Model-based approaches to nonparametric Bayesian quantile regression

Athanasios Kottas[1], Milovan Krnjajić[2], Matthew Taddy[3]
University of California, Santa Cruz[1]
University of California, Lawrence Livermore National Laboratory[2]
University of California, Santa Cruz[3]

Abstract

In several regression applications, a different structural relationship might be anticipated for the higher or lower responses than the average responses. In such cases, quantile regression analysis can uncover important features that would likely be overlooked by mean regression. We develop two distinct Bayesian approaches to fully nonparametric model-based quantile regression. The first approach utilizes an additive regression framework with Gaussian process priors for the quantile regression functions and a scale uniform Dirichlet process mixture prior for the error distribution, which yields flexible unimodal error density shapes. Under the second approach, the joint distribution of the response and the covariates is modeled with a Dirichlet process mixture of multivariate normals, with posterior inference for different quantile curves emerging through the conditional distribution of the response given the covariates. The proposed nonparametric prior probability models allow the data to uncover non-linearities in the quantile regression function and non-standard distributional features in the response distribution. Inference is implemented using a combination of posterior simulation methods for Dirichlet process mixtures. We illustrate the performance of the proposed models using simulated and real data sets.

Keywords: Dirichlet process mixture models; Gaussian process priors; Multivariate normal mixtures; Scale uniform mixtures

1. Introduction

Quantile regression can be used to quantify the relationship between a set of quantiles of the response distribution and available covariates. Because, in general, a set of quantiles provides a more complete description of the response distribution than the mean, quantile regression offers a practically important alternative to traditional mean regression. In many regression examples (e.g., in econometrics, educational studies, and environmental applications), we might expect a different structural relationship for the higher (or lower) responses than the *average* responses. In such applications, mean, or median, regression approaches would likely overlook important features that could be uncovered by a more general quantile regression analysis.

There is a fairly extensive literature on classical estima-tion for the standard p-th quantile regression model, $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i$, where y_i denotes the response observations, $\boldsymbol{x}_i$ the corresponding covariate vectors, and ϵ_i the errors, which are typically assumed independent from a distribution (with density, say, $f_p(\cdot)$) that has p-th quantile equal to 0. (See, e.g., the review paper by Yu, Lu and Stander, 2003, and the book by Koenker, 2005.) This literature is dominated by *semiparametric* techniques where the error density $f_p(\cdot)$ is left unspecified (apart from the restriction $\int_{-\infty}^0 f_p(\epsilon) \mathrm{d}\epsilon = p$). Hence, since there is no probability model for the response distribution, point estimation for $\boldsymbol{\beta}$ proceeds by optimization of some *loss* function. For instance, under the standard setting with independent and uncensored responses, the point estimates for $\boldsymbol{\beta}$ min-imize $\sum_{i=1}^n \rho_p(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})$, where $\rho_p(u) = up - u1_{(-\infty,0)}(u)$; this form yields the least absolute deviations criterion for $p = 0.5$, i.e., for the special case of median regres-sion. Any inference beyond point estimation is based on asymptotic arguments or resampling methods. The clas-sical literature includes also work that relaxes the para-metric (linear) regression form for the quantile regression function (see, e.g., He, Ng and Portnoy, 1998; Horowitz and Lee, 2005).

By comparison with the existing volume of classical work, the Bayesian literature on quantile regression is relatively limited. The special case of median regression has been considered in Walker and Mallick (1999), Kot-tas and Gelfand (2001), and Hanson and Johnson (2002). This work is based on a parametric form for the me-dian regression function and nonparametric modeling for the error distribution, using either Pólya tree priors or Dirichlet process mixture priors. (See, e.g., Müller and Quintana, 2004, for reviews of these nonparametric prior models.) Regarding quantile regression, based again on parametric regression functions, Yu and Moyeed (2001) and Tsionas (2003) discuss parametric inference based on the asymmetric Laplace distribution for the errors; Kot-tas and Krnjajić (2005) develop Bayesian semiparametric models using Dirichlet process mixtures for the error dis-tribution; Hjort and Petrone (2005) study nonparametric inference for the quantile function based on Dirichlet pro-cesses, including brief discussion of the semiparametric extension to quantile regression. Moreover, Dunson and Taylor (2005) propose an approximate semi-Bayesian in-ference method for quantile regression, which, in contrast to the work discussed above, does not involve probabilis-tic modeling for the response distribution.

A practical limitation of the Bayesian semiparametric approaches developed in Walker and Mallick (1999), Kottas and Gelfand (2001), Hanson and Johnson (2002), and Kottas and Krnjajić (2005) is that, although they all provide flexible shapes for the error distribution, they are based on parametric (in fact, linear) quantile regression functions. Regarding inference for non-linear quantile regression functions, Scaccia and Green (2003) model the conditional distribution of the response given a single continuous covariate with a discrete normal mixture with covariate-dependent weights. Moreover, Yu (2002) discusses a semi-Bayesian estimation method based on a piecewise polynomial representation for the quantile regression function corresponding, again, to a single continuous covariate, but without a probability model for the error distribution. We note that both of these approaches involve rather complicated Markov chain Monte Carlo (MCMC) methods for inference, in particular, certain forms of reversible jump MCMC techniques.

To our knowledge, this paper presents the first attempt to develop a model-based, fully inferential framework for Bayesian nonparametric quantile regression. We argue for the utility of Bayesian modeling approaches, since they enable exact and full inference, given the data, for the quantile regression function as well as for any functional of the response distribution that may be of interest. But then the fexibility of such inference under nonparametric prior models becomes attractive. We propose two distinct modeling approaches to nonparametric quantile regression. The first approach (presented in Section 2) utilizes an additive regression framework with different nonparametric priors for the quantile regression functions and for the error distribution. Under the second approach (developed in Section 3), the joint distribution of the response and the covariates is modeled with a flexible nonparametric mixture model, and then inference for different quantile curves is obtained from the induced conditional distribution of the response given the covariates. We discuss MCMC posterior simulation methods under both approaches, and illustrate inferences with either simulated or real data sets. Section 4 concludes with a comparative discussion of the two proposed approaches to quantile regression.

2. Bayesian nonparametric modeling in an additive quantile regression framework

2.1 Model formulation

Here, we employ the additive nonparametric regression setting in the context of quantile regression. Hence, with $\boldsymbol{x} = (x_1, ..., x_L)$ denoting the covariate vector (comprising continuous variables), the p-th quantile regression function is expressed as a sum of covariate-specific regression functions $h_\ell(x_\ell)$, $\ell = 1, ..., L$. Without loss of generality for the model formulation, we consider the case with two (continuous) covariates, and therefore with y_i denoting

the response observations, the model becomes

$$y_i = h_1(x_{i1}) + h_2(x_{i2}) + \epsilon_i, \ \ i = 1, ..., n. \qquad (1)$$

The error terms ϵ_i are assumed independent from a distribution with p-th quantile equal to 0, i.e., $\int_{-\infty}^{0} f_p(\epsilon)\mathrm{d}\epsilon = p$, with $f_p(\cdot)$ denoting the error density.

Our objective is to combine flexible nonparametric prior models for the quantile regression function and the random error density.

In particular, we work with independent Gaussian process (GP) priors for $h_1(\cdot)$ and $h_2(\cdot)$. (See, e.g., Neal, 1997, 1998, on GP regression under parametric error distributions.) To avoid identifiability issues, we set the GP mean functions to zero, $\mathrm{E}(h_\ell(x)) = 0$, for all x. We have also observed empirically that an intercept in (1) is weakly identifiable, especially, for data sets that support skewness in the error distributions. Rather than working with a random intercept with a very informative prior, we opt to fix it to zero centering the response observations before fitting the model. For the GP covariance functions, we use the simple isotropic specification,

$$\mathrm{Cov}(h_\ell(x), h_\ell(x') \mid \tau_\ell^2, \phi_\ell) = \tau_\ell^2 \exp(-\phi_\ell|x - x'|^{a_\ell}), \ \ (2)$$

with random variances τ_ℓ^2 and *range* parameters $\phi_\ell > 0$, and fixed power parameters $a_\ell \in [1, 2)$, $\ell = 1, 2$.

For the random error density $f_p(\cdot)$ we use the scale uniform Dirichlet process (DP) mixture prior from Kottas and Krnjajić (2005). The key result for its construction is a *representation* for non-increasing densities on the positive real line. Specifically, for any non-increasing density $u(\cdot)$ on R^+ there exists a distribution function G, with support on R^+, such that $u(t) \equiv u(t; G) = \int \theta^{-1} 1_{[0,\theta)}(t)\mathrm{d}G(\theta)$, i.e., $u(\cdot)$ can be expressed as a scale mixture of uniform densities. The result requires a general mixing distribution G and thus, for Bayesian modeling, invites the use of a nonparametric prior for G; see, e.g., Brunner and Lo (1989), Brunner (1995), Lavine and Mockus (1995), and Kottas Gelfand (2001) for DP-based modeling involving variations of this representation.

