoothing parameters, and $a_k \leq x_{ik} \leq b_k$ are endpoints covering the values of each covariate. The basic penalty approach adopted here is described in Green and Silverman Green and Silverman [10].

2. Methodology

Let $\mathcal{F}(\boldsymbol{\eta}; y, \mathbf{x})$ be a 1-parameter statistical model as in (4) parametrized by $\theta \in \Theta \subset \mathbb{R}$ for some parameter space Θ residing in $(-\infty, \infty)$. Also let $Q_y(\tau | \mathbf{x})$ be the corresponding quantile function with $\tau \in (0, 1)$. Crucially, note that (5) handles suitable transformations of θ in the linear predictor by parameter link functions. In contrast our proposal focusses on directly modelling $Q_y(\tau | \mathbf{x}, \theta)$ via a smooth and one-to-one function \mathcal{G} , in the form of

$$\eta_\tau = \mathcal{G}(Q_y(\tau | \mathbf{x}, \theta)) = \mathcal{G}^*(\theta | \mathbf{x}, \tau), \quad (13)$$

which is to be incorporated in the VGLM/VGAM log-likelihood, namely, (9) and (12). Here, $\boldsymbol{\tau} = (\tau_1, \dots, \tau_L)^T$ is a prespecified vector of quantiles of interest. Examples of (13) are $\log \xi_\tau$, $\text{logit} \xi_\tau$, and ξ_τ .

Equation (13) is central to this work. It allows modelling choices via \mathcal{G} for the quantile function Q_y , and it represents a new modification to the VGLM/VGAM framework. Note that \mathcal{G} resembles a *link function* within the VGLM/VGAM framework as in Table 1. Two notes: first, without any loss of generality, (13) can be seen (strictly) as a function of θ since the quantiles $\boldsymbol{\tau}$ and the covariates \mathbf{x} are known. Secondly, \mathcal{G}^* is monotonic and one-to-one, as a result of the composite of \mathcal{G} and Q_y , which also hold such properties. However, during the fitting process, the IRLS algorithm internally requires the inverse of \mathcal{G}^* . Working with 1-parameter distribution

at this stage eases the implementation via Fisher scoring because the inverse $(\mathcal{G}^*)^{-1}$ can be derived manually and then incorporated in the IRLS algorithm. In a few cases, the inverse of \mathcal{G}^* does not have a closed form, such as the 1-parameter gamma distribution, and an alternative iterative method is employed to approximate $(\mathcal{G}^*)^{-1}$. To achieve this efficiently, two choices are available. These are (a) `newtonRaphson.basic()` from `VGAMextra` and (b) `VGAM::bisection.basic()`, two *vectorized* implementations of the well-known Newton–Raphson and bisection algorithms, to solve the roots of a real-valued function in a given interval (a, b) . Further details are given in Section 2.2.

One advantage of this work is that the VGLM/VGAM framework can circumvent the quantile crossing problem (e.g., [2, 11], Sect. 2.5) by choosing $\mathbf{H}_1 = \mathbf{I}_M$ and $\mathbf{H}_2 = \mathbf{H}_3 = \dots = \mathbf{H}_d = \mathbf{I}_M$ (an M -vector of ones). Under this parallelism assumption the method borrows strength across the entire data set so that the additive models for η_j with respect to x_2, \dots, x_d are parallel. Each family function in Tables 2 and 3 has a `parallel` argument which is `FALSE` by default. Using the syntax of `VGAM` based on Chambers and Hastie [12], setting `parallel = TRUE` (or `parallel = FALSE ~ 1`) results in $\mathbf{H}_1 = \mathbf{I}_M$ and $\mathbf{H}_2 = \mathbf{H}_3 = \dots = \mathbf{I}_M$; i.e., it is false for only the intercept.

It is noted that for some distributions such as the exponential and Maxwell the η_j are naturally parallel with respect to x_2, \dots, x_d because $\log \xi_\tau$ has the form $h_1(\tau) + h_2(x_2, \dots, x_d)$. If this is so then only the intercepts will change as a function of τ and the MLEs for h_2 are the same. Other distributions such as the 1-parameter gamma do not possess this property, and then it is necessary to constrain $\mathbf{H}_2 = \dots = \mathbf{H}_d = \mathbf{I}_M$ to avoid the quantile crossing problem.

2.1. Two Derivations. Ideally the link transforms the support of Y to \mathbb{R} because η_j should be unbounded. The three most common cases are as follows. For $Y \in (0, \infty)$ a log link is recommended, for $Y \in (0, 1)$ a logit link is a good choice, and $Y \in (-\infty, \infty)$ means that an identity link is natural. These cases have been implemented for seven 1-parameter distributions. The selection of the function \mathcal{G} for each Q is shown in the 5th column of Table 2, whilst the resulting quantile links as functions of θ are shown in the last column.

We now describe the quantile links for the exponential and the Topp–Leone distributions as examples. Firstly, for $Y \sim \text{exponential}(\theta)$, with a rate parameter $\theta > 0$, the density and CDF are given by $f(y; \theta) = \theta e^{-\theta y}$ and $F(y; \theta) = 1 - e^{-\theta y}$. With a slight change in notation, the quantile function is given by F^{-1} , i.e.,

$$Q_Y(\tau; \theta) = -\frac{1}{\theta} \log(1 - \tau), \quad (14)$$

which lies in $(0, \infty)$ regardless of the values of τ and θ . Given that values of τ are known (prespecified by the user), (14) becomes a function of θ . Thus, the new quantile link for the exponential distribution as shown in Table 2 is simply

obtained by taking \mathcal{G} as the logarithmic transformation, as follows:

$$\begin{aligned} \eta(\theta; \tau) &= \log \left[-\frac{1}{\theta} \log(1 - \tau) \right] \\ &= \log \log \left[(1 - \tau)^{-1/\theta} \right]. \end{aligned} \quad (15)$$

This quantile link has been implemented in `VGAMextra` via the function `expQlink()`, as shown in Table 3. Its inverse (denoted as $\theta(\eta; \tau)$) can be manually obtained from the inverse of (15). Note that the corresponding family function (`exponential()`) implemented in `VGAM` includes a (known) location parameter A , which gives the density $f(y; \theta) = \theta e^{-\theta(y-A)}$. By default $A = 0$, and it is handled by the argument `location`.

Secondly, consider the Topp–Leone distribution $Y \sim \text{Topp} - \text{Leone}(s)$ whose support is $(0, 1)$ and

$$Q_Y(s; \tau_j) = 1 - \sqrt{1 - \tau_j^{1/s}}, \quad (16)$$

with $0 < \tau_j < 1$. Here, $\theta = s$. To verify this restriction note that $1/s > 1$, for any shape parameter $s \in (0, 1)$, and hence for any $\tau_j \in (0, 1)$,

$$\begin{aligned} 0 < \tau_j^{1/s} < 1 &\iff \\ 0 < 1 - \tau_j^{1/s} < 1 &\iff \\ 0 < 1 - \sqrt{1 - \tau_j^{1/s}} < 1. \end{aligned} \quad (17)$$

Thus, to allow the quantile function to be modelled by covariates, we take the logit transformation as \mathcal{G} . The resulting quantile link for this distribution is simply $\eta(s; \tau) = \log Q_Y(s; \tau)$, shown in Table 2. The distribution has CDF $F(y; s) = [y \cdot (2 - y)]^s$ for $0 < y < 1$, and density $f(y; s) = 2s(1 - y) \cdot [y(2 - y)]^{s-1}$. The quantile function derives from solving the equation $\tau_0 = F(y; s) = [y \cdot (2 - y)]^s$, for $0 < \tau_0 < 1$, which leads to the quadratic equation $y^2 - 2y + \tau_0^{1/s} = 0$. The solution must lie in $(0, 1)$ and is in fact (16), as a function of s . The family function `topple()` from `VGAM` estimates s , where the default link is $\eta(s) = \text{logit}(s)$.

2.2. Software Implementation. For practical use by others, we have implemented seven VGLM–quantile links, η_τ in the R package `VGAMextra`. They are summarized in Table 2. The package `VGAM` is a requirement of `VGAMextra` because the modelling functions `vglm()` and `vgam()`, and all but the last family function of Table 2, reside there. For this paper `VGAMextra` 0.0-2 and `VGAM` 1.1-0 or later are required; they are available at www.stat.auckland.ac.nz/~vmir178 and www.stat.auckland.ac.nz/~yee/VGAM/prerelease/ whilst older versions of both are available on CRAN (<http://CRAN.R-project.org>).

One special case is `gamma1Qlink()`, for the 1-parameter (shape) gamma distribution, defined as

$$\eta(\theta; \tau) = \log \text{gamma}(\tau, \text{shape} = \theta), \quad (18)$$

TABLE 3: Inverse of the quantile links and names in VGAMextra. “Approximate” means that Newton–Raphson or bisection is used to approximate the inverse. All family functions except for `normal1sdff()`, which is in VGAMextra, are in VGAM.

Distribution	θ	Inverse $L[\theta(\eta; \boldsymbol{\tau})]$	Family function	Link in VGAMextra
Exponential	λ	$\frac{-\log(1-\boldsymbol{\tau})}{e^\eta}$	<code>exponential()</code>	<code>expQlink()</code>
Benini	s	$\frac{\log(1-\boldsymbol{\tau})}{(\eta - \log y_0)^2}$	<code>benini1()</code>	<code>benini1Qlink()</code>
Rayleigh	b	$\frac{\exp(\eta)}{\sqrt{-2\log(1-\boldsymbol{\tau})}}$	<code>rayleigh()</code>	<code>rayleighQlink()</code>
Gamma	s	Approximate	<code>gamma1()</code>	<code>gamma1Qlink()</code>
Maxwell	a	$\frac{2\text{qgamma}(\boldsymbol{\tau}, 1.5)}{\exp(2\eta)}$	<code>maxwell()</code>	<code>maxwellQlink()</code>
Topp–Leone [†]	s	$\frac{\log \boldsymbol{\tau}}{\log\{1 - [1 - \text{logit}^{-1}(\eta)]^2\}}$	<code>topple()</code>	<code>toppleQlink()</code>
$N(\mu = 0, \sigma)$	σ	$ (\eta - \mu_0)/\sqrt{2} \cdot \kappa(\boldsymbol{\tau}) $	<code>normal1sdff()</code>	<code>normal1sdQlink()</code>

[†]`logit-1()` denotes the inverse of the `logit()` transformation.

whose primary arguments are $\boldsymbol{\tau}$ and θ . Its inverse (Table 3) does not admit a closed form and it is approximated by the function `VGAM::newtonRaphson.basic()`, a *vectorized* implementation of the Newton–Raphson algorithm. Almost all implementations elsewhere of this are for a scalar argument, but we operate on vectors of length n . It works as follows. Our data is effectively $\mathbf{y}_i = \{y_i, \mathbf{x}_{i,d}\}$, $i = 1, \dots, n$, whilst the quantiles of interest, $\boldsymbol{\tau}$ or \mathbf{p} , must be entered by the user. The shape parameter θ is estimated by IRLS and therefore it is available at each iteration. Thus, for each $\eta(\theta; \mathbf{p})_0$, the ‘inverse’ is given by the root, θ , of the function

$$f(\theta; \mathbf{p}, \eta) = \eta(\theta; \mathbf{p})_0 - \log \text{gamma}(\boldsymbol{\tau}, \text{shape} = \theta). \quad (19)$$

Finally, the inverse of all the VGLM–quantile links is shown in Table 3, as well as the name of the corresponding implementation in VGAMextra. The inverse–links are required at different stages of the IRLS by Fisher scoring, which internally switches between $\eta(\theta; \boldsymbol{\tau})$ (namely, Table 2) and $\theta(\eta; \boldsymbol{\tau})$ (namely, Table 3). Specifically, the algorithm requires the score vector and the EIMs at each

IRLS iteration, which are given by the following chain–rule formulas:

$$\begin{aligned} \frac{\partial \ell}{\partial \eta} &= \frac{\partial \ell}{\partial \theta} \cdot \frac{\partial \theta}{\partial \eta}, \\ -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \eta^2} \right] &= -\mathbb{E} \left[\frac{\partial^2 \ell}{\partial \theta^2} \right] \left(\frac{\partial \theta}{\partial \eta} \right)^2. \end{aligned} \quad (20)$$

Internally, the functions utilized to compute the inverse are `VGAM::eta2theta()` or `VGAM::theta2eta()`. The VGAMextra Manual and Miranda-Soberanis [13] give further details about the derivation of the quantile links, whilst Yee [6] describes in the IRLS and Fisher scoring algorithms for estimating VGLMs and VGAMs. Complements at the second author’s homepage give additional details on link functions.

2.3. Software Use. For the user, this methodology runs as usual by calling the modelling functions `VGAM::vglm()` and `VGAM::vgam()`, except for two modifications that are described below.

To start with, we give the following output that shows the central arguments handled by `VGAM::vglm()`:

```
> library("VGAM")
> args(vglm)

function (formula, family = stop("argument 'family' needs to be assigned"),
  data = list(), weights = NULL, subset = NULL, na.action = na.fail,
  etastart = NULL, mustart = NULL, coefstart = NULL, control = vglm.control(...),
  offset = NULL, method = "vglm.fit", model = FALSE, x.arg = TRUE,
  y.arg = TRUE, contrasts = NULL, constraints = NULL, extra = list(),
  form2 = NULL, qr.arg = TRUE, smart = TRUE, ...)
NULL
```

The first adjustment takes place with the argument `formula`, a symbolic description of the model to be fit. Usually, an expression like $y \sim x_2 + x_3$ should suffice

for a response y and covariates x_2 and x_3 . This effectively works for univariate and even for multiple responses say y_1, y_2 , and y_3 , where the only change is to set

TABLE 4: Arguments handled by the function `VGAMextra::Q.reg()`.

Argument	Description
<code>y</code>	Numeric, a vector or a matrix. It is the response or dependent variable in the formula of the model to be fitted.
<code>pvector</code>	A prototype vector. Entries are the conditional p -quantiles in the fitting process.
<code>length.arg</code>	A unit-length positive integer. It is the number of p -quantiles to be modelled.

`cbind(y1, y2, y3) ~ x2 + x3`. Here, the right hand side (RHS) of the formula is applied to each linear predictor.

For quantile modelling using VGLMs and VGAMs, `Q.reg()` must be incorporated in the formula, whose arguments are shown in Table 4. For a given set of quantiles of interest, entered through $\boldsymbol{\tau} = (\tau_1, \dots, \tau_L)^T$, `Q.reg()` replicates the response matrix \mathbf{Y} into $NOS \cdot \dim(\boldsymbol{\tau})$ columns, where NOS denotes the number of columns of \mathbf{Y} . Then, the

```
> p.quantiles <- c(0.25, 0.5, 0.75) ## Quantiles of interest
> fit.quantile <-
  vglm(Q.reg(y, pvector = p.quantile) ~ x2 + x3,
       family = rayleigh(lscale = rayleighQlink(p = p.quantile)),
       data = mydata)
```

Further fitting variants can be incorporated here, e.g., categorical covariates and the use of smoothers such as regression splines. These and a few other features are illustrated in the following section.

3. Examples

3.1. Maxwell Data. We use simulation to generate $n = 200$ random variates from a Maxwell distribution whose rate parameter is a function of a single covariate x_2 . To account for a nonlinear trend in the dataset, additive models with cubic smoothing splines appear to be a better choice over linear schemes such as with VGLMs. In this example we perform the following steps to confirm the performance of the methodology.

- (1) Generate random deviates from the Maxwell distribution.
- (2) Run conditional VGAM-quantile modelling using `maxwellQlink()` based on the VGAM family function `VGAM::maxwell()`, which estimates the Maxwell distribution by Fisher scoring.

RHS of the formula applies to every set of columns according to the number of quantiles of interest. Ordinarily the response is a vector so that $NOS = 1$ and $L = M$.

As an example, suppose that we have two responses Y_1 and Y_2 sampled from a prespecified distribution \mathcal{F} , as per Table 3, and the quantiles of interest are $p = (0.25, 0.50, 0.75)^T$. Then `Q.reg(cbind(Y1, Y2), pvector = p)` will return a matrix with six columns, with the first three columns being Y_1 , one for each quantile, and similarly the last three columns equal Y_2 . Thus `vglm()` handles this model as a multiple responses fit.

The second adjustment is related to the argument family, a function that describes the statistical model to be fitted. Each family has at least one argument for the link functions to be used in the fitting process (the name changes from family to family). For example, for `VGAM::exponential()` this is called `link`, whilst for the family function `VGAM::benini1()` (see the third column of Table 3), it is called `lshape`. When VGLM-quantile modelling is to be performed, the corresponding link (last column of Table 3) must be entered into the family accordingly. All the quantile links manage the same arguments, including `p`, the vector of quantiles, except by `benini1Qlink()` which has the additional argument `y0`.

With both modifications, a typical call has the following form:

- (3) Perform ordinary quantile regression using `VGAM::alaplace1()` that estimates the 1-parameter ALD by Fisher scoring. Here, the special argument `tau` will be employed.
- (4) Plot the artificial data with the estimated quantile functions, $Q_y(\tau | \mathbf{x}, \hat{\theta})$ (from (2)), and the estimated quantile curves (from (3)) superimposed.

We will consider the quantiles 25%, 50%, and 75% for simplicity, so that $\boldsymbol{\tau} = (1/4, 1/2, 3/4)^T$.

Regarding (1), the data is generated by `VGAM::rmaxwell()`, which gives random deviates from the Maxwell distribution whose density is $f(y; a) = \sqrt{2/\pi a}^{3/2} y^2 \exp(-ay^2/2)$. We use the rate function

$$a = \exp \left\{ 2 - \frac{6 \sin(2x_{i2} - 1/5)}{(x_{i2} + 1/2)^2} \right\}, \quad (21)$$

where $X_{i2} \stackrel{\text{i.i.d}}{\sim} \text{Unif}(0, 1)$, $i = 1, \dots, n$. The following code chunk sets things up and the dataset is saved as `maxdata`.

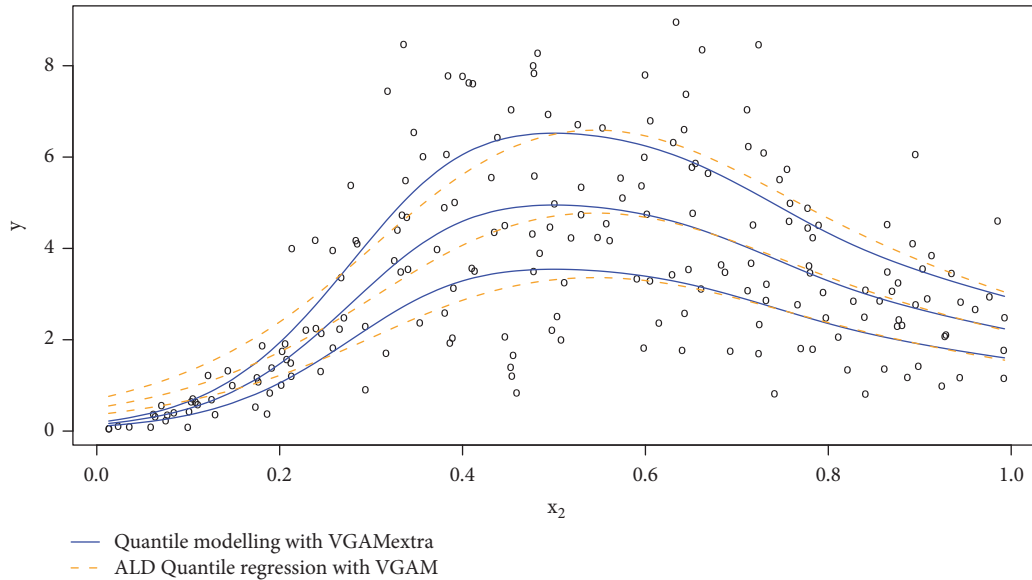


FIGURE 1: Simulated Maxwell data (21) including (a) the fitted quantile functions from `fit.Qmodelling` (VGAMextra) and (b) the fitted quantile curves from `fit.Qregression` (VGAM). The quantile curves in both cases derive from vector smoothing spline fits.

```
set.seed(1)
maxdata <- data.frame(x2 = sort(runif(n <- 200))) # Sorted for plotting
ratefun <- function(x) exp(2 - 6 * sin(2 * x - 0.2) / (x + 0.5)^2)
# Generate the data:
maxdata <- transform(maxdata, y = rmaxwell(n, rate = ratefun(x2)))
my.tau <- c(0.25, 0.50, 0.75) # Use these quantiles
```

The following code chunk performs steps (2) and (3). Note the fitting of additive models via `VGAM::vgam()` with smooth terms defined by `VGAM::s()` where x_2 is

to be smoothed. To compare both fits, they are saved in `fit.Qmodelling` (from (2)) and `fit.Qregression` (from (3)).

```
### Quantile modelling with 'maxwellQlink()'
mydof <- 4 # Effective degrees of freedom of the smoothing spline
fit.Qmodelling <-
  vgam(Q.reg(y, pvector = my.tau) ~ s(x2, df = mydof), data = maxdata,
       maxwell(link = maxwellQlink(p = my.tau),
              type.fitted = "Qlink"))

### Quantile regression with alaplace2() from VGAM
fit.Qregression <- vgam(y ~ s(x2, df = mydof), alaplace1(tau = my.tau,
               llocation = "loge", parallel.locat = TRUE),
                    data = maxdata, maxit = 100)
```

Figure 1 shows the simulated data, the estimated quantile functions, and the fitted quantile curves from `fit.Qmodelling` and `fit.Qregression`, obtained from vector smoothing spline fits [14]. The results are similar for $x_2 > 0.3$, but our present work performs better at the bottom LHS tail. The data coverage from each modelling framework

is summarized in Table 5. Once again our work outperforms the ALD method.

We conclude with a few remarks.

- (1) The argument `p` is available for *all* quantile links in Table 3 and not only for `maxwellQlink()`.

It can be assigned any vector of percentile values.

- (2) Under the conditional VGAM-quantile modelling framework, the arguments to handle the parallelism assumption such as the arguments `parallel.locat` and `parallel.scale` in family functions are no longer required. This is internally managed by the new quantile links rather than being managed by the family function.

- (3) If `fit` is a Qlink fit, then `fitted(fit)` returns the fitted quantiles. This is in the form of a $n \times (L \cdot NOS)$ matrix. Similarly, `predict(fit)` returns a $n \times M$ matrix where the i th row is η_i^T .

3.2. *Comparison with the Quantreg Package.* For checking purposes, the results are compared with `quantreg` too. Figure 2 gives the results based on the following code.

```
> qrfit25 <- rq(y ~ bs(x2), data = maxdata, tau = 0.25)
> qrfit50 <- rq(y ~ bs(x2), data = maxdata, tau = 0.50)
> qrfit75 <- rq(y ~ bs(x2), data = maxdata, tau = 0.75)

> mylwd <- 1.5
> plot(y ~ x2, maxdata, main = "", xlab = expression(x[2]), las = 1,
       ylab = "y", cex = 0.85)
> with(maxdata, matlines(x2, fitted(fit.Qmodelling), col = "blue",
                          lwd = mylwd, lty = 1))
> lines(predict(qrfit25) ~ x2, maxdata, lty = 2, lwd = mylwd^2,
        col = "orange")
> lines(predict(qrfit50) ~ x2, maxdata, lty = 2, lwd = mylwd^2,
        col = "orange")
> lines(predict(qrfit75) ~ x2, maxdata, lty = 2, lwd = mylwd^2,
        col = "orange")
> legend("topleft", lwd = c(mylwd, mylwd^2),
        c("Quantile modelling with VGAMextra",
          "Quantile regression with quantreg"),
        cex = 0.95, col = c("blue", "orange"), lty = 1:2)
```

The results should be the similar to Section 3.1 because the ALD and the classical quantile regression method are essentially the same. It can be seen that the bottom LHS corner is not modelled well with `quantreg` either. Once again our method performs best, which is not surprising given the strong distributional assumption.