In the context of the quantile regression setting in (1), this result can be used to provide a mixture representation for any unimodal density on the real line with p-th quantile (and mode) equal to zero, $\iint k_p(\epsilon; \sigma_1, \sigma_2)\mathrm{d}G_1(\sigma_1)\mathrm{d}G_2(\sigma_2)$. Here G_1 and G_2 are general mixing distributions, supported on R^+, and

$$k_p(\epsilon; \sigma_1, \sigma_2) = \frac{p}{\sigma_1}1_{(-\sigma_1,0)}(\epsilon) + \frac{(1-p)}{\sigma_2}1_{[0,\sigma_2)}(\epsilon),$$

with $0 < p < 1$, and $\sigma_r > 0$, $r = 1, 2$. Assuming independent DP priors, $\mathrm{DP}(\alpha_r, G_{r0})$, for G_r, $r = 1, 2$, we obtain the DP mixture model

$$f_p(\epsilon; G_1, G_2) = \iint k_p(\epsilon; \sigma_1, \sigma_2)\mathrm{d}G_1(\sigma_1)\mathrm{d}G_2(\sigma_2), \quad (3)$$

for the error density in (1).

Recall that the DP was developed by Ferguson (1973) as a prior probability model for random distributions

(equivalently, distribution functions) G. A $\mathrm{DP}(\alpha, G_0)$ prior for G is defined in terms of two parameters, a parametric base distribution G_0 (the mean of the process) and a positive scalar parameter α, which can be interpreted as a precision parameter; larger values of α result in realizations G that are closer to G_0. We will write $G \sim \mathrm{DP}(\alpha, G_0)$ to indicate that a DP prior is used for the random distribution G. In fact, DP-based modeling typically utilizes mixtures of DPs (Antoniak 1974), i.e., a more general version of the DP prior that involves hyperpriors for α and/or the parameters of G_0. The most commonly used DP definition is its constructive definition (Sethuraman and Tiwari, 1982; Sethuraman, 1994), which characterizes DP realizations as countable mixtures of point masses (and thus as random discrete distributions). Specifically, a random distribution G generated from $\mathrm{DP}(\alpha, G_0)$ is (almost surely) of the form

$$G(\cdot) = \sum_{\ell=1}^{\infty} w_\ell \, \delta_{\vartheta_\ell}(\cdot)$$

where $\delta_y(\cdot)$ denotes a point mass at y. The locations of the point masses, ϑ_ℓ, are i.i.d. realizations from G_0; the corresponding weights, w_ℓ, arise from a *stick-breaking* mechanism based on i.i.d. draws $\{\zeta_k : k = 1, 2, ...\}$ from a $\mathrm{Beta}(1, \alpha)$ distribution. In particular, $w_1 = \zeta_1$, and, for each $\ell = 2, 3, ...$, $w_\ell = \zeta_\ell \prod_{k=1}^{\ell-1}(1 - \zeta_k)$. Moreover, the sequences $\{\vartheta_\ell, \ell = 1, 2, \dots\}$ and $\{\zeta_k : k = 1, 2, ...\}$ are independent.

To complete the DP mixture model specification in (3), we take inverse gamma distributions for G_{r0} with fixed shape parameters c_r and random scale parameters d_r, $r = 1, 2$, which are assigned gamma priors. (We use the parameterization of the inverse gamma distribution with mean $d_r/(c_r - 1)$, provided $c_r > 1$.) Gamma priors are also placed on the DP precision parameters α_r, $r = 1, 2$.

To break the mixture in (3), we introduce latent mixing parameters σ_{1i} and σ_{2i}, corresponding to y_i, and express the full Bayesian model in the following hierarchical form:

$$
\begin{aligned}
y_i \mid h_1, h_2, \sigma_{1i}, \sigma_{2i} &\overset{ind}{\sim} k_p(y_i - \textstyle\sum_{\ell=1}^{2} h_\ell(x_{i\ell}); \sigma_{1i}, \sigma_{2i}) \\
h_\ell(\cdot) \mid \tau_\ell^2, \phi_\ell &\overset{ind}{\sim} \mathrm{GP}(0, C_\ell(\tau_\ell^2, \phi_\ell)) \\
\sigma_{ri} \mid G_r &\overset{iid}{\sim} G_r \\
G_r \mid \alpha_r, d_r &\overset{ind}{\sim} \mathrm{DP}(\alpha_r, G_{r0}(d_r)),
\end{aligned}
$$

$$(4)$$

for $i = 1, ..., n$, $\ell = 1, 2$, and $r = 1, 2$, where $C_\ell(x, x') = \mathrm{Cov}(h_\ell(x), h_\ell(x') \mid \tau_\ell^2, \phi_\ell)$ denotes the covariance function in (2). Moreover, priors are placed on the GP and DP hyperparameters $\boldsymbol{\psi} = (\tau_1^2, \phi_1, \tau_2^2, \phi_2, \alpha_1, d_1, \alpha_2, d_2)$ (see Section 2.2 regarding the priors for the τ_ℓ^2, ϕ_ℓ).

To obtain posterior predictive inference, we work with the version of (4) that involves the normal finite dimensional distributions for the $h_\ell(x_{i\ell})$ induced by the GP priors, and the priors for the σ_{ri} induced by marginalizing the random distributions G_r over their DP priors (Blackwell and MacQueen, 1973). Let $\delta_{i1} = h_1(x_{i1})$ and

$\delta_{i2} = h_2(x_{i2})$, for $i = 1, ..., n$. Then, the nonparametric model (4) induces the following hierarchical model that involves only finite dimensional parameters:

$$
\begin{aligned}
y_i \mid \delta_{i1}, \delta_{i2}, \sigma_{1i}, \sigma_{2i} &\overset{ind}{\sim} k_p(y_i - \delta_{i1} - \delta_{i2}; \sigma_{1i}, \sigma_{2i}) \\
(\delta_{1\ell}, ..., \delta_{n\ell}) \mid \tau_\ell^2, \phi_\ell &\overset{ind}{\sim} \mathrm{N}_n(\mathbf{0}, S_\ell(\tau_\ell^2, \phi_\ell)) \\
(\sigma_{r1}, ..., \sigma_{rn}) \mid \alpha_r, d_r &\overset{ind}{\sim} p(\sigma_{r1}, ..., \sigma_{rn} \mid \alpha_r, d_r),
\end{aligned}
$$

$$(5)$$

for $i = 1, ..., n$, $\ell = 1, 2$, and $r = 1, 2$, again, with priors for the components of $\boldsymbol{\psi}$. Here, the covariance matrix $S_\ell(\tau_\ell^2, \phi_\ell)$ is induced by the covariance function $C_\ell(\tau_\ell^2, \phi_\ell)$, i.e., its (i, j)-th element is given by $\mathrm{Cov}(h_\ell(x_{i\ell}), h_\ell(x_{j\ell}) \mid \tau_\ell^2, \phi_\ell) = \tau_\ell^2 \exp(-\phi_\ell |x_{i\ell} - x_{j\ell}|^{a_\ell})$. Moreover, the priors $p(\sigma_{r1}, ..., \sigma_{rn} \mid \alpha_r, d_r)$ are constructed based on the standard DP Polya urn scheme. Specifically, for $r = 1, 2$, $p(\sigma_{r1}, ..., \sigma_{rn} \mid \alpha_r, d_r)$ can be represented in terms of successive complete conditionals, with $\sigma_{r1} \sim G_{r0}$, and for each $i = 2, ..., n$, $p(\sigma_{ri} \mid \sigma_{r1}, ..., \sigma_{r,i-1}, \alpha_r, d_r)$ given by a mixed distribution with point masses $(\alpha_r + i - 1)^{-1}$ at the σ_{rj}, for $j = 1, ..., i - 1$, and continuous mass $\alpha_r(\alpha_r + i - 1)^{-1}$ on $G_{r0}(d_r)$.

2.2 MCMC posterior simulation

We describe here an MCMC algorithm to obtain posterior samples from $p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid \text{data})$, where $\boldsymbol{\delta}_\ell = (\delta_{1\ell}, ..., \delta_{n\ell})$, $\ell = 1, 2$, $\boldsymbol{\sigma}_r = (\sigma_{r1}, ..., \sigma_{rn})$, $r = 1, 2$, and data $= \{(y_i, x_{i1}, x_{i2}) : i = 1, ..., n\}$. These posterior samples can then be used to estimate posterior predictive distributions as is also discussed in this section.