3.3. *Exponential Data.* Feigl and Zelen [15] fit an exponential distribution to a data set comprising the time to death

(in weeks) and white blood cell counts for two groups of leukaemia patients, and a binary variable for AG-positive and AG-negative. The two groups were not created by random allocation. The variable AG is the morphological variable, the AG factor; a numeric vector where 1 means AG-positive and 2 means AG-negative. We create AG01 which is $AG - 1$. We take the log of the white blood cell count (WBC) because it is very highly skewed. The data are found in `GLMsData` on CRAN, which supports Dunn and Smyth [16].

```
> leukwbc <- transform(leukwbc, AG01 = as.factor(AG - 1), # Creating AG01
                      logWBC = log(WBC) - 9) # Symmetric and centred
> tauvec <- c(25, 75) / 100 # Quantiles of interest
> expfit1 <- vglm(Q.reg(Time, tauvec) ~ logWBC + AG01, data = leukwbc,
                 exponential(expQlink(p = tauvec), ishrink = 0.05))
> coef(expfit1, matrix = TRUE)
              expQlink(rate1; 0.25) expQlink(rate2; 0.75)
(Intercept)          2.84757          4.42010
logWBC              -0.30442          -0.30442
AG011                -1.01765          -1.01765
```

One benefit of quantile modelling with VGLMs is that it easily allows comparisons of the effect of AG01 or any other

indicator variable, at different quantiles. First note that, for AG-positive patients with $\log WBC = 9$, the 25% percentile for

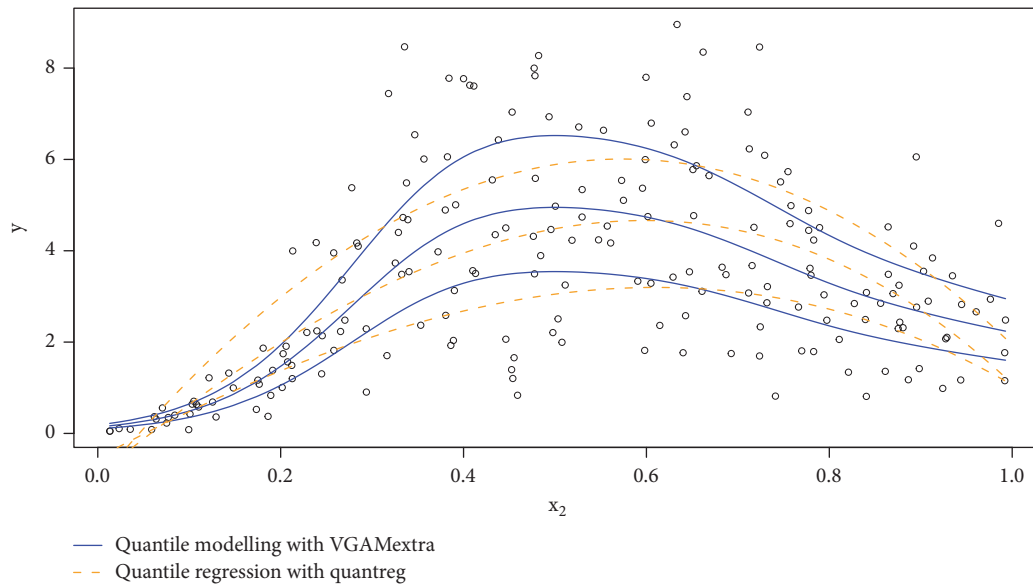


FIGURE 2: Simulated Maxwell data (21) including (a) the fitted quantile functions from `fit.Qmodelling` (VGAMextra) and (b) the fitted quantile curves from `qrfit50`, etc. from `quantreg`.

time to death is around $\exp 2.848 \approx 17.25$ weeks, whilst the 75% percentile is about $\exp 4.42 \approx 83.1$ weeks. Secondly, the coefficient of AG011 measures the influence of the AG factor on the time to death. Keeping the levels of WBC constant, for patients at either the 25% or 75% percentiles, the time to death for AG–negatives compared to AG–positives is multiplicative by a factor of $\exp(-1.02) \approx 0.361$, i.e., a 63.9% reduction in lifetime.

For further illustration's sake, we fit a 1-parameter gamma distribution to these data and interpret the results. Unlike the Maxwell and exponential distributions, where simple mathematics shows that different quantiles are parallel because their logarithm is additive with respect to τ , the 1-parameter gamma does not possess this property.

```
> # Fit a gamma model
> gamma1fit <- vglm(Q.reg(Time, tauvec) ~ logWBC + AG01, data = leukwbc,
  gamma1(gamma1Qlink(p = tauvec),
    parallel = FALSE ~ AG01))
> coef(gamma1fit, matrix = TRUE)
      gamma1Qlink(shape1; 0.25) gamma1Qlink(shape2; 0.75)
(Intercept)          3.67717          3.86380
logWBC              -0.58688          -0.58688
AG011               -1.46180          -1.26843
```

Here, keeping the level of WBC constant, for patients at the 25% percentile, the time to death for AG–negatives compared to AG–positives is multiplicative by a factor of $\exp(-1.46) \approx 0.232$, i.e., a 76.8% reduction. In comparison, for patients at the 75% percentile, the time to death for AG–negatives compared to AG–positives is multiplicative by a factor of

$\exp(-1.27) \approx 0.281$, i.e., a 71.9% reduction. This suggests that the effect of AG is greater for more severe cases than those who live longer in general.

Finally, just to check, we obtain the constraint matrices for each predictor:

```
> constraints(gamma1fit, matrix = TRUE)
      (Intercept):1 (Intercept):2 logWBC AG011:1 AG011:2
gamma1Qlink(shape1; 0.25)      1      0      1      1      0
gamma1Qlink(shape2; 0.75)      0      1      1      0      1
```

TABLE 5: Empirical data coverage from quantile–modelling using VGAMextra (QM–VGAMextra) and quantile–regression from VGAM (QR–VGAM), after fitting (21).

τ_j	QM–VGAMextra coverage	QR–VGAM coverage
25%	26%	28.5%
50%	50%	54%
75%	73.5%	78.5%

There is a parallelism assumption made for $\log\text{WBC}$ but not for any of the other explanatory variables.

4. Discussion and Future Work

This work in parametric quantile regression is blighted by the strong assumption of the assumed distribution. In theory, this might be ameliorated somewhat by implementing as many distributions as possible. Some of the distributions

```
> fitted(maxwellfit, percentile = 20, type.fitted = "quantile")
```

worked like many other VGAM models. The difficulty here is that the `@linkinv S4` slot of a VGAM family function has `eta` as an argument, and in our implementation this could only possibly be created by supplying the new percentiles to `predict()` beforehand.

Another minor deficiency in our software implementation is that the response vector is replicated $\dim(\boldsymbol{\tau})$ times so that is a form of recycling. Possibly this could be avoided because the memory requirement might be excessive when either $\dim(\boldsymbol{\tau})$ or n are very large.

At present, the VGAM framework has infrastructure to afford 1–parameter quantile links. For quantile functions depending on 2 or more parameter, such as the two-parameter gamma distribution, the quantiles will be bivariate functions whose inverse would probably not admit a closed form. Nevertheless, future work includes being able to write links for two-parameter distributions, of which the normal distribution would be the most important. For this, the methodology behind Yee and Miranda-Soberanis [17] could be employed; they solve a decades-old problem implementing the two-parameter canonical link function $\log(\mu/(\mu + k))$ of the negative binomial distribution. We have already commenced work in this direction, e.g., with the 2-parameter gamma distribution.

Data Availability

The data used to support the findings of this study are available in the Supplementary Materials.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

listed in Table 2 have real applications, for example, in kinetic-molecular theory the speed of individual molecules of idealized gases follows the Maxwell distribution and the average kinetic speed is directly related to Kelvin temperature. In experiments that do not satisfy the various postulates made (such as the effects of the container) one might model the median particle speed with \boldsymbol{x} comprising temperature and other covariates such as volume of the container and density of the container walls. Different forms of gases, such as plasmas and rarefied gases, could be modelled as such too. Another example is the Rayleigh distribution which is similar to the Maxwell distribution. In two-dimensions, and in applications of magnetic resonance imaging (MRI), complex images are often viewed in terms of the background data, which is Rayleigh distributed. Nonstandard background information could be included in \boldsymbol{x} and their effects on the distribution examined.

In the current software implementation there are limitations due to its internal design. For example, it would be good if

Acknowledgments

VM’s work was supported by University of Auckland Doctoral Scholarship.

Supplementary Materials

The file is a text file that comprises R commands that can be copied into R. The file contains information about what R packages need to be installed and loaded to run properly. There are no copyright issues. Some of the data is contained in the R packages, and some of the data is simulated data. (*Supplementary Materials*)

References

- [1] R. Koenker and G. Bassett Jr., “Regression quantiles,” *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [2] R. Koenker, *Quantile Regression*, vol. 38, Cambridge University Press, New York, USA, 2005.
- [3] A. Noufaily and M. C. Jones, “Parametric quantile regression based on the generalized gamma distribution,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 62, no. 5, pp. 723–740, 2013.
- [4] R. A. Rigby and D. M. Stasinopoulos, “Generalized additive models for location, scale and shape,” *Journal of Applied Statistics*, vol. 54, no. 3, pp. 507–554, 2005.
- [5] T. J. Cole and P. J. Green, “Smoothing reference centile curves: the LMS methods and penalized likelihood,” *Statistics in Medicine*, vol. 110, no. 5, pp. 1305–1319, 1992.
- [6] T. W. Yee, *Vector Generalized Linear and Additive Models with an Implementation in R*, Springer, New York, USA, 2015.

- [7] T. W. Yee and T. J. Hastie, "Reduced-rank vector generalized linear models," *Statistical Modelling. An International Journal*, vol. 3, no. 1, pp. 15–41, 2003.
- [8] J. Nelder and R. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society*, vol. 1350, no. 3, pp. 370–384, 1972.
- [9] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Statistical Science*, vol. 11, no. 2, pp. 89–121, 1996.
- [10] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London, UK, 1994.
- [11] X. He, "Quantile Curves without Crossing," *The American Statistician*, vol. 51, no. 2, pp. 186–192, 1997.
- [12] J. M. Chambers and T. J. Hastie, *Statistical Models in S*, Chapman & Hall, New York, USA, 1993.
- [13] V. Miranda-Soberanis, *Vector Generalized Linear Time Series Models with an Implementation in R [Ph.D. thesis]*, Department of Statistics, University of Auckland, New Zealand, 2018.
- [14] T. W. Yee and C. J. Wild, "Vector generalized additive models," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 3, pp. 481–493, 1996.
- [15] P. Feigl and M. Zelen, "Estimation of exponential survival probabilities with concomitant information," *Biometrics*, vol. 21, no. 4, pp. 826–838, 1965.
- [16] P. K. Dunn and G. K. Smyth, *Generalized Linear Models with Examples in R*, Springer, Springer, New York, 2018.
- [17] T. W. Yee and V. F. Miranda-Soberanis, "Vector generalized linear models and negative binomial regression," *Australian & New Zealand Journal of Statistics*, 2018.

Research Article

Group Identification and Variable Selection in Quantile Regression

Ali Alkenani ¹ and Basim Shlaibah Msallam²

¹Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq

²Supervision & Scientific Evaluation Office, Ministry of Higher Education and Scientific Research, Iraq

Correspondence should be addressed to Ali Alkenani; ali.alkenani@qu.edu.iq

Received 24 November 2018; Revised 21 February 2019; Accepted 28 March 2019; Published 10 April 2019

Guest Editor: Himel Mallick

Copyright © 2019 Ali Alkenani and Basim Shlaibah Msallam. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Using the Pairwise Absolute Clustering and Sparsity (PACS) penalty, we proposed the regularized quantile regression QR method (QR-PACS). The PACS penalty achieves the elimination of insignificant predictors and the combination of predictors with indistinguishable coefficients (IC), which are the two issues raised in the searching for the true model. QR-PACS extends PACS from mean regression settings to QR settings. The paper shows that QR-PACS can yield promising predictive precision as well as identifying related groups in both simulation and real data.

1. Introduction

The regression model is one of the most important statistical models. The ordinary least squares regression (OLS) estimates the conditional mean function of the response. The least absolute deviation regression (LADR) estimates the conditional median function and it is resistant to outliers. The QR was introduced by Koenker and Bassett [1] as a generalization of LADR to estimate the conditional quantile function of the response. Consequently, QR gives us much more information about the conditional distribution of the response. QR has attracted a vast amount of interest in literature. It is applied in many different areas such as economics, finance, survival analysis, and growth chart.

Variable selection (VS) is very important in the process of model building. In many applications, the number of variables is huge. However, keeping irrelevant variables in the model is undesirable because it makes the model difficult to interpret and may affect negatively its ability of prediction. Many different penalties were suggested to achieve VS. For example, Lasso [2], SCAD [3], fused Lasso [4], elastic-net [5], group Lasso [6], adaptive Lasso [7], adaptive elastic-net [8], and MCP [9].

Under QR framework, Koenker [10] combined the Lasso with the mixed-effect QR model to encourage shrinkage in estimating the random effects. Wang, Li, and Jiang [11] combined LADR with the adaptive Lasso penalty. Li and Zhu [12] proposed L1-norm penalized QR (PQR) by combining QR with Lasso penalty. Wu and Liu [13] introduced PQR with the SCAD and the adaptive Lasso penalties. Slawski [14] proposed the structured elastic-net regularizer for QR.

In the setting $p > n$, where p represents the number of predictors and n represents the sample size, Belloni and Chernozhukov [15] studied the theory of PQR for the Lasso penalty. They considered QR in high-dimensional sparse models. Wang, Wu, and Li [16] investigated the methodology of PQR in ultrahigh dimension for the nonconvex penalties such as SCAD or MCP. Peng and Wang [17] proposed and studied a new iterative coordinate descent algorithm for solving nonconvex PQR in high dimension.

The search for the true model focuses on two issues: deleting irrelevant predictors and merging predictors with IC [18]. Although the above penalties can achieve the first issue, they fail in achieving the second one. The two issues can be achieved through Pairwise Absolute Clustering and

Sparsity (PACS) [18]. Moreover, PACS is an oracle method for simultaneous group identification and VS.

The limitations of existing variable selection methods motivate the authors to write this paper. The aim of the current research is to find an effective procedure for simultaneous group identification and VS under QR framework.

In this paper, we suggested the QR-PACS to get the advantages over the existing PQR methods. QR-PACS benefits from the ability of PACS on achieving the mentioned issues of the discovery of the true model which is unavailable in Lasso, Adaptive Lasso, SCAD, MCP, Elastic-net, and structured elastic-net.

The rest of the paper is organized as follows. In Section 2, penalized linear QR is reviewed briefly. QR-PACS is introduced in Section 3. The numerical results of simulations and real data are presented in Sections 4 and 5, respectively. The conclusions are reported in Section 6.

2. Penalized Linear QR

QR is a widespread technique used to describe the distribution of an outcome variable (y_i), given a set of predictors (\mathbf{x}_i). Let \mathbf{x}_i be a $p \times 1$ vector of predictors for the i th observation and $q_\tau(\mathbf{x}_i)$ be the inverse cumulative distribution function of y_i given \mathbf{x}_i . Then, $q_\tau(\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_\tau = \sum_{j=1}^p x_{ij} \beta_{j,\tau}$, where $\boldsymbol{\beta}_\tau$ is a vector of p unknown parameters and τ is the level of quantile.

Koenker and Bassett [1] suggested estimating $\boldsymbol{\beta}_\tau$ as follows:

$$\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^n \rho_\tau \left(y_i - \sum_{j=1}^p x_{ij} \beta_{j,\tau} \right), \quad (1)$$

where $\rho_\tau(\cdot)$ is the check loss function defined as

$$\rho_\tau(u) = \tau u I_{[0,\infty)}(u) - (1-\tau) u I_{(-\infty,0)}(u). \quad (2)$$

Under regularization framework, Li and Zhu [12], Wu and Liu [13], Slawski [14], and Wang, Wu, and Li [16] among others proposed the penalized versions of (1) by adding different penalties as follows:

$$\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^n \rho_\tau \left(y_i - \sum_{j=1}^p x_{ij} \beta_{j,\tau} \right) + \lambda \sum_{j=1}^p P(\beta_{j,\tau}) \quad (3)$$

where $\lambda > 0$ is the penalization parameter and $P(\beta_k)$ is the penalty function.

For the rest of this paper, the subscript τ is omitted for notational convenience.

3. Penalized Linear QR through PACS (QR-PACS)

In this section, we incorporate PACS into the optimization of (1) to propose QR-PACS. Under the QR setup, the predictors x_{ij} are standardized and the response y_i is centered, $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, p$. The QR-PACS is proposed for simultaneous group identification and VS in QR. The QR-PACS encourages correlated variable to have equal coefficient

values. The equality of coefficients is attained by adding group identification penalty to the pairwise differences and sums of coefficients. The QR-PACS estimates are proposed as minimizers of

$$\begin{aligned} & \sum_{i=1}^n \rho \left(y_i - \sum_{j=1}^p x_{ij} \beta_j \right) + \lambda \left\{ \sum_{j=1}^p \omega_j |\beta_j| \right. \\ & \left. + \sum_{1 \leq j < k \leq p} \omega_{jk(-)} |\beta_k - \beta_j| + \sum_{1 \leq j < k \leq p} \omega_{jk(+)} |\beta_k + \beta_j| \right\}, \quad (4) \\ & \dots \end{aligned}$$

where ω are the nonnegative weights.

The PACS penalty in (4) consists of $\lambda \{ \sum_{j=1}^p \omega_j |\beta_j| \}$ that encourages sparseness, $\lambda \{ \sum_{1 \leq j < k \leq p} \omega_{jk(-)} |\beta_k - \beta_j| \}$ and $\{ \sum_{1 \leq j < k \leq p} \omega_{jk(+)} |\beta_k + \beta_j| \}$, which are employed for the group identification and encourage equality of coefficients. The second term of the penalty encourages the same sign coefficients to be set as equal, while the third term encourages opposite sign coefficients to be set as equal in magnitude.

Choosing appropriate adaptive weights is very important for PACS. In QR-PACS, we employed the adaptive weights that incorporate correlations into the weights as suggested by Sharma et al. [18] with a small modification as follows:

$$\begin{aligned} \omega_j &= |\tilde{\beta}_j|^{-1}, \\ \omega_{jk(-)} &= (1 - r_{bjk})^{-1} |\tilde{\beta}_k - \tilde{\beta}_j|^{-1} \\ \text{and } \omega_{jk(+)} &= (1 + r_{bjk})^{-1} |\tilde{\beta}_k + \tilde{\beta}_j|^{-1} \end{aligned} \quad (5)$$

for $1 \leq j < k \leq p, \dots$

where $\tilde{\beta}$ is a \sqrt{n} consistent estimator of β , such as the PACS [18] estimates or other shrinkage QR estimates, and r_{bjk} is the biweight midcorrelation (j, k)th pair of predictors. We propose to employ the biweight midcorrelation [19, 20] instead of Pearson correlation which is used in the adaptive weights in [18] to obtain robust correlation and robust weights.

In this paper, ridge quantile estimates were employed as initial estimates for β 's to obtain weights performing well in studies with collinear predictors.

4. Simulation Study

In this section, five examples were carried out to assess QR-PACS method by comparing it with existing selection approaches under QR setting in both prediction precision and model discovery. A regression model was generated as follows.

$$y = X\beta + \sigma\epsilon \quad (6)$$

In all examples, predictors X and the error term ϵ were standard normal.

TABLE 1: ME, SA, GA, and SGA results of Example 1 for n=100.

τ	Criterion	Ridge QR	QR-LASSO	QR-SCAD	QR-adaptive LASSO	QR-elastic-net	QR-PACS
0.10	ME(S.E)	0.1144(0.0176)	0.0834(0.0182)	0.0684(0.0165)	0.0673(0.0153)	0.0623(0.0112)	0.0432(0.0093)
	SA	0	62	79	80	87	79
	GA	0	0	0	0	0	88
	SGA	0	0	0	0	0	72
$\sigma=1$ 0.50	ME(S.E)	0.0832(0.0056)	0.0524(0.0065)	0.0443(0.0056)	0.0431(0.0050)	0.0404(0.0041)	0.0143(0.0020)
	SA	0	64	81	82	86	80
	GA	0	0	0	0	0	89
	SGA	0	0	0	0	0	73
0.75	ME(S.E)	0.1140(0.0173)	0.0829(0.0177)	0.0672(0.0150)	0.0668(0.0146)	0.0619(0.0105)	0.0429(0.0082)
	SA	0	63	81	81	88	80
	GA	0	0	0	0	0	89
	SGA	0	0	0	0	0	72
0.10	ME(S.E)	0.1471(0.0183)	0.1168(0.0188)	0.1062(0.0162)	0.1056(0.0162)	0.0953(0.0120)	0.0755(0.0106)
	SA	0	59	77	77	84	77
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	70
$\sigma=3$ 0.50	ME(S.E)	0.1153(0.0066)	0.0855(0.0071)	0.0763(0.0060)	0.0756(0.0060)	0.0730(0.0049)	0.0473(0.0034)
	SA	0	61	80	80	83	78
	GA	0	0	0	0	0	87
	SGA	0	0	0	0	0	71
0.75	ME(S.E)	0.1466(0.0179)	0.1161(0.0185)	0.1056(0.0161)	0.1044(0.0158)	0.0945(0.0117)	0.0744(0.0099)
	SA	0	61	79	79	85	78
	GA	0	0	0	0	0	87
	SGA	0	0	0	0	0	70

We compared QR-PACS with ridge QR, QR-Lasso, QR-SCAD, QR-adaptive Lasso, and QR-elastic-net. The performance of the methods was compared using model error (ME) criterion for prediction accuracy which was defined by $(\hat{\beta} - \beta)' V (\hat{\beta} - \beta)$ where V represents the population covariance matrix of X and the resulting model complexity for model discovery. The median and standard error (SE) of ME were reported. Also, selection accuracy (SA, % of true models identified), grouping accuracy (GA, % of true groups identified), and % of both selection and grouping accuracy (SGA) were computed and reported. Note that none of ridge QR, QR-Lasso, QR-SCAD, QR-adaptive Lasso, and QR-elastic-net perform grouping. The sample size was 100 and the simulated model was replicated 100 times. Some typical examples are reported as follows.

Example 1. In this example, we assumed the true parameters for the model of study as $\beta = (2, 2, 2, 0, 0, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^8$. The first three predictors were highly correlated with correlation equal to 0.7 and their coefficients were equal in magnitude, while the rest were uncorrelated.

Example 2. In this example, the true coefficients were assumed as $\beta = (0.5, 1, 2, 0, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^8$. The first three predictors were highly correlated with correlation equal to 0.7 and their coefficients were different in magnitude, while the rest were uncorrelated.

Example 3. In this example, the true parameters were $\beta = (1, 1, 1, 0.5, 1, 2, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three predictors were highly correlated with correlation equal to 0.7 and their coefficients were equal in magnitude, while the correlation for the second three predictors was equal to 0.3 and their coefficients were different in magnitudes. The remaining predictors were uncorrelated.

Example 4. In this example, the true parameters were $\beta = (1, 1, 1, 0.5, 1, 2, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three predictors were correlated with correlation equal to 0.3 and their coefficients were equal in magnitude, while the correlation for the second three predictors was 0.7 and their coefficients were different in magnitudes. The remaining predictors were uncorrelated.

Example 5. In this example, the true parameters were assumed as $\beta = (2, 2, 2, 1, 1, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^{10}$. The first three and the second two predictors were highly correlated with correlation equal to 0.7 and their coefficients were different in magnitude, while the rest were uncorrelated.

For all the values of τ and σ , Table 1 shows that the QR-PACS method has the lowest ME. Although the QR-elastic-net and QR-adaptive Lasso have the highest SA, it is clear that all the considered methods do not perform grouping except QR-PACS. The QR-PACS method successfully identifies the groups of predictors as seen in the GA and SGA rows.

TABLE 2: ME, SA, GA, and SGA results of Example 2 for n=100.