Updating the σ_{ri}, conditionally on the $\delta_{i\ell}$ and all other relevant parameters, proceeds exactly as in the semiparametric model developed in Kottas and Krnjajić (2005), the only difference being that $\boldsymbol{x}_i^T \boldsymbol{\beta}$ from the semiparametric model is replaced with $h_1(x_{i1}) + h_2(x_{i2})$. The approach is also the same for updating the α_r and d_r.

For each $i = 1, ..., n$, the posterior full conditional for δ_{i1} is given by

$$p(\delta_{i1} \mid ..., \text{data}) \propto \mathrm{N}(\delta_{i1}; m_1, V_1) k_p(u_i - \delta_{i1}; \sigma_{1i}, \sigma_{2i})$$

where $u_i = y_i - \delta_{i2}$, and $\mathrm{N}(\delta_{i1}; m_1, V_1)$ is the prior full conditional for δ_{i1} corresponding to the n-variate normal prior for $\boldsymbol{\delta}_1$. Hence, letting $\boldsymbol{\delta}_{1(-i)} = \{\delta_{i'1} : i' \neq i\}$, we have $m_1 = \boldsymbol{b}B^{-1}\boldsymbol{\delta}_{1(-i)}$ and $V_1 = \tau_1^2 - \boldsymbol{b}B^{-1}\boldsymbol{b}^T$. Here, $\boldsymbol{b}$ is the $1 \times (n-1)$ vector with elements $\tau_1^2 \exp(-\phi_1 |x_{i1} - x_{i'1}|^{a_1})$ for $i' \neq i$, and B is the $(n-1) \times (n-1)$ matrix with elements $\tau_1^2 \exp(-\phi_1 |x_{j1} - x_{k1}|^{a_1})$ for (j, k) with $j \neq i$ and $k \neq i$. Now, $p(\delta_{i1} \mid ..., \text{data}) \propto \mathrm{N}(\delta_{i1}; m_1, V_1) \{(1 - p)\sigma_{2i}^{-1} 1_{(u_i - \sigma_{2i}, u_i]}(\delta_{i1}) + p\sigma_{1i}^{-1} 1_{(u_i, u_i + \sigma_{1i})}(\delta_{i1})\}$, and therefore letting $E_1 = \int_{u_i - \sigma_{2i}}^{u_i} \mathrm{N}(\delta; m_1, V_1) d\delta$ and $E_2 = \int_{u_i}^{u_i + \sigma_{1i}} \mathrm{N}(\delta; m_1, V_1) d\delta$, we obtain that $p(\delta_{i1} \mid ..., \text{data})$ is a two-component *mixture* of truncated normal distributions, $\mathrm{TN}(m_1, V_1; \delta_{i1} \in (u_i - \sigma_{2i}, u_i])$ and $\mathrm{TN}(m_1, V_1; \delta_{i1} \in (u_i, u_i + \sigma_{1i}))$ with weight, associated with the first component, given by $q =$
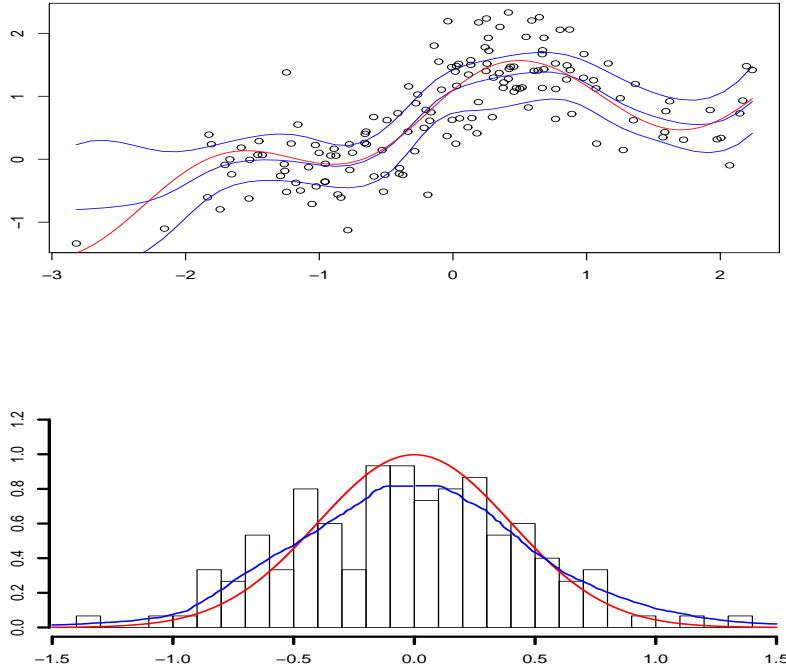
Figure 1: Median regression simulation example with normal errors. The blue lines denote point and interval estimates for the regression function (top panel) and the posterior predictive error density (bottom panel).

$((1-p)\sigma_{2i}^{-1}E_1)/((1-p)\sigma_{2i}^{-1}E_1 + p\sigma_{1i}^{-1}E_2)$. Note that although this is not a standard mixture, in that the components have different support, it can be easily sampled, since it only requires a truncated normal random variate.

For each $i = 1, ..., n$, the posterior full conditional for δ_{i2}, $p(\delta_{i2} \mid ..., \text{data}) \propto \text{N}(\delta_{i2}; m_2, V_2)k_p(v_i - \delta_{i2}; \sigma_{1i}, \sigma_{2i})$, where $v_i = y_i - \delta_{i1}$, and $\text{N}(\delta_{i2}; m_2, V_2)$ is the prior full conditional for δ_{i2} resulting from the $\text{N}_n(\mathbf{0}, S_2(\tau_2^2, \phi_2))$ prior for $\boldsymbol{\delta}_2$. Hence, the approach to update the δ_{i2} is similar to the one above for the δ_{i1}.

For $\ell = 1, 2$, assume an inverse gamma prior for τ_ℓ^2 with shape parameter a_{τ_ℓ} and scale parameter b_{τ_ℓ}. Moreover, write $S_\ell(\tau_\ell^2, \phi_\ell) = \tau_\ell^2 R(\phi_\ell)$, where $R(\phi_\ell)$ is the correlation matrix with (i,j)-th element $\exp(-\phi_\ell |x_{i\ell} - x_{j\ell}|^{a_\ell})$. Then, the posterior full conditional for τ_ℓ^2 is an inverse gamma distribution with shape parameter $a_{\tau_\ell} + 0.5n$ and scale parameter $b_{\tau_\ell} + 0.5\boldsymbol{\delta}_\ell^T R^{-1}(\phi_\ell)\boldsymbol{\delta}_\ell$.

Finally, for $\ell = 1, 2$, the posterior full conditional for ϕ_ℓ is proportional to

$$p(\phi_\ell)|R(\phi_\ell)|^{-1/2} \exp\{-(\boldsymbol{\delta}_\ell^T R^{-1}(\phi_\ell)\boldsymbol{\delta}_\ell)/(2\tau_\ell^2)\}$$

where $p(\phi_\ell)$ is the prior for ϕ_ℓ, $\ell = 1, 2$, which is taken to be uniform on $(0, b_\phi)$. Evidently, this form does not lead to an expression that can be sampled directly. Metropolis-Hastings steps can be used to update the ϕ_ℓ. Alternatively, the full conditional can be discretized and sampled directly as a discrete distribution (in which case,

the prior for ϕ_ℓ is discrete uniform on $(0, b_\phi)$). This latter approach is, in general, more efficient computationally, since the computations involving matrix $R(\phi_\ell)$ can be performed before the beginning of the MCMC steps. Regardless, because the data typically can not inform stronly about this GP hyperparameter, we need to specify the priors for ϕ_ℓ with some care. The interpretation of the range parameter ϕ_ℓ can be used to specify a plausible upper bound b_ϕ for its uniform prior. For instance, under exponential covariance functions (i.e., $a_\ell = 1$ in (2)), $3/\phi_\ell$ is the *range of dependence* (the value of the distance between covariate values that yields correlation about 0.05). If z_ℓ is the range of the observed $x_{i\ell}$, we can use, say, $z_\ell/2$ as a guess at the range of dependence, and thus set accordingly the prior upper bound for the ϕ_ℓ.

To obtain inference for the quantile regression functions, of direct interest is the posterior predictive distribution for $h_\ell(x_{0\ell})$, $\ell = 1, 2$, for any new covariate values x_{01} and x_{02}. Of interest is also the posterior predictive error density and the posterior predictive distribution for a new response observation y_0. Denote by $(\sigma_{10}, \sigma_{20})$ the new latent mixing parameters corresponding to y_0, and also let $\delta_{0\ell} = h_\ell(x_{0\ell})$, for $\ell = 1, 2$.