τ	Criterion	Ridge QR	QR-LASSO	QR-SCAD	QR-adaptive LASSO	QR-elastic-net	QR-PACS
0.10	ME(S.E)	0.1170(0.0201)	0.0848(0.0195)	0.0696(0.0177)	0.0665(0.0165)	0.0639(0.0106)	0.0445(0.0101)
	SA	0	60	78	78	84	78
	NG	100	100	100	100	100	100
	SNG	0	69	70	70	78	69
$\sigma=1$ 0.50	ME(S.E)	0.0857(0.0075)	0.0538(0.0068)	0.0456(0.0062)	0.0439(0.0057)	0.0421(0.0055)	0.0156(0.0035)
	SA	0	63	81	81	84	79
	NG	100	100	100	100	100	100
	SNG	0	70	71	71	80	70
0.75	ME(S.E)	0.1155(0.0185)	0.0841(0.0179)	0.0681(0.0161)	0.0679(0.0155)	0.0630(0.0111)	0.0437(0.0090)
	SA	0	61	80	81	84	79
	NG	100	100	100	100	100	100
	SNG	0	70	70	71	79	70
0.10	ME(S.E)	0.1275(0.0213)	0.0968(0.0199)	0.0788(0.0185)	0.0745(0.0183)	0.0728(0.0106)	0.0419(0.0117)
	SA	0	57	77	77	83	77
	NG	100	100	100	100	100	100
	SNG	0	67	69	69	76	67
$\sigma=3$ 0.50	ME(S.E)	0.0876(0.0078)	0.0549(0.0072)	0.0464(0.0064)	0.0447(0.0062)	0.0430(0.0060)	0.0156(0.0040)
	SA	0	63	80	80	83	79
	NG	100	100	100	100	100	100
	SNG	0	69	70	70	79	69
0.75	ME(S.E)	0.1164(0.0190)	0.0850(0.0187)	0.0692(0.0170)	0.0688(0.0162)	0.0639(0.0120)	0.0446(0.0098)
	SA	0	60	79	80	84	79
	NG	100	100	100	100	100	100
	SNG	0	69	70	70	78	69

In Table 2, the percentage of no-grouping (NG, no groups found) and percentage of selection and no-grouping (SNG) were reported instead of GA and SGA, respectively. In terms of prediction and selection, the QR-PACS method does not perform well, while the QR-elastic-net, QR-adaptive Lasso, and QR-SCAD perform the best, respectively. All the methods under consideration perform well in terms of not identifying the group. Thus, the QR-PACS is not a recommended method when there is high correlation and the significant variables do not form a group.

Table 3 demonstrates that the QR-elastic-net and QR-adaptive Lasso have the best SA, respectively; however, the QR-PACS performs better in terms of ME. It is obvious that QR-PACS identifies the important group with high GA and SGA.

Table 4 shows that the QR-elastic-net, QR-adaptive Lasso, and QR-SCAD have the best SA. It is clear that the QR-PACS has the best results among the other methods in terms of ME. In terms of GA and SGA, it can be observed that the QR-PACS performs well.

From Table 5, it can be noticed that the QR-elastic-net has the best SA. The QR-PACS has excellent GA. Also, it is clear that QR-PACS successfully identifies the groups of predictors as seen in the GA and SGA.

5. NCAA Sports Data

In this section, the behavior of the QR-PACS with ridge QR, QR-Lasso, QR-SCAD, QR-adaptive Lasso, and QR-elastic-net

was illustrated in the analysis of NCAA sports data [21]. We standardized the predictors and centered the response before the data analysis.

In each repetition, the authors randomly split the data into a training and a testing dataset, the percentage of the testing data was 20%, and models were fit onto the training set. The NCAA sports data were randomly split 100 times each to allow for more stable comparisons. We reported the average and SE of the ratio of test error (RTE) over QR of all methods and the effective model size (MZ) after accounting for equality of absolute coefficient estimates.

The NCAA data was taken from a study of the effects of sociodemographic indicators and the sports programs on graduation rates.

The data size is n=94 and p=19 predictors. The dataset and its description are available from the website (<http://www4.stat.ncsu.edu/~boos/var.select/ncaa.html>). The predictors are students in top 10% HS (x_1), ACT COMPOSITE 25TH (x_2), On living campus (x_3), first-time undergraduates (x_4), Total Enrolment/1000 (x_5), courses taught by TAs (x_6), composite of basketball ranking (x_7), in-state tuition/1000 (x_8), room and board/1000 (x_9), avg BB home attendance (x_{10}), Professor Salary (x_{11}), Ratio of Student/faculty (x_{12}), white (x_{13}), Assistant professor salary (x_{14}), population of city (x_{15}), faculty with PHD (x_{16}), Acceptance rate (x_{17}), receiving loans (x_{18}), and Out of state (x_{19}).

From Table 6, the results indicate that QR-PACS does significantly better than ridge QR, QR-Lasso, QR-SCAD,

TABLE 3: ME, SA, GA, and SGA results of Example 3 for n=100.

τ	Criterion	Ridge QR	QR-LASSO	QR-SCAD	QR-adaptive LASSO	QR-elastic-net	QR-PACS
0.10	ME(S.E)	0.1383(0.0195)	0.1176(0.0196)	0.0933(0.0164)	0.0908(0.0151)	0.0902(0.0119)	0.0558(0.0098)
	SA	0	58	79	79	82	78
	GA	0	0	0	0	0	85
	SGA	0	0	0	0	0	70
$\sigma=1$ 0.50	ME(S.E)	0.1139(0.0071)	0.0843(0.0062)	0.0662(0.0059)	0.0643(0.0054)	0.0638(0.0051)	0.0367(0.0029)
	SA	0	62	80	81	84	79
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	71
0.75	ME(S.E)	0.1289(0.0180)	0.1083(0.0171)	0.0861(0.0161)	0.0843(0.0147)	0.0834(0.0108)	0.0479(0.0081)
	SA	0	61	80	80	84	79
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	71
0.10	ME(S.E)	0.1490(0.0199)	0.1279(0.0198)	0.1030(0.0169)	0.1016(0.0155)	0.0998(0.0124)	0.0651(0.0104)
	SA	0	57	78	78	80	77
	GA	0	0	0	0	0	83
	SGA	0	0	0	0	0	69
$\sigma=3$ 0.50	ME(S.E)	0.1244(0.0076)	0.0945(0.0065)	0.0867(0.0062)	0.0742(0.0056)	0.0735(0.0054)	0.0464(0.0032)
	SA	0	61	79	80	83	78
	GA	0	0	0	0	0	85
	SGA	0	0	0	0	0	70
0.75	ME(S.E)	0.1399(0.0185)	0.1189(0.0171)	0.0958(0.0166)	0.0946(0.0160)	0.0930(0.0111)	0.0567(0.0086)
	SA	0	60	79	80	83	78
	GA	0	0	0	0	0	85
	SGA	0	0	0	0	0	70

TABLE 4: ME, SA, GA, and SGA results of Example 4 for n=100.

τ	Criterion	Ridge QR	QR-LASSO	QR-SCAD	QR-adaptive LASSO	QR-elastic-net	QR-PACS
0.10	ME(S.E)	0.1380(0.0193)	0.1149(0.0195)	0.0939(0.0165)	0.0937(0.0153)	0.0832(0.0130)	0.0560(0.0105)
	SA	0	57	79	79	82	78
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	71
$\sigma=1$ 0.50	ME(S.E)	0.1132(0.0069)	0.0816(0.0062)	0.0672(0.0060)	0.0672(0.0056)	0.0568(0.0062)	0.0369(0.0036)
	SA	0	62	81	81	85	80
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	72
0.75	ME(S.E)	0.1280(0.0178)	0.1056(0.0171)	0.0873(0.0164)	0.0872(0.0149)	0.0764(0.0119)	0.0481(0.0088)
	SA	0	61	80	80	84	79
	GA	0	0	0	0	0	86
	SGA	0	0	0	0	0	71
0.10	ME(S.E)	0.1495(0.0200)	0.1270(0.0197)	0.1016(0.0169)	0.1008(0.0154)	0.0970(0.0120)	0.0639(0.0100)
	SA	0	57	78	78	81	77
	GA	0	0	0	0	0	84
	SGA	0	0	0	0	0	70
$\sigma=3$ 0.50	ME(S.E)	0.1239(0.0072)	0.0928(0.0064)	0.0859(0.0060)	0.0738(0.0054)	0.0730(0.0051)	0.0458(0.0029)
	SA	0	62	80	80	83	79
	GA	0	0	0	0	0	85
	SGA	0	0	0	07	0	70
0.75	ME(S.E)	0.1391(0.0181)	0.1182(0.0167)	0.0953(0.0163)	0.0941(0.0158)	0.0921(0.0106)	0.0559(0.0081)
	SA	0	60	79	80	83	78
	GA	0	0	0	0	0	85
	SGA	0	0	0	0	0	70

TABLE 5: ME, SA, GA, and SGA results of Example 5 for n=100.

τ	Criterion	Ridge QR	QR-LASSO	QR-SCAD	QR-adaptive LASSO	QR-elastic-net	QR-PACS	
$\sigma=1$	0.10	ME(S.E)	0.1330(0.0161)	0.1106(0.0168)	0.0914(0.0154)	0.0910(0.0142)	0.0818(0.0122)	0.0548(0.0981)
		SA	0	58	79	79	84	79
		GA	0	0	0	0	0	87
		SGA	0	0	0	0	0	72
	0.50	ME(S.E)	0.1119(0.0061)	0.0807(0.0053)	0.0666(0.0054)	0.0663(0.0051)	0.0561(0.0054)	0.0358(0.0031)
		SA	0	63	82	82	86	81
		GA	0	0	0	0	0	87
		SGA	0	0	0	0	0	73
	0.75	ME(S.E)	0.1274(0.0169)	0.1048(0.0162)	0.0867(0.0157)	0.0865(0.0143)	0.0756(0.0109)	0.0476(0.0079)
		SA	0	62	81	81	84	79
		GA	0	0	0	0	0	87
		SGA	0	0	0	0	0	72
$\sigma=3$	0.10	ME(S.E)	0.1486(0.0188)	0.1260(0.0189)	0.1001(0.0156)	0.1000(0.0148)	0.0960(0.0110)	0.0625(0.0088)
		SA	0	57	77	77	81	78
		GA	0	0	0	0	0	84
		SGA	0	0	0	0	0	70
	0.50	ME(S.E)	0.1227(0.0069)	0.0922(0.0059)	0.0850(0.0053)	0.0731(0.0047)	0.0724(0.0045)	0.0445(0.0020)
		SA	0	63	80	80	83	80
		GA	0	0	0	0	0	85
		SGA	0	0	0	0	0	70
	0.75	ME(S.E)	0.1385(0.0177)	0.1172(0.0154)	0.0940(0.0154)	0.0935(0.0151)	0.0926(0.0100)	0.0552(0.0075)
		SA	0	60	80	80	83	79
		GA	0	0	0	0	0	85
		SGA	0	0	0	0	0	70

TABLE 6: The test error and the effective model size values for the methods under consideration based on the NCAA sport data.

Method	$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.75$	
	RTE	MZ	RTE	MZ	RTE	MZ
Ridge QR	1.135 (0.035)	19	1.131 (0.037)	19	1.134 (0.034)	19
QR-LASSO	1.126(0.033)	6	1.123(0.032)	6	1.124(0.033)	6
QR-SCAD	1.115 (0.036)	6	1.111 (0.035)	6	1.113 (0.035)	6
QR-adaptive LASSO	1.110 (0.029)	5	1.101 (0.030)	5	1.108 (0.029)	5
QR-elastic-net	1.029(0.029)	5	1.022 (0.026)	5	1.025(0.030)	5
QR-PACS	0.913(0.014)	5	0.896(0.012)	5	0.912(0.015)	5

QR-adaptive Lasso, and QR-elastic-net in test error. In fact, ridge QR, QR-Lasso, QR-SCAD, QR-adaptive Lasso, and QR-elastic-net perform worse than the QR in test error. The effective model size is 5 for QR-PACS, although it includes all variables in the models.

6. Conclusions

In this paper, QR-PACS for group identification and VS under QR settings is developed, which combines the strength of QR and the ability of PACS for consistent group identification and VS. QR-PACS can achieve the two goals simultaneously. QR-PACS extends PACS from mean regression settings to QR settings. It is proved computationally that it can be simply

carried out with an effective computational algorithm. The paper shows that QR-PACS can yield promising predictive precision as well as identifying related groups in both simulation and the real data. Future direction or extension of the current paper is QR-PACS under Bayesian framework. Also, robust QR-PACS is another possible extension of the current paper.

Data Availability

The data which is studied in our paper is the NCAA sports data from Mangold et al. [21]. It is public and available from the website (<http://www4.stat.ncsu.edu/~boos/var.select/ncaa.html>), [21].

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.
- [3] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [4] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 67, no. 1, pp. 91–108, 2005.
- [5] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.
- [6] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [7] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [8] H. Zou and H. H. Zhang, "On the adaptive elastic-net with a diverging number of parameters," *The Annals of Statistics*, vol. 37, no. 4, pp. 1733–1751, 2009.
- [9] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.
- [10] R. Koenker, "Quantile regression for longitudinal data," *Journal of Multivariate Analysis*, vol. 91, no. 1, pp. 74–89, 2004.
- [11] H. Wang, G. Li, and G. Jiang, "Robust regression shrinkage and consistent variable selection through the LAD-Lasso," *Journal of Business & Economic Statistics*, vol. 25, no. 3, pp. 347–355, 2007.
- [12] Y. Li and J. Zhu, " L_1 -norm quantile regression," *Journal of Computational and Graphical Statistics*, vol. 17, no. 1, pp. 163–185, 2008.
- [13] Y. Wu and Y. Liu, "Variable selection in quantile regression," *Statistica Sinica*, vol. 19, no. 2, pp. 801–817, 2009.
- [14] M. Slawski, "The structured elastic net for quantile regression and support vector classification," *Statistics and Computing*, vol. 22, no. 1, pp. 153–168, 2012.
- [15] A. Belloni and V. Chernozhukov, L_1 -Penalized Quantile Regression in High-Dimensional Sparse Models, 2011.
- [16] L. Wang, Y. Wu, and R. Li, "Quantile regression for analyzing heterogeneity in ultra-high dimension," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 214–222, 2012.
- [17] B. Peng and L. Wang, "An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression," *Journal of Computational and Graphical Statistics*, vol. 24, no. 3, pp. 676–694, 2015.
- [18] D. B. Sharma, H. D. Bondell, and H. H. Zhang, "Consistent group identification and variable selection in regression with correlated predictors," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 319–340, 2013.
- [19] R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, San Diego, Calif, USA, 2nd edition, 1997.
- [20] A. Alkenani and K. Yu, "A comparative study for robust canonical correlation methods," *Journal of Statistical Computation and Simulation*, vol. 83, no. 4, pp. 690–720, 2013.
- [21] W. D. Mangold, L. Bean, and D. Adams, "The impact of inter-collegiate athletics on graduation rates among major NCAA division I universities: implications for college persistence theory and practice," *Journal of Higher Education*, vol. 74, no. 5, pp. 540–563, 2003.

Research Article

Retirement Consumption Puzzle in Malaysia: Evidence from Bayesian Quantile Regression Model

Ros Idayuwati Alaudin,¹ Noriszura Ismail ,² and Zaidi Isa²

¹*School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah DA, Malaysia*

²*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor DE, Malaysia*

Correspondence should be addressed to Noriszura Ismail; ni@ukm.edu.my

Received 12 August 2018; Accepted 13 December 2018; Published 1 January 2019

Guest Editor: Rahim Alhamzawi

Copyright © 2019 Ros Idayuwati Alaudin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The objective of this study is to use the Bayesian quantile regression for studying the retirement consumption puzzle, which is defined as the drop in consumption upon retirement, using the cross-sectional data of the Malaysian Household Expenditure Survey (HES) 2009/2010. Three different measures of consumption, namely, total expenditure, work-related expenditure, and nonwork-related expenditure, are suggested for studying the retirement consumption puzzle. The results show that the drop in consumption upon retirement is significant and has a regressive distributional effect as indicated by larger drops at lower percentiles and smaller drops at higher percentiles. The smaller drops among higher consumption retirees (or higher income retirees) may imply that they have more savings and/or retirement benefits than the smaller consumption retirees (or lower income retirees). Comparison between the three types of consumption shows that the work-related expenditure has a uniform drop across the distribution. The drop under the nonwork-related expenditure varies across the distribution, implying that it is the source behind the variation of the consumption drop.

1. Introduction

A life cycle theory is a major economic theory that relates consumption and saving behavior, which states that individuals desire to maintain their level of consumption in their entire lifetime [1]. The marginal utility consumption should remain smooth throughout retirement transition because the change in income during retirement should be predictable [2]. On the contrary, a number of previous studies found a one-time significant drop in consumption in the early years of retirement, a situation which is known as retirement consumption puzzle.

Over the past three decades, several studies focusing on the smoothing or stable path of consumption have been carried out. Hamermesh [3], who was among the first to study retirement consumption puzzle in the United States, showed that an individual was unable to sustain the level of real consumption prior to retirement due to inadequate

retirement savings. Later, several more studies have been conducted in other countries such as Banks et al. [4], Hurd and Rohwedder [5], Schwerdt [6], Wakabayashi [7], and Battistin et al. [8]. A review on several studies related to the retirement consumption puzzle can be found in Attanasio and Weber [9]. There are a number of studies which found that the decline in consumption upon retirement is due to several reasons. As examples, Haider and Stephens [10], Smith [11], and Blau [12] found that the consumption drop is due to the unexpected retirement which resulted from illnesses, disabilities, or involuntarily unemployment. Blau [12] also developed a modified life cycle model which incorporated the uncertainty in the timing of retirement.

In terms of consumption measures, several studies utilized food expenditure as a prominent substitution for the actual consumption during retirement. However, Aguiar and Hurst [13] found that food expenditure is a poor proxy for the actual consumption because retirees consumed the

same quantity of food, as well as its quality, even when food expenditure has declined. The main reason is that the retirees consumed home production food in their retirement since they have more time to prepare their meal and survey for cheaper food. Later, Hurst et al. [14] and Fisher et al. [15] proposed a broader measure of consumption which can be categorized according to several different types of expenditure and showed that a broader consumption measure can be used to eliminate retirement consumption puzzle.

Previous studies on the drop in retirement consumption were mostly concentrated at the mean distribution, which covered only certain parts of distribution, and may lead to poor estimation of parameters especially in long-tailed distributions. The traditional mean model estimates only the average effects of the whole data and do not allow for an understanding of any potential distributional impacts (or heterogeneity potential). Furthermore, the mean regression model is based on least squares estimation and thus has a significant sensitivity (or is not robust) to outliers. Several studies have been carried out to study the distributional aspects of retirement consumption puzzle, including Bernheim et al. [16] who estimated the retirement consumption drop using subgroups of wealth and income replacement rate and Aguila et al. [17] who used low and high consumption households to examine the drop in retirement expenditures. Recently, Fisher and Marchand [18] used the quantile regression model to investigate the drop in consumption upon retirement.

Quantile regression model is different from the traditional mean regression model as it uses the Least Absolute Deviation (LAD), instead of the least square error, which is able to rectify the weaknesses prevalent in the usual regression framework. Quantile regression model also allows the impact of each regression parameter to be analysed based on different selected quantiles. In short, quantile regression model has several advantages; it is a distribution-free model which does not adhere to any distributional assumptions, it is robust to outliers, it does not require independence assumption, and it allows the analysis of regression parameters to be extended beyond central locations.

Quantile regression model is based on the works of Koenker and Bassett [19] and Koenker and Hallock [20] and is gaining rapid interests by other researchers. The model has been developed for linear models with continuous responses and has been applied in various fields such as finance and economics [21], ecology [22], environmental epidemiology [23], criminology [24], and climate change [25]. Several extensions to the quantile regression model have been suggested and applied in other areas, such as Yu and Moyeed [26] who proposed Bayesian quantile regression model, Machado and Silva [27] who proposed quantile regression model for discrete data, Hewson and Yu [28] who suggested quantile regression model for binary data within the Bayesian framework, Reich et al. [25] who introduced Bayesian spatial quantile regression model, and Fuzi et al. [29] who applied Bayesian quantile regression model for claim count data in insurance area.

In this paper, we use the Bayesian quantile regression model to examine the drop in consumption upon retirement. The Bayesian quantile regression has the combined

advantages of both the quantile regression and the Bayesian approach. The quantile regression is a distribution-free model and robust to data, while the Bayesian approach allows the complete univariate and joint posterior distribution of each parameter to be generated by the MCMC simulations. Several motivational examples are worth mentioning here. For instance, in the field of environmental study, the quantile regression model allows the investigation of whether the effects of environmental exposure change depending on the level of respiratory health of the population [23], while in the area of insurance pricing and ratemaking, the Bayesian quantile regression models handle the parameters of covariates (or the risk factors) as random variables [29]. We use a cross-sectional data of Household Expenditure Survey (HES) 2009/2010 in Malaysia to investigate the changes in consumption drop across the population. We also expand the consumption measure into three categories, namely, total expenditure (TE), work-related expenditure (WRE), and nonwork-related expenditure (NWRE), to identify the source behind the variation of consumption drop.

2. Materials and Methods

2.1. Mean Regression. In our study, the drop in mean consumption upon retirement is estimated using the ordinary least square (OLS):

$$\ln(C_i) = \alpha + \beta \cdot Retired_i + \mathbf{x}_i^T \boldsymbol{\gamma} + \varepsilon_i \quad (1)$$

where $\ln(C_i)$ is the log of response variable (consumption measures), $Retired_i$ is the binary variable which equals one for retirees and zero for working households, and \mathbf{x}_i is the vector of control variables consisting of demographic and socioeconomic variables (gender, marital status, ethnic, etc.). The regression parameter, β , represents the mean consumption difference between working and retired households.

2.2. Frequentist Quantile Regression. The frequentist quantile regression model from Koenker and Bassett [19] is used in our study for comparison purposes. The same equation shown in (1) is used, but the quantile regression is now fitted to the conditional differences in the log of consumption between working households (preretirement) and retirees (postretirement) at the θ th quantile.

Let \mathbf{Y} be the vector of continuous response variables, which in our case can be represented by the three different consumption measures, and \mathbf{x}_i be the associated row-vector of covariates consisting of demographic and socioeconomic characteristics of households. The classical regression model focuses on the expectation of variable Y , conditional on the values of variables \mathbf{X} , which can be summarized as $E(Y_i | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}$. On the other hand, the quantile regression model extends this approach to $Q_{Y_i}(\theta | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_\theta$, where the quantile θ are fixed values between 0 and 1, allowing us to study the conditional distribution of Y on \mathbf{X} at different locations.

The regression parameter, $\boldsymbol{\beta}_\theta$, can be obtained using

$$\min_{\boldsymbol{\beta}_\theta} \sum p_\theta [\ln(C_i) - \mathbf{x}_i^T \boldsymbol{\beta}_\theta] \quad (2)$$

where the loss function $p_\theta(u) = u(\theta - I[u < 0])$ is a piecewise linear function and $I[\cdot]$ is the indicator function. Equivalently, the loss function can be written as

$$p_\theta(u) = u(\theta I[u > 0] - (1 - \theta) I[u < 0]). \quad (3)$$

Equation (3) can be minimized using linear programming, while the confidence interval can be obtained using bootstrap method. R statistical program with *quantreg* package [30] is used in this study to fit the frequentist quantile regression model.

2.3. Bayesian Quantile Regression. The Bayesian approach for quantile regression model was introduced by Yu and Moyeed [26] who formed the likelihood function using the asymmetric Laplace distribution (ALD). A random variable Y follows the ALD when the density function is given by

$$f_\theta(y | \mu, \sigma) = \frac{\theta(1 - \theta)}{\sigma} \exp \left\{ -p_\theta \left(\frac{y - \mu}{\sigma} \right) \right\} \quad (4)$$

where $0 < \theta < 1$ and $p_\theta(\cdot)$ is the loss function in equation (3). Under the Bayesian approach, the MCMC can be used to obtain the posterior distributions of the unknown parameters. The posterior distribution of parameter β is given by

$$\pi(\beta | y) \propto L(y | \beta) p(\beta) \quad (5)$$

where $p(\beta)$ is the prior distribution of β and $L(y | \beta)$ is the likelihood function which is formed by joining the independently distributed ALD. The joint likelihood function can be written as

$$L(y | \beta) = \frac{\theta^n (1 - \theta)^n}{\sigma(\theta)^n} \exp \left\{ -\sum_{i=1}^n p_\theta \left(\frac{y_i - \mathbf{x}_i^T \beta(\theta)}{\sigma(\theta)} \right) \right\}. \quad (6)$$

Since there is no specific conjugate prior distribution for generating the posterior distribution, we use the uniform prior distribution for all $\beta(\theta)$ in our study. The prior distribution for the scale parameter is the inverse-gamma distribution, or $\sigma(\theta) \sim \text{inverse-gamma}(a, b)$, which allows the Gibbs sampling algorithm in the MCMC to update and tune $\sigma(\theta)$ for obtaining good acceptance rates.