Adding $(\sigma_{10}, \sigma_{20})$, δ_{01}, δ_{02} and y_0 to model (4), and marginalizing the G_r over their DP priors, the corresponding joint posterior $p(y_0, \sigma_{10}, \sigma_{20}, \delta_{01}, \delta_{02}, \boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid x_{01}, x_{02}, \text{data})$ is
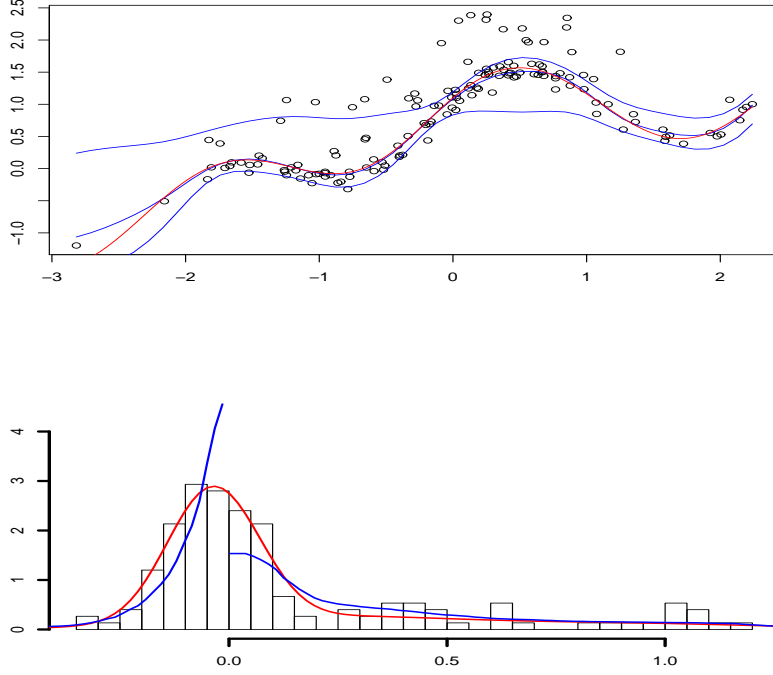
Figure 2: Median regression simulation example with skewed errors. The blue lines denote posterior point and interval estimates for the regression function (top panel) and the posterior predictive error density (bottom panel).

given by

$$
\begin{aligned}
& k_p(y_0 - \delta_{01} - \delta_{02}; \sigma_{10}, \sigma_{20}) \\
& \times p(\delta_{01} \mid \boldsymbol{\delta}_1, \tau_1^2, \phi_1) p(\delta_{02} \mid \boldsymbol{\delta}_2, \tau_2^2, \phi_2) \\
& \times p(\sigma_{10} \mid \boldsymbol{\sigma}_1, \alpha_1, d_1) p(\sigma_{20} \mid \boldsymbol{\sigma}_2, \alpha_2, d_2) \\
& \times p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid \text{data}).
\end{aligned}
$$

Here, $p(\delta_{0\ell} \mid \boldsymbol{\delta}_\ell, \tau_\ell^2, \phi_\ell)$ are the conditional normal distributions that result from the $(n+1)$-variate normal distributions for $(\delta_{0\ell}, \boldsymbol{\delta}_\ell)$ induced by the GP priors, for $\ell = 1, 2$. Hence, the mean of $p(\delta_{0\ell} \mid \boldsymbol{\delta}_\ell, \tau_\ell^2, \phi_\ell)$ is given by $\boldsymbol{b}_{0\ell} S_\ell^{-1}(\tau_\ell^2, \phi_\ell) \boldsymbol{\delta}_\ell$ and the variance by $\tau_\ell^2 - \boldsymbol{b}_{0\ell} S_\ell^{-1}(\tau_\ell^2, \phi_\ell) \boldsymbol{b}_{0\ell}^T$, where $\boldsymbol{b}_{0\ell}$ is the $1 \times n$ vector with elements $\tau_\ell^2 \exp(-\phi_\ell |x_{0\ell} - x_{i\ell}|^{a_\ell})$, for $i = 1, ..., n$. Moreover, $p(\sigma_{r0} \mid \boldsymbol{\sigma}_r, \alpha_r, d_r)$ arises from the DP Polya urn structure, i.e., for $r = 1, 2$, $p(\sigma_{r0} \mid \boldsymbol{\sigma}_r, \alpha_r, d_r)$ can be expressed as

$$
\frac{\alpha_r}{\alpha_r + n} G_{r0}(\sigma_{r0} \mid d_r) + \frac{1}{\alpha_r + n} \sum_{j=1}^{n_r^*} n_{rj} \delta_{\sigma_{rj}^*}(\sigma_{r0}),
$$

where σ_{rj}^*, $j = 1, ..., n_r^*$, are the distinct values in the vector $(\sigma_{r1}, ..., \sigma_{rn})$ and n_{rj} is the size of the j-th cluster.

Now, the posterior predictive density for y_0, $p(y_0 \mid$

$x_{01}, x_{02}, \text{data})$, is given by

$$
\begin{aligned}
& \iint \iint \int k_p(y_0 - \delta_{01} - \delta_{02}; \sigma_{10}, \sigma_{20}) p(\delta_{01} \mid \boldsymbol{\delta}_1, \tau_1^2, \phi_1) \\
& p(\delta_{02} \mid \boldsymbol{\delta}_2, \tau_2^2, \phi_2) p(\sigma_{10} \mid \boldsymbol{\sigma}_1, \alpha_1, d_1) p(\sigma_{20} \mid \boldsymbol{\sigma}_2, \alpha_2, d_2) \\
& p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid \text{data}) \\
& d\delta_{01} d\delta_{02} d\sigma_{10} d\sigma_{20} d\boldsymbol{\delta}_1 d\boldsymbol{\delta}_2 d\boldsymbol{\sigma}_1 d\boldsymbol{\sigma}_2 d\boldsymbol{\psi}
\end{aligned}
$$

and the posterior predictive error density,

$$
\begin{aligned}
p(\epsilon_0 \mid \text{data}) = & \iint \int k_p(\epsilon_0; \sigma_{10}, \sigma_{20}) p(\sigma_{10} \mid \boldsymbol{\sigma}_1, \alpha_1, d_1) \\
& p(\sigma_{20} \mid \boldsymbol{\sigma}_2, \alpha_2, d_2) p(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid \text{data}) \\
& d\sigma_{10} d\sigma_{20} d\boldsymbol{\sigma}_1 d\boldsymbol{\sigma}_2 d\boldsymbol{\psi}.
\end{aligned}
$$

Hence, both can be easily sampled using the expressions for $p(\delta_{0\ell} \mid \boldsymbol{\delta}_\ell, \tau_\ell^2, \phi_\ell)$ and $p(\sigma_{r0} \mid \boldsymbol{\sigma}_r, \alpha_r, d_r)$ and the posterior samples from $p(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \boldsymbol{\psi} \mid \text{data})$.

Finally, for $\ell = 1, 2$, the posterior predictive distribution for $\delta_{0\ell}$,

$$
\begin{aligned}
p(\delta_{0\ell} \mid x_{0\ell}, \text{data}) = & \iint \int p(\delta_{0\ell} \mid \boldsymbol{\delta}_\ell, \tau_\ell^2, \phi_\ell) \\
& p(\boldsymbol{\delta}_\ell, \tau_\ell^2, \phi_\ell \mid \text{data}) d\tau_\ell^2 d\phi_\ell d\boldsymbol{\delta}_\ell.
\end{aligned}
$$

As illustrated in Section 2.3, samples from $p(\delta_{0\ell} \mid x_{0\ell}, \text{data})$ over a grid of points $x_{0\ell}$ yield (pointwise) point and interval estimates for the quantile regression functions $h_\ell(\cdot)$, $\ell = 1, 2$.