3. Results and Discussion

3.1. Sample Data. The sample data from Household Expenditure Survey (HES) 2009/2010 is used in our study. The data contains information on monthly expenditure, together with demographic and socioeconomic characteristics of each household. The selected sample is a cross-sectional data consisting of 6480 household heads.

Three different measures of consumption are used for the response variable, namely, total expenditure (TE), work-related expenditure (WRE), and nonwork-related expenditure (NWRE). TE consists of food at home, alcohol and cigarettes, home appliances and furniture, clothing, education, entertainment and recreation, health, insurance, outside

food (restaurant and café), transportation (own vehicle and public transport), personal care, rental, utility, and other services (such as legal services, tax services, and government agency). WRE and NWRE are constructed using the definitions in Aguiar and Hurst [31] and Fisher and Marchand [18]. WRE consists of outside food, personal care, public transport, and clothing, whereas NWRE contains food at home, alcohol, utilities, and entertainment.

3.2. Model Development. The MCMC simulations via Gibbs sampling algorithm are used to generate 5,000 posterior samples of each regression parameter. The first 1,000 runs of posterior samples are discarded as burn-ins to lessen the effect of initial simulations, and the process resulted in 4,000 final posterior samples for each regression parameter. The values of β are initialized at zero, while the inverse-gamma parameters are set at $a = 0.01$ and $b = 0.01$.

3.3. Summary Statistics. Table 1 provides the summary statistics for the sample. More than half of household heads (69%) are located in urban areas, a large majority (78%) are married, and more than half (64%) belongs to Bumiputera ethnic. In terms of educational attainment, 14% of household heads are university or college graduates, 35% are high school graduates, and the balances are below high school education (16%), and others (35%) (attend informal or religious education). For occupational groups, almost 13% of household heads are professionals, while 19% and 13% are administrative supports and technicians, respectively. In terms of employment type, almost half of household heads are private sector employees, followed by self-employed (19%) and government employees (13%). The proportions of age group are quite equally divided, with the exception of younger household heads (age less than 25). With regard to the status of living quarters, more than half (67%) are homeowners, followed by renters (24%).

3.4. Mean Regression. Table 2 provides the estimate, β , for the mean regression model, which can be used to show the difference (in mean consumption) between working and retired households. Three response variables are considered, namely, total expenditure (TE), work-related expenditure (WRE), and nonwork-related expenditure (NWRE).

The mean consumption among retirees is lower compared to the working households for all expenditure measures as shown by the negative coefficient. The regression estimate for TE is significant and indicates that the consumption for retirees is 14% lower than the working households. The estimates for WRE and NWRE are also significant, showing that the retirees' consumptions are 39% and 5% lower than the working households, respectively. As expected, the WRE has a relatively larger drop than the NWRE, indicating that retirees are no longer involved in employment. The results also agree with studies from Fisher and Marchand [18] and Fisher et al. [15] who found that the WRE has the highest consumption drop. The results indicate that a broader measure of consumption may diminish the retirement consumption puzzle, as indicated by the smaller drop in the NWRE.

TABLE 1: Summary statistics for HES Sample 2009.

Variable	Proportion of households (%)
Region	
1 (Kelantan, Pahang, Terengganu)	15
2 (Johor, Melaka, Negeri Sembilan)	16
3 (Kedah, Perak, Perlis)	20
4 (P.Pinang, Selangor, Kuala Lumpur, Putrajaya)	27
5 (Sabah, Sarawak)	22
Strata	
Urban	69
Rural	31
Marital status	
Married	78
Single Female	12
Single Male	10
Ethnic	
Bumiputera	64
Non Bumiputera	36
Educational level	
College/University	14
High School Grad	35
Less than High School	16
Others (not attending formal education, religious education, not finishing school)	35
Occupational group	
Professionals and Legislators	13
Administrative Supports	19
Technicians	13
Agriculture and Fishery	10
Craft and Repair	9
Elementary Occupations	10
Operators	11
Others (housewife, unemployed, disabled)	15
Employment type	
Employer	2.8
Government Professional and Administrative	7.0
Government Technicians and below	6.3
Private Professional and Administrative	16.4
Private Technicians and below	33.4
Self-employed	19.0
Others (e.g. pensioners)	15.1
Subjective life expectancy	
live \leq 25	4
25 \leq live \leq 34	18
35 \leq live \leq 44	26
45 \leq live \leq 54	26
55 \leq live	26
Status living quarters	
Owned	67
Rented	24
Quarters	5
Others (e.g. squatters owned, squatters rented)	4

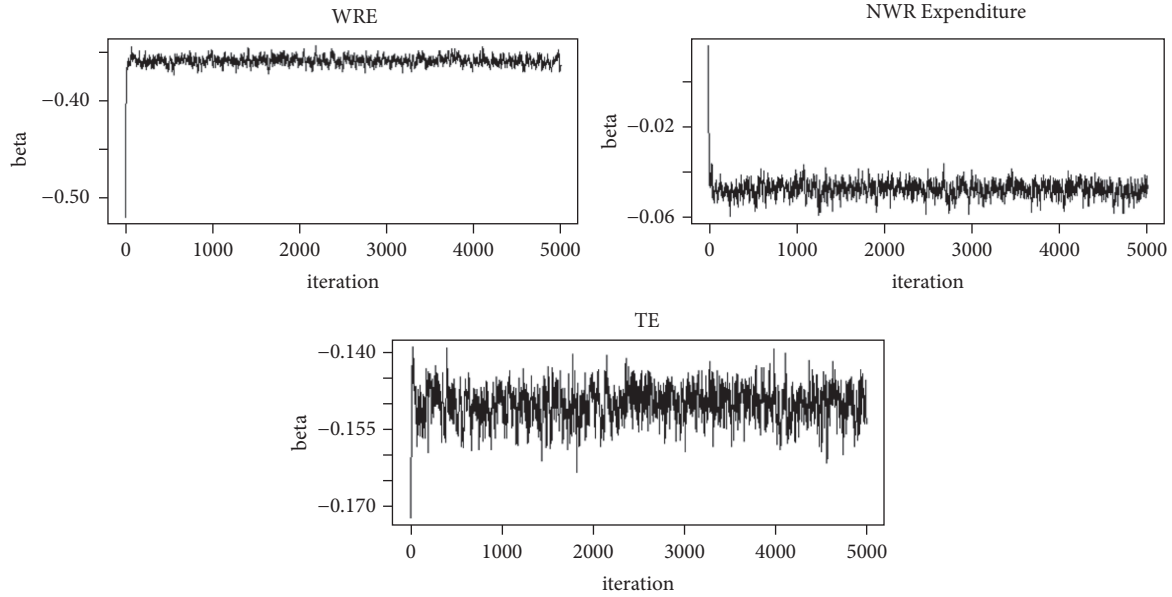


FIGURE 1: Trace plots for regression estimate under Bayesian median regression ($\theta = 0.5$).

TABLE 2: Regression estimate for mean regression model.

Expenditure Measure	Estimate	SE
Work-Related Expenditures (WRE)	-0.385	0.026
Non Work-Related Expenditures (NWR)	-0.050	0.017
Total Expenditures (TE)	-0.139	0.018

3.5. *Bayesian Quantile Regression.* The trace plots for the regression estimate under the Bayesian median regression ($\theta = 0.5$) are provided in Figure 1. The trace plots for other quantiles are also obtained but are not shown here. The trace plots show that the MCMC simulations mix well, the convergence of the posterior distributions took place, and there were no significant problems in the chain simulations.

Table 3 shows the regression estimates (and standard deviations) under the Bayesian quantile regression. It can be seen that the estimates for the three consumption measures are significant at all quantiles. The drops in consumption also differ across the distribution. All three consumption measures show larger drops at lower percentiles and smaller drops at higher percentiles (regressive trend), where the smallest and highest drops are at the highest and lowest percentiles respectively. The regressive trend disagrees with the results of Fisher and Marchand [18] who found that the estimates are more negative (progressive trend) when moving towards the upper distribution. However, the regressive trend is consistent with the results of Aguila et al. [17] who found that the retirement consumption has larger drops at lower percentiles.

The TE displays the largest drop (22%) at the 10th percentile, a drop of 15% at the median, and the smallest drop (3%) at the 90th percentile. Similar patterns are also seen in the WRE and NWR: 60%, 36%, and 17% drops, respectively, at the 10th, median, and 90th percentiles for the WRE and

11%, 5%, and 2% drops, respectively, at the 10th, median, and 90th percentiles for the NWR. In terms of magnitude, the WRE shows the largest drop at all quantiles, and the result agrees with the mean model which shows that the largest drop is from the WRE. As expected, the consumption drops under the Bayesian median model ($\theta=0.50$) that is quite comparable to the OLS model. The drops for TE, WRE, and NWR are 14%, 39%, and 5%, respectively, under the OLS and 15%, 36%, and 5%, respectively, under the Bayesian median model, indicating that the median model can be used as an alternative to the mean model (OLS).

It can be observed that the drop in WRE follows a uniform trend across the distribution, while the NWR has more variations at lower percentiles (from 0.10 to 0.50). The variations show that the NWR is the source behind the variation in the consumption drop. Our result agrees with Fisher and Marchand [18] who found that the WRE displays a uniform drop across the distribution, and the NWR is the source behind the variation of the consumption drop.

Figure 2 exhibits the plots of regression estimates with their respective 95% credible intervals under the Bayesian quantile regression. For comparison purpose, the estimates from the OLS are also included, represented by the dashed horizontal line. It can be seen that the estimates under the NWR have more variations in the lower percentiles (from 0.10 to 0.50), while the estimates under the WRE are uniformly increasing. The consistently small widths of the credible intervals throughout the quantiles indicate that the estimates for the three consumption measures are significant throughout the distribution.

3.6. *Frequentist Quantile Regression.* For comparison purpose, Table 4 shows the estimates (and standard errors) under the frequentist quantile regression model for the three different measures of consumption. Comparison between

TABLE 3: Regression estimate for Bayesian quantile regression model.

Consumption Measure	$\theta = 0.10$				$\theta = 0.25$				$\theta = 0.50$			
	Est	LB	UB	SD	Est	LB	UB	SD	Est	LB	UB	SD
WRE	-0.596	-0.608	-0.584	0.006	-0.480	-0.491	-0.470	0.005	-0.359	-0.367	-0.352	0.004
NWRE	-0.112	-0.123	-0.101	0.006	-0.044	-0.051	-0.035	0.004	-0.047	-0.054	-0.041	0.003
TE	-0.223	-0.234	-0.211	0.006	-0.150	-0.157	-0.143	0.004	-0.150	-0.157	-0.145	0.003

Consumption Measure	$\theta = 0.75$				$\theta = 0.90$			
	Est	LB	UB	SD	Est	LB	UB	SD
WRE	0.004	-0.281	-0.289	-0.273	0.004	-0.167	-0.185	-0.151
NWRE	0.003	-0.021	-0.029	-0.014	0.004	-0.016	-0.007	-0.027
TE	0.003	-0.123	-0.130	-0.115	0.004	-0.029	-0.044	-0.016

TABLE 4: Regression estimate for frequentist quantile regression model.

Consumption measures	$\theta = 0.1$		$\theta = 0.25$		$\theta = 0.5$		$\theta = 0.75$		$\theta = 0.9$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
WRE	-0.595	0.060	-0.483	0.041	-0.361	0.033	-0.281	0.032	-0.162	0.045
NWRE	-0.108	0.030	-0.043	0.023	-0.047	0.019	-0.020	0.021	-0.020	0.032
TE	-0.222	0.036	-0.150	0.022	-0.148	0.020	-0.124	0.025	-0.024	0.036

Tables 3 and 4 shows that the estimates under the Bayesian and frequentist quantile regression models are similar for all quantiles. The main difference between both models is shown by the standard deviations and standard errors; the standard deviations under the Bayesian model are different from the standard errors under the frequentist model. The differences are expected since the estimates under both approaches are obtained under different estimation methods; the frequentist intervals are estimated via bootstrap method, whereas the Bayesian intervals are obtained from the MCMC simulation.