2.3 Simulation study

Here, we consider simulated data examples to study the performance of the model developed in Section 2.1, fo-
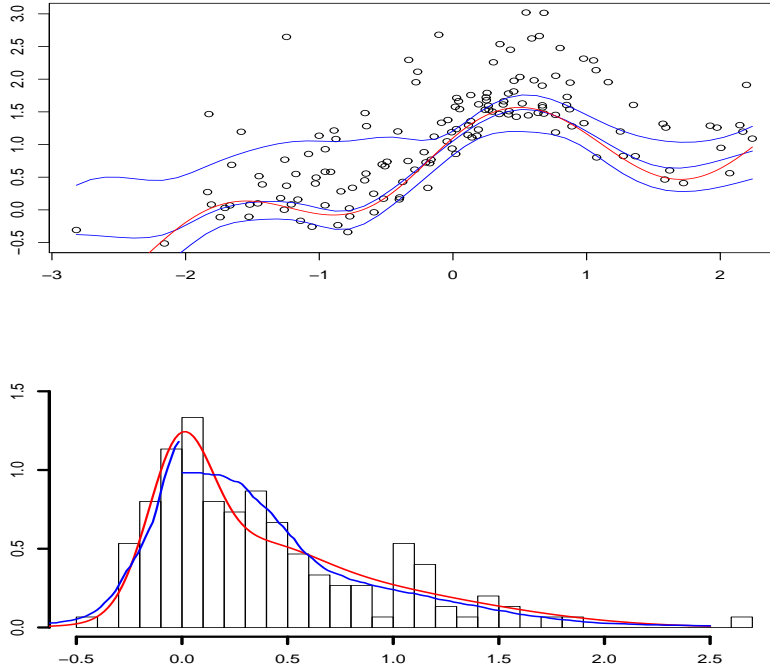
Figure 3: Simulation example for 25-th quantile regression. The blue lines denote posterior point and interval estimates for the regression function (top panel) and the posterior predictive error density (bottom panel).

cusing on the case with a single covariate, x. In all examples, we take $n = 150$, and work with the non-linear regression function used by Neal (1997) in the context of mean regression with GP priors and either normal or t error distributions. Specifically, the true quantile regression function for all simulated data sets is given by

$$h(x) = 0.4x + 0.5\sin(2.7x) + 1.1(1 + x^2)^{-1}.$$

The covariate values are generated from a standard normal distribution, while response values y are drawn by adding errors to $h(x)$ based on four different choices for the error distribution, resulting in four simulated data sets. In each case, the simulated response and covariate values are plotted in the top panel of Figures $1 - 4$, where also plotted is the true regression function (denoted by the red line). The bottom panels of Figures $1 - 4$ include histograms of the simulated errors as well as the corresponding true error density (again, denoted by the red line).

The first data set (Figure 1) involves normal errors in the median regression setting. The other three data sets are based on errors generated from mixtures of normal distributions chosen to yield skewed error densities with a specific quantile (and mode) equal to 0. In particular, the error distribution in the second data example (Figure 2) is right skewed; here, again, we focus on the median regression case. In the remaining two data examples, we

consider the cases of 25-th quantile regression (Figure 3) and 75-th quantile regression (Figure 4), with right skewed and left skewed error distributions, respectively.

We apply the single-covariate version of model (4) to the simulated data. For all four data examples, we take an exponential covariance function for the GP prior, i.e., the version of (2) given by $\tau^2 \exp(-\phi|x - x'|)$, with a discrete uniform prior on $(0, 5)$ for ϕ, and an inverse gamma prior for τ^2 with mean 1 (and shape parameter 2). Regarding the DP prior specification, we take $c_r = 2$ for the shape parameters of the inverse gamma base distributions G_{r0}, $r = 1, 2$, and specify the mean of the gamma priors for the scale parameters d_r using a rough range for the response values. Sensitivity analysis with regard to the dispersion of the prior distributions for the d_r revealed robustness for the resulting posterior predictive inference. Finally, a gamma$(4, 0.5)$ prior (with mean 8) was used for α_1 and α_2.

For each of the data examples, the top panels of Figures $1 - 4$ plot pointwise posterior mean estimates and 95% posterior interval estimates for the respective quantile regression functions. The bottom panels of the figures show the corresponding posterior predictive error densities. In all cases, the model captures quite well the shape of both the non-linear regression function and the error density. Particularly encouraging for the potential of the methodology is the fact that the posterior predictive esti-
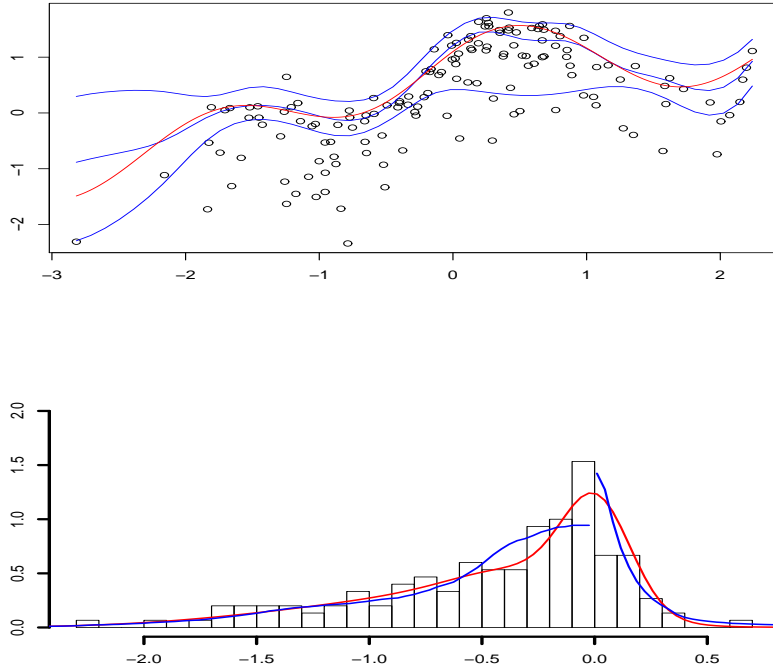
Figure 4: Simulation example for 75-th quantile regression. The blue lines denote posterior point and interval estimates for the regression function (top panel) and the posterior predictive error density (bottom panel).

mates identify successfully the two random functions (the quantile regression function and the error density function) under different types and amounts of skewness in the true error distributions.

3. A fully nonparametric approach to inference for quantile regression

The starting point for the approach of Section 2 (and for most existing approaches to quantile regression) is the standard additive regression framework $y = h(\boldsymbol{x}) + \epsilon$, where again the errors ϵ are assumed independent from a distribution with p-th percentile equal to 0. Note that, under this framework (and regardless of the formulation for the regression function), if inference is sought for more than one quantile regression, the particular model needs to be fitted separately for each corresponding p. In particular, note that estimated quantile regression functions for nearby values of p might not satisfy the explicit ordering of the corresponding percentiles, especially with small sample sizes and/or for extreme percentiles. And this attribute of the additive formulation is shared by any approach that utilizes a (proper) probability model for the error distribution, regardless of the estimation method (likelihood or Bayesian). There are certain approaches that allow, in the context of this framework, *simultaneous* estimation for more than one quantile regression

(e.g., Dunson & Taylor, 2005); however, this is only possible because they do not involve modeling for the errors, but are rather based on approximate methods (e.g., certain pseudo-likelihoods). Hence, the additive quantile regression framework is more suitable for applications where interest lies in explaining one percentile (or a few well-separated percentiles) of the response distribution in terms of available covariates. For such settings, by separating the quantile regression function from the errors, the model formulation in (1) allows readily interpretable inference, incorporation of different types of covariates, and extensions of the inference methods to handle censored observations.

In this section, we develop an alternative approach to inference for quantile regression. This approach does not build on a structured regression model formulation as the one in Section 2, but it yields flexible, fully nonparametric inference for quantile regression.

3.1 The modeling approach

The starting point for this approach is to consider a model for the joint distribution of the response, y, and the set of covariates, $\boldsymbol{x} = (x_1, ..., x_L)$ (again, we focus here on covariate information on continuous variables). Then, inference for any set of quantile curves can be obtained based on the posterior of the implied conditional response

distribution given the covariates. Clearly, the richness of the resulting inference relies on the flexibility of the prior probability model for the distribution of $\boldsymbol{z} = (y, \boldsymbol{x})$. We employ a DP mixture of multivariate normals, which models the joint density for $\boldsymbol{z}$ through

$$f(\boldsymbol{z}; G) = \int N_{L+1}(\boldsymbol{z}; \boldsymbol{\mu}, \Sigma) dG(\boldsymbol{\mu}, \Sigma), \; G \sim DP(\alpha, G_0) \tag{6}$$

with G_0 built from independent $N_{L+1}(\boldsymbol{m}, V)$ and IWish(ν, S) components for $\boldsymbol{\mu}$ and Σ, respectively. We work with random $\boldsymbol{m}$, V and S and fixed ν. Here, IWish(ν, S) denotes the inverse Wishart distribution for the $(L+1) \times (L+1)$ (positive definite) matrix Σ with density proportional to $|\Sigma|^{-(\nu+L+2)/2} \exp\{-0.5 \text{tr}(S\Sigma^{-1})\}$.

This model has been studied in the context of multivariate density estimation as well as curve fitting (see, e.g., Müller, Erkanli and West, 1996). However, the scope of inference has been limited to posterior point estimates, obtained through posterior predictive distributions, i.e., $p(\boldsymbol{z}_0|\text{data}) \equiv E(f(\boldsymbol{z}_0; G)|\text{data})$, where, again, data comprises $\{\boldsymbol{z}_i = (y_i, \boldsymbol{x}_i) : i = 1, ..., n\}$. Our application to quantile regression requires the entire posterior of $f(\boldsymbol{z}_0; G)$ at any $\boldsymbol{z}_0$, and we thus employ a more general approach to MCMC inference (discussed in Section 3.2) that includes sampling from the posterior of G.

The hierarchical model formulation involves, again, latent mixing parameters associated with each vector of response/covariate observations,

$$\begin{aligned}
\boldsymbol{z}_i|\boldsymbol{\mu}_i, \Sigma_i &\overset{ind}{\sim} N_{L+1}(\boldsymbol{z}_i; \boldsymbol{\mu}_i, \Sigma_i) \\
(\boldsymbol{\mu}_i, \Sigma_i)|G &\overset{iid}{\sim} G \\
G|\alpha, \boldsymbol{\psi} &\sim DP(\alpha, G_0(\boldsymbol{\psi}))
\end{aligned} \tag{7}$$

for $i = 1, ..., n$, with hyperpriors for the DP precision parameter α and for the parameters, $\boldsymbol{\psi} = (\boldsymbol{m}, V, S)$, of G_0. In particular, we use a gamma prior for α, a $N_{L+1}(a_{\boldsymbol{m}}, B_{\boldsymbol{m}})$ prior for $\boldsymbol{m}$, an IWish(a_V, B_V) prior for V, and a Wish(a_S, B_S) prior for the $(L+1) \times (L+1)$ positive definite matrix S with density proportional to $|S|^{(a_S - L - 2)/2} \exp\{-0.5 \text{tr}(SB_S^{-1})\}$ (provided $a_S \geq L+1$).

3.2 Posterior inference for quantile regression

We describe here the approach to estimate quantile curves based on the posterior for the conditional density $f(y|\boldsymbol{x}; G)$ implied by DP mixture model (6).

The first step involves MCMC sampling from the posterior of model (7) with G marginalized over its DP prior. We update each $(\boldsymbol{\mu}_i, \Sigma_i)$ using algorithm 5 from Neal (2000), which is based on Metropolis-Hastings steps with the prior full conditional of $(\boldsymbol{\mu}_i, \Sigma_i)$ as the proposal. Updating all the $(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1, ..., n$, generates a partition with $n^* (\leq n)$ distinct components $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, $j = 1, ..., n^*$. The vector of the $(\boldsymbol{\mu}_i, \Sigma_i)$ is determined through configuration indicators $\boldsymbol{w} = (w_1, ..., w_n)$ such that $w_i = j$ if and only if $(\boldsymbol{\mu}_i, \Sigma_i) = (\boldsymbol{\mu}_j^*, \Sigma_j^*)$. Evidently, the Metropolis-Hastings approach to update the

$(\boldsymbol{\mu}_i, \Sigma_i)$ can lead to poor mixing. However, it is straightforward to implement and, combined with the standard trick from Bush and MacEachern (1996) to resample the $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, yields an efficient MCMC method. For each $j = 1, ..., n^*$, the posterior full conditional for $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$ is proportional to $g_0(\boldsymbol{\mu}_j^*, \Sigma_j^*; \boldsymbol{\psi}) \prod_{\{i:w_i=j\}} N_{L+1}(\boldsymbol{z}_i; \boldsymbol{\mu}_j^*, \Sigma_j^*)$, where g_0 denotes the density of G_0. We sample from this full conditional by drawing from the full conditionals for $\boldsymbol{\mu}_j^*$ and Σ_j^*. The former is $(L+1)$-variate normal with mean vector $(V^{-1} + n_j \Sigma_j^{*-1})^{-1}(V^{-1}\boldsymbol{m} + n_j \Sigma_j^{*-1}\tilde{\boldsymbol{z}}_j)$ and covariance matrix $(V^{-1} + n_j \Sigma_j^{*-1})^{-1}$, where $n_j = |\{i : w_i = j\}|$ and $\tilde{\boldsymbol{z}}_j = n_j^{-1} \sum_{\{i:w_i=j\}} \boldsymbol{z}_i$. The latter is inverse Wishart with scalar parameter $\nu + n_j$ and matrix parameter $S + \sum_{\{i:w_i=j\}} (\boldsymbol{z}_i - \boldsymbol{\mu}_j^*)(\boldsymbol{z}_i - \boldsymbol{\mu}_j^*)^T$.

Regarding the DP hyperparameters, we update α using the auxiliary variable method from Escobar and West (1995). The posterior full conditional for $\boldsymbol{m}$ is $(L+1)$-variate normal with mean vector $(B_{\boldsymbol{m}}^{-1} + n^* V^{-1})^{-1}(B_{\boldsymbol{m}}^{-1} a_{\boldsymbol{m}} + n^* V^{-1} \tilde{\boldsymbol{\mu}}^*)$, with $\tilde{\boldsymbol{\mu}}^* = n^{*-1} \sum_{j=1}^{n^*} \boldsymbol{\mu}_j^*$, and covariance matrix $(B_{\boldsymbol{m}}^{-1} + n^* V^{-1})^{-1}$. The full conditional for V is inverse Wishart with scalar parameter $a_V + n^*$ and matrix parameter $B_V + \sum_{j=1}^{n^*} (\boldsymbol{\mu}_j^* - \boldsymbol{m})(\boldsymbol{\mu}_j^* - \boldsymbol{m})^T$. Finally, the full conditional for S is given by a Wishart distribution with scalar parameter $a_S + \nu n^*$ and matrix parameter $(B_S^{-1} + \sum_{j=1}^{n^*} \Sigma_j^{*-1})^{-1}$.

Next, note that the full posterior corresponding to model (7) is given by

$$p(G, \boldsymbol{\theta}, \alpha, \boldsymbol{\psi}|\text{data}) = p(G|\boldsymbol{\theta}, \alpha, \boldsymbol{\psi})p(\boldsymbol{\theta}, \alpha, \boldsymbol{\psi}|\text{data}) \tag{8}$$

where $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_i, \Sigma_i) : i = 1, ..., n\}$, or equivalently, $\boldsymbol{\theta} = (n^*, \{(\boldsymbol{\mu}_j^*, \Sigma_j^*) : j = 1, ..., n^*\}, \boldsymbol{w})$ (Antoniak, 1974). Here, the distribution for $G|\boldsymbol{\theta}, \alpha, \boldsymbol{\psi}$ corresponds to a DP with updated precision parameter $\alpha + n$ and mean $\tilde{G}_0(\cdot; \boldsymbol{\theta}, \alpha, \boldsymbol{\psi})$, which is a mixed distribution with point masses $n_j(\alpha + n)^{-1}$ at the $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, $j = 1, ..., n^*$, and continuous mass $\alpha(\alpha + n)^{-1}$ on $G_0(\boldsymbol{\psi})$.