The smaller standard deviations under the Bayesian regression suggest that the model has more significant estimates. The estimates are statistically significant at all quantiles for all three consumption measures under the Bayesian model, while the frequentist model has several insignificant estimates at several quantiles.

Similar to the Bayesian quantile regression, all three consumption measures under the frequentist quantile regression show larger drops at lower percentiles and smaller drops at higher percentiles (regressive trend). The TE displays the largest drop (22%) at the 10th percentile and the smallest drop (12%) at the 75th percentile. The drop at the 90th percentile is insignificant. The WRE displays the largest drop (60%) at the 10th percentile and the smallest drop (16%) at the 90th percentile, while the NWRE displays the largest drop (11%) at the 10th percentile and the smallest drop (5%) at the median. The drops after the median are insignificant. The drop in WRE also follows a uniform trend, while the drop in NWRE has more variations.

4. Conclusions

In this study, we applied the Bayesian quantile regression to investigate the consumption drop upon retirement which is an area where, currently, the quantile regression model is of limited utilization. The Bayesian quantile regression

has the combined advantages of both quantile regression and Bayesian approach. In particular, the quantile regression is a distribution-free model and robust to data, while the Bayesian approach allows the complete univariate and joint posterior distribution of each parameter to be generated by the MCMC simulations. Our study also compared the estimates from the Bayesian quantile regression with the OLS (mean) and the frequentist quantile regression. We also considered three different consumption measures, namely, total expenditure (TE), work-related expenditure (WRE), and nonwork-related expenditure (NWRE).

The consumption drops in TE, WRE, and NWRE are 14%, 39%, and 5%, respectively, under the OLS (mean model), which agree with studies from Fisher and Marchand [18] and Fisher et al. [15] who found that the WRE has the highest drop, and the NWRE has the lowest drop. The results also prove that a broader measure of consumption may diminish the retirement consumption puzzle, as indicated by the smaller drop in the NWRE.

As expected, the drops in TE, WRE, and NWRE under the Bayesian median regression model ($\theta = 0.50$) are quite comparable to the OLS, indicating that the median model may be used as a substitute for the mean model.

The consumption drops upon retirement are statistically significant at all quantiles under the Bayesian quantile regression, where larger drops at lower percentiles and smaller drops at higher percentiles indicate a regressive distributional effect (regressive trend). The WRE shows a relatively uniform drop, while the drops in NWRE have more variations at lower percentiles (from 0.10 to 0.50), indicating that the NWRE drops are the source behind the variations of drops. Our study agrees with Aguila et al. [17] who found larger consumption drops at lower percentiles (regressive trend) but disagrees with Fisher and Marchand [18] who found larger consumption drops at higher percentiles (progressive trend).

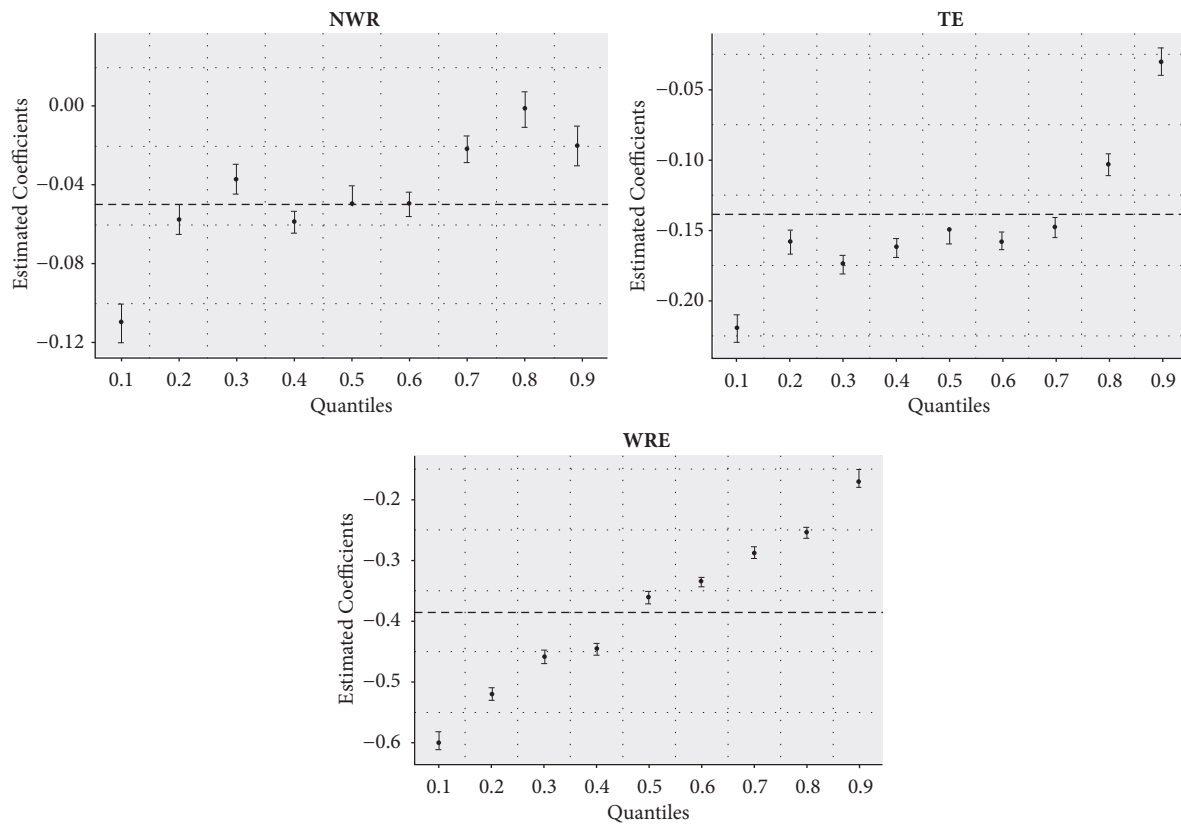


FIGURE 2: Regression estimates and 95% credible intervals for Bayesian quantile regression.

It should be noted that different data may provide different results.

The smaller drop among higher consumption retirees (or retirees with higher income) under the Bayesian quantile regression imply that the retirees with higher consumption have more savings and/or retirement benefits. The results are consistent with the expectations of life cycle theory which states that higher income households save more than lower income households. The larger drop at lower percentiles under the Bayesian quantile regression implies that the smaller consumption responses (or retirees with lower income) are exposed to larger consumption shocks.

Comparison between the Bayesian and the frequentist quantile regressions shows that the estimates are similar at all quantiles. The main difference between both models is that the standard deviations under the Bayesian model are different than the standard errors under the frequentist model. The differences are expected since the estimates under both approaches are obtained under different estimation methods.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This article uses sample data from Malaysian Household Expenditure Survey (HES) 2009/10 conducted by Department of Statistics Malaysia (DOSM). The authors gratefully acknowledge the financial support received in the form of research grants (GUP-2017-011) from Universiti Kebangsaan Malaysia.

References

- [1] F. Modigliani and R. Brumberg, "Utility analysis and the consumption function: An interpretation of cross-section data," *Post Keynesian Economics*, pp. 388–436, 2013.
- [2] A. Ando and F. Modigliani, "The Life Cycle Hypothesis of Saving?: Aggregate Implications and Tests," *American Economic Review*, vol. 53, pp. 55–84, 1963.
- [3] D. Hamermesh, "Consumption During Retirement: The Missing Link in the Life Cycle," National Bureau of Economic Research w0930, 1982.
- [4] J. Banks, R. Blundell, and S. Tanner, "Is there a retirement-savings puzzle?" *American Economic Review*, vol. 88, no. 4, pp. 769–788, 1998.
- [5] M. Hurd and S. Rohwedder, "The Retirement-Consumption Puzzle: Anticipated and Actual Declines in Spending at Retirement," National Bureau of Economic Research w9586, 2003.
- [6] G. Schwerdt, "Why does consumption fall at retirement? Evidence from Germany," *Economics Letters*, vol. 89, no. 3, pp. 300–305, 2005.

- [7] M. Wakabayashi, "The retirement consumption puzzle in Japan," *Journal of Population Economics*, vol. 21, no. 4, pp. 983–1005, 2008.
- [8] E. Battistin, A. Brugiavini, E. Rettore, and G. Weber, "The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach," *American Economic Review*, vol. 99, no. 5, pp. 2209–2226, 2009.
- [9] O. P. Attanasio and G. Weber, "Consumption and saving: Models of intertemporal allocation and their implications for public policy," *Journal of Economic Literature (JEL)*, vol. 48, no. 3, pp. 693–751, 2010.
- [10] S. J. Haider and M. Stephens, "Is There a Retirement-Consumption Puzzle? Evidence Using Subjective Retirement Expectations," *Review of Economics and Statistics*, vol. 89, no. 2, pp. 247–264, 2007.
- [11] S. Smith, "The Retirement-Consumption Puzzle and Involuntary Early Retirement: Evidence from the British Household Panel Survey," *The Economic Journal*, vol. 116, no. 510, pp. C130–C148, 2006.
- [12] D. Blau, "Retirement and Consumption in a Life Cycle Model," *Journal of Labor Economics*, vol. 26, no. 1, pp. 35–71, 2008.
- [13] M. Aguiar and E. Hurst, "Consumption versus Expenditure," *Journal of Political Economy*, vol. 113, no. 5, pp. 919–948, 2005.
- [14] E. Hurst, "The Retirement of a Consumption Puzzle," National Bureau of Economic Research w13789, 2008.
- [15] J. D. Fisher, D. S. Johnson, J. Marchand, T. M. Smeeding, and B. B. Torrey, "The retirement consumption conundrum: Evidence from a consumption survey," *Economics Letters*, vol. 99, no. 3, pp. 482–485, 2008.
- [16] B. D. Bernheim, J. Skinner, and S. Weinberg, "What accounts for the variation in retirement wealth among U.S. households?" *American Economic Review*, vol. 91, no. 4, pp. 832–857, 2001.
- [17] E. Aguila, O. Attanasio, and C. Meghir, "Changes in consumption at retirement: Evidence from panel data," *Review of Economics and Statistics*, vol. 93, no. 3, pp. 1094–1099, 2011.
- [18] J. D. Fisher and J. T. Marchand, "Does the retirement consumption puzzle differ across the distribution?" *The Journal of Economic Inequality*, vol. 12, no. 2, pp. 279–296, 2014.
- [19] R. Koenker and G. Bassett Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
- [20] R. Koenker and K. F. Hallock, "Quantile regression," *Journal of Economic Perspectives (JEP)*, vol. 15, no. 4, pp. 143–156, 2001.
- [21] R. F. Engle and S. Manganelli, "CAViaR," *Journal of Business & Economic Statistics*, vol. 22, no. 4, pp. 367–381, 2004.
- [22] B. S. Cade and B. R. Noon, "A gentle introduction to quantile regression for ecologists," *Frontiers in Ecology and the Environment*, vol. 1, no. 8, pp. 412–420, 2003.
- [23] D. Lee and T. Neocleous, "Bayesian quantile regression for count data with application to environmental epidemiology," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 59, no. 5, pp. 905–920, 2010.
- [24] M. Wang and L. Zhang, "A Bayesian Quantile Regression Analysis of Potential Risk Factors for Violent Crimes in USA," *Open Journal of Statistics*, vol. 02, no. 05, pp. 526–533, 2012.
- [25] B. J. Reich, M. Fuentes, and D. B. Dunson, "Bayesian Spatial Quantile Regression," *Journal of the American Statistical Association*, vol. 106, no. 493, pp. 6–20, 2011.
- [26] K. Yu and R. A. Moyeed, "Bayesian quantile regression," *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.
- [27] J. A. Machado and J. M. Silva, "Quantiles for Counts," *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1226–1237, 2005.
- [28] P. Hewson and K. Yu, "Quantile regression for binary performance indicators," *Applied Stochastic Models in Business and Industry*, vol. 24, no. 5, pp. 401–418, 2008.
- [29] M. F. Fuzi, A. A. Jemain, and N. Ismail, "Bayesian quantile regression model for claim count data," *Insurance: Mathematics and Economics*, vol. 66, pp. 124–137, 2016.
- [30] R. Koenker, "quantreg," 2015.
- [31] M. Aguiar and E. Hurst, "Deconstructing life cycle expenditure," *Journal of Political Economy*, vol. 121, no. 3, pp. 437–492, 2013.