Hence, after we obtain posterior samples from $p(\boldsymbol{\theta}, \alpha, \boldsymbol{\psi}|\text{data})$, using the MCMC algorithm described above, we can draw from (8) by augmenting each posterior sample with a draw from $p(G|\boldsymbol{\theta}, \alpha, \boldsymbol{\psi})$. The latter requires simulation from the DP with parameters given above, which we implement using the DP constructive definition (discussed in Section 2.1) with a truncation approximation (Gelfand and Kottas, 2002; Kottas, 2006). Therefore, this approach yields samples $\{G_b, \boldsymbol{\theta}_b, \alpha_b, \boldsymbol{\psi}_b : b = 1, ..., B\}$ from the full posterior (8). Each posterior realization G_b is a discrete distribution with point masses at $\boldsymbol{\vartheta}_{rb} = (\tilde{\boldsymbol{\mu}}_{rb}, \tilde{\Sigma}_{rb})$, $r = 1, ..., R_b$, drawn i.i.d. from $\tilde{G}_0(\cdot; \boldsymbol{\theta}, \alpha, \boldsymbol{\psi})$, and corresponding weights ω_{rb}, $r = 1, ..., R_b - 1$, generated using the stick-breaking construction based on i.i.d. Beta$(1, \alpha_b)$ draws (we set $\omega_{R_b b} = 1 - \sum_{r=1}^{R_b - 1} \omega_{rb}$). Here, R_b is the number of terms used in the truncation series approximation to the countable series representation for the DP. In general, R_b depends on the particular posterior realization, and the approximation can be specified up to any desired accuracy (see Kottas, 2006, for a specific rule to choose R_b).
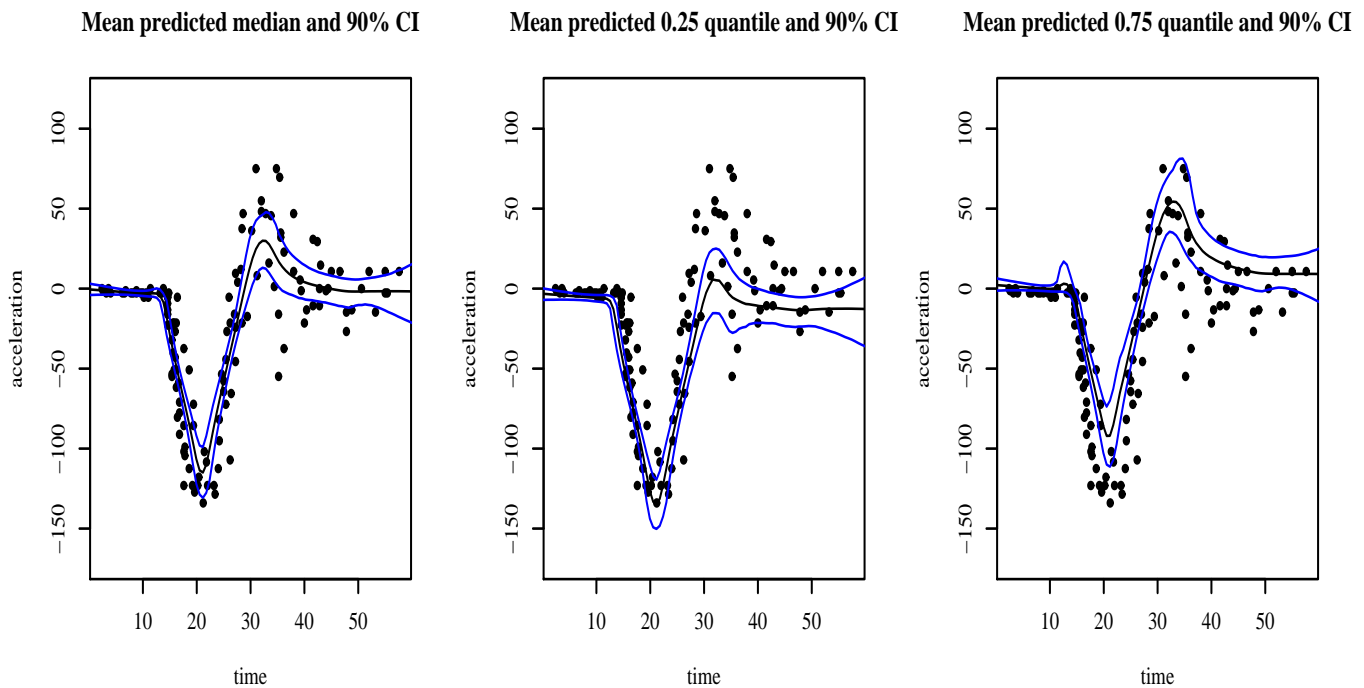
Figure 5: Posterior estimates for the median and quartiles for the motorcycle data. In each panel, the black line denotes the posterior mean and the blue lines contain a 90% posterior credible interval.

Now, for any specific combination of response and covariate values, say, $(y_0, \boldsymbol{x}_0)$,

$$
\begin{aligned}
f(y_0, \boldsymbol{x}_0; G_b) &= \int \mathrm{N}_{L+1}(y_0, \boldsymbol{x}_0; \boldsymbol{\mu}, \Sigma) \mathrm{d}G_b(\boldsymbol{\mu}, \Sigma) \\
&= \sum_{r=1}^{R_b} \omega_{rb} \mathrm{N}_{L+1}(y_0, \boldsymbol{x}_0; \tilde{\boldsymbol{\mu}}_{rb}, \tilde{\Sigma}_{rb})
\end{aligned}
$$

is a realization from the posterior of the random mixture density $f(y, \boldsymbol{x}; G)$, corresponding to the DP mixture model (6), at point $(y, \boldsymbol{x}) = (y_0, \boldsymbol{x}_0)$. Analogously, we can compute $f(\boldsymbol{x}_0; G_b)$ a draw from the posterior of the marginal $f(\boldsymbol{x}; G)$ at point $\boldsymbol{x} = \boldsymbol{x}_0$, and thus, also, $f(y_0 \mid \boldsymbol{x}_0; G_b) = f(y_0, \boldsymbol{x}_0; G_b)/f(\boldsymbol{x}_0; G_b)$, i.e., a realization from the posterior of the conditional density $f(y \mid \boldsymbol{x}; G)$, at point $(y, \boldsymbol{x}) = (y_0, \boldsymbol{x}_0)$. Repeating over a grid in y, that covers the range of response values of interest, we obtain a posterior realization from the random conditional density function $f(\cdot \mid \boldsymbol{x}_0; G)$ for the specific covariate values $\boldsymbol{x}_0$. Note that this is a posterior realization for the entire function, obtained, of course, up to the grid approximation. Now, for any $0 < p < 1$, the conditional quantile $q_p(\boldsymbol{x}_0) \equiv q_p(\boldsymbol{x}_0; G)$ satisfies $\int^{q_p(\boldsymbol{x}_0)} f(y \mid \boldsymbol{x}_0; G) \mathrm{d}y = p$. Hence, using numerical integration (with interpolation) of the posterior realizations from the conditional density $f(\cdot \mid \boldsymbol{x}_0; G)$, yields draws from the posterior of $q_p(\boldsymbol{x}_0)$ for any set of percentiles that might be of interest (i.e., for any set of $p \in (0, 1)$).

Therefore, for any specified $\boldsymbol{x}_0$, and for any $0 < p < 1$,

we obtain samples from $p(q_p(\boldsymbol{x}_0) \mid \text{data})$ that can be used to summarize the information from these conditional quantiles in any desired form. In particular, for any set of p values, working with a grid in $\boldsymbol{x}$, i.e., over the covariate space, we can compute point and interval estimates for the corresponding quantile curves $q_p(\cdot; G)$. Evidently, graphical depiction of these estimates for the entire curve is not feasible for problems with more than two covariates. For such applications, one can focus on illustrations involving the quantile regression function given subsets of the covariate vector including specific choices of one or two covariates.

Note that this approach to inference for quantile regression allows both non-linear quantile curves as well as non-standard shapes for the conditional distribution of the response given the covariates. As illustrated in Section 3.3, the latter can be explored through the posterior of $f(\cdot \mid \boldsymbol{x}_0; G)$ for any set of covariate values $\boldsymbol{x}_0$ of interest. Moreover, the model does not rely on the additive nonparametric regression formulation and therefore can uncover interactions between covariates that might influence certain quantile regression curves. Finally, an important feature of the approach is that it enables simultaneous inference for any set of quantile regressions that might be of interest in a particular application.
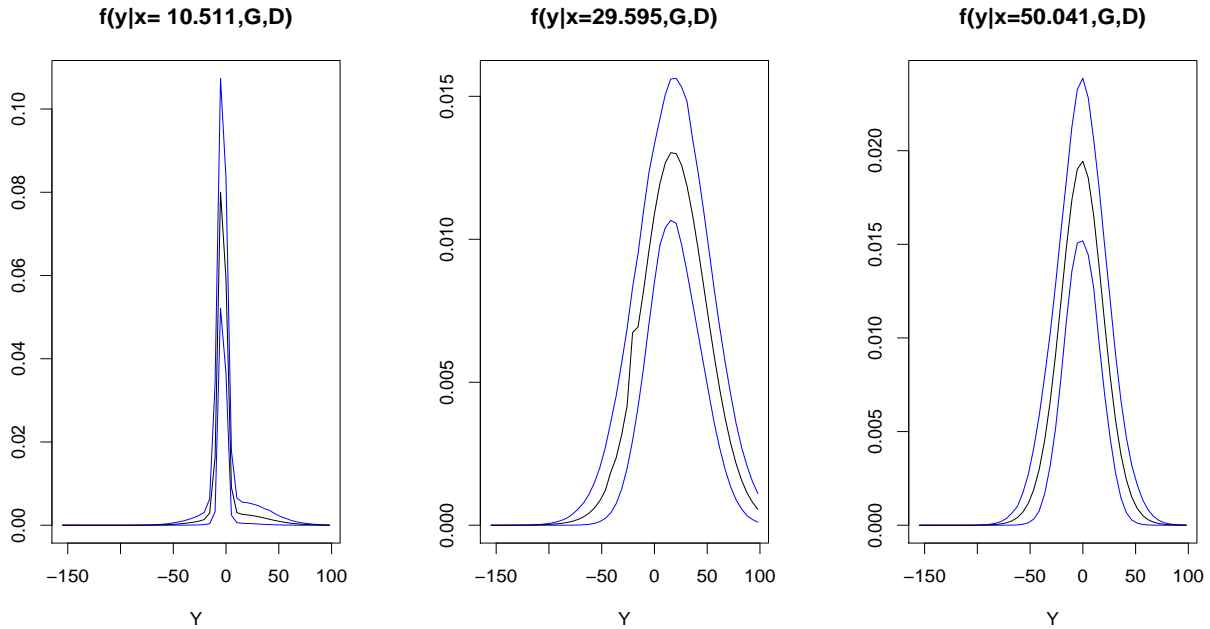
Figure 6: For the motorcycle data, posterior estimates of the conditional density $f(\cdot \mid x; G)$ at three values of x. In each panel, the black line denotes the posterior mean and the blue lines contain a 90% posterior interval.

3.3 Data example

We illustrate the methodology of Section 3.2 with the motorcycle data (see, e.g., Silverman, 1985), which consist of 133 measurements of velocity in time for the helmet of a motorcycle crash victim after impact. (The data set is available from the MASS package for R.) Note that model (6) incorporates uncertainty in the covariates, which is arguably not required in this application. However, this example is still useful for illustrative purposes, since the motorcycle data is used as one of the standard tests for nonparametric regression approaches, including quantile regression estimation methods (see, e.g., Koenker, 2005, chapter 7).

We used a rough center and range for the response and covariate values to specify the hyperparameters for G_0. In particular, we choose the prior mean for $\boldsymbol{m}$ using the center from the data. Setting $\nu = a_V = a_S = 6$, and working with diagonal matrices for $B_{\boldsymbol{m}}$, B_V, B_S, we specify their diagonal elements using a rough range for acceleration and time values. Posterior inference was robust to less (as well as more) informative choices, e.g., involving range values larger (and smaller) than the data ranges. Finally, an exponential prior with mean 5 was placed on α.

Figure 5 plots point (posterior mean) and 90% interval estimates for three quantile curves, specifically, for the median regression, 25-th quantile regression, and 75-th quantile regression functions. All posterior estimates capture very well the shape suggested by the data, in par-

ticular, adapting to its heteroscedasticity, the key feature of the motorcycle data. This is also reflected in posterior inference for $f(\cdot \mid x; G)$ for different values x of time. For instance, Figure 6 shows posterior mean estimates and 90% posterior interval estimates for this conditional density at three values of x.

4. Discussion

We have developed two different model-based, fully inferential approaches for quantile regression. First, we have discussed a practically important extension of the semiparametric model from Kottas and Krnjajić (2005), combining Gaussian process priors for the regression functions and a Dirichlet process mixture prior for the errors in an additive quantile regression framework. Under the second approach, the joint distribution of the response and the (continuous) covariates is assigned a general Dirichlet process mixture prior model with inference for quantile regressions emerging from the induced conditional posterior distribution of the response given the covariates. We have presented MCMC posterior simulation methods under both approaches. Finally, we have provided illustrations for both models using either simulated or real data examples.

The MCMC algorithms for both models can be readily extended to handle censored response observations that arise, for instance, in survival analysis applications. For such applications, more structured versions for the kernel of the nonparametric mixture model of Section 3 can

be utilized. For example, the kernel can be built from a Weibull or gamma distribution for the response and, independently, a multivariate normal distribution for the covariates.

An alternative method for posterior simulation from the model of Section 3.1 involves direct approximation of G in model (7), using the constructive definition of its $DP(\alpha, G_0)$ prior, and then application of an MCMC technique for the induced discrete mixture model (see, e.g., Ishwaran and James, 2001). Results from comparison of this method with the approach of Section 3.2 will be reported elsewhere.

Regarding the approach of Section 2, we note that there has been relatively limited work in the Bayesian nonparametrics literature on combination of nonparametric priors for different sets of functions within the same model. Further study, involving either analytical work or empirical work through simulation, is needed for this approach regarding the extent of identifiability issues.

Finally, implementation of inference under both models becomes computationally intensive for moderate number of covariates, and seems prohibitive for large number of covariates with the currently used MCMC algorithms. It will therefore be of interest to study alternative to MCMC inferential methods (e.g., variational approximation techniques for Dirichlet process mixtures) and/or more efficient MCMC posterior simulation methods.

References

Antoniak, C.E. (1974), "Mixtures of Dirichlet processes with applications to nonparametric problems," *The Annals of Statistics*, **2**, 1152-1174.

Blackwell, D., and MacQueen, J.B. (1973), "Ferguson distributions via Pólya urn schemes." *The Annals of Statistics*, **1**, 353-355.

Brunner, L.J. (1995), " Bayesian linear regression with error terms that have symmetric unimodal densities." *Journal of Nonparametric Statistics*, **4**, 335-348.

Brunner, L.J., and Lo, A.Y. (1989), "Bayes methods for a symmetric unimodal density and its mode." *The Annals of Statistics*, **17**, 1550-1566.

Bush, C.A., and MacEachern, S.N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs." *Biometrika*, **83**, 275-285.

Dunson, D.B., and Taylor, J.A. (2005), "Approximate Bayesian inference for quantiles." *Journal of Nonparametric Statistics*, **17**, 385-400.

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

Ferguson, T.S. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, **1**, 209-230.

Gelfand, A.E., and Kottas, A. (2002), "A Computational Approach for Full Nonparametric Bayesian Inference under Dirichlet Process Mixture Models." *Journal of Computational and Graphical Statistics*, **11**, 289-305.

Hanson, T., and Johnson, W.O. (2002), "Modeling regression error with a mixture of Pólya trees." *Journal of the American Statistical Association*, **97**, 1020-1033.

He, X., Ng, P., and Portnoy, S. (1998), "Bivariate quantile smoothing splines." *Journal of the Royal Statistical Society, Series B*, **60**, 537-550.

Hjort, N.L., and Petrone, S. (2005), "Nonparametric quantile inference using Dirichlet processes." In *Festschrift for Kjell Doksum, IMS Lecture Notes Series* (eds D. Dabrowska & V. Nair).

Horowitz, J.L., and Lee, S. (2005), "Nonparametric estimation of an additive quantile regression model." *Journal of the American Statistical Association*, **100**, 1238-1249.

Ishwaran, H., and James, L.F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors." *Journal of the American Statistical Association*, **96**, 161-173.

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

Kottas, A. (2006), "Nonparametric Bayesian Survival Analysis using Mixtures of Weibull Distributions." *Journal of Statistical Planning and Inference*, **136**, 578-596.

Kottas, A., and Gelfand, A.E. (2001), "Bayesian semiparametric median regression modeling." *Journal of the American Statistical Association*, **96**, 1458-1468.

Kottas, A., and Krnjajić, M. (2005), "Bayesian nonparametric modeling in quantile regression." Technical Report AMS 2005-06, University of California, Santa Cruz.

Lavine, M., and Mockus, A. (1995), "A nonparametric Bayes method for isotonic regression." *Journal of Statistical Planning and Inference*, **46**, 235-248.

Müller, P., and Quintana, F.A. (2004), "Nonparametric Bayesian data analysis." *Statistical Science*, **19**, 95-110.

Müller, P., Erkanli, A., and West, M. (1996), "Bayesian curve fitting using multivariate normal mixtures." *Biometrika*, **83**, 67-79.

Neal, R.M. (1997), "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification." Technical Report No. 9702, Department of Statistics, University of Toronto.

Neal, R.M. (1998), "Regression and classification using Gaussian process priors." In *Bayesian statistics 6: Proceedings of the sixth Valencia international meeting* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid & A.F.M. Smith). Oxford University Press, pp. 475-501.

Neal, R.M. (2000), "Markov chain sampling methods for Dirichlet process mixture models." *Journal of Computational and Graphical Statistics*, **9**, 249-265.

Scaccia, L., and Green, P.J. (2003), "Bayesian growth curves using normal mixtures with nonparametric weights." *Journal of Computational and Graphical Statistics*, **12**, 308-331.

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

Sethuraman, J., and Tiwari, R.C. (1982), "Convergence of Dirichlet Measures and the Interpretation of their Parameter." In *Statistical Decision Theory and Related Topics III*, eds. S. Gupta and J.O. Berger, New York: Springer-Verlag, 2, pp. 305-315.

Silverman, B.W. (1985), "Some aspects of the spline smoothing approach to non-parametric curve fitting," *Journal of the Royal Statistical Society, Series B*, **47**, 1-52.

Tsionas, E.G. (2003), "Bayesian quantile inference." *Journal of Statistical Computation and Simulation*, **73**, 659-674.

Walker, S.G., and Mallick, B.K. (1999), "A Bayesian semiparametric accelerated failure time model." *Biometrics*, **55**, 477-483.

Yu, K. (2002), "Quantile regression using RJMCMC algorithm." *Computational Statistics & Data Analysis*, **40**, 303-315.

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

Yu, K., Lu, Z., and Stander, J. (2003), "Quantile regression: applications and current research areas." *The Statistician*, **52**, 331-350.