Stat Comput (2015) 25:651–668
DOI 10.1007/s11222-013-9446-9

Bayesian quantile regression for partially linear additive models

Yuao Hu · Kaifeng Zhao · Heng Lian

Received: 9 July 2013 / Accepted: 11 December 2013 / Published online: 11 March 2014

Abstract In this article, we develop a semiparametric Bayesian estimation and model selection approach for partially linear additive models in conditional quantile regression. The asymmetric Laplace distribution provides a mechanism for Bayesian inferences of quantile regression models based on the check loss. The advantage of this new method is that nonlinear, linear and zero function components can be separated automatically and simultaneously during model fitting without the need of pre-specification or parameter tuning. This is achieved by spike-and-slab priors using two sets of indicator variables. For posterior inferences, we design an effective partially collapsed Gibbs sampler. Simulation studies are used to illustrate our algorithm. The proposed approach is further illustrated by applications to two real data sets.

Keywords Additive models · Markov chain Monte Carlo · Quantile regression · Variable selection

1 Introduction

Partially linear additive models (PLAMs) generalize multiple linear regression models. They can also be regarded as a special case of generalized additive regression models. PLAMs, containing both linear and nonlinear additive components, are more flexible than stringent linear models, and are more parsimonious than general nonparametric regression models and they circumvent the difficulty brought by the problem known as "curse of dimensionality". PLAMs have been widely applied in practice because of these advantages. For example, Liang et al. (2008) applied PLAMs to study the relationship between environmental chemical exposures and semen quality, and Panagiotelis and Smith (2008) applied the models for intra-day electricity load analysis.

In this article, we propose a Bayesian quantile regression approach for partially linear additive models. At a given quantile level $\tau \in (0, 1)$, a partially linear additive model has the following form,

$$y_i = \mu_\tau + \sum_{j=1}^{p} f_{\tau,j}(x_{ij}) + \epsilon_{\tau,i}, \quad i = 1, \ldots, n, \tag{1}$$

where $(y_i, \boldsymbol{x}_i)$ are independent and identically distributed pairs. Here y_i is the response, $\boldsymbol{x_i} = (x_{i1}, \ldots, x_{ip})^T$ is the p-dimensional predictor, μ_τ is the intercept, $\epsilon_i, i = 1, \ldots, n$, are random errors with their τth quantile equal to 0 and $f_{\tau,j}$ is a univariate component function which might be nonlinear, linear or zero. Since we assume that we don't know which components are linear before statistical analysis, we do not explicitly write out the linear terms in (1). With a smaller number of parameters to be estimated, existence of linear components will generally improve efficiency of estimation and inferences, if we can identify the linear components. Quantile regression (Koenker and Bassett 1978) has been demonstrated to be valuable by a rapidly expanding literature in economics, social sciences, and biomedical studies (Buchinsky 1994; Abrevaya 2001; Cade and Noon 2003; Yoshida 2014). It provides more robust analyses and more complete descriptions of data structure than traditional mean regression. Quantile regression for additive models has previously been considered in the literature. In the frequentist con-

Electronic supplementary material The online version of this article (doi:10.1007/s11222-013-9446-9) contains supplementary material, which is available to authorized users.

Y. Hu · K. Zhao · H. Lian (✉)
Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore
e-mail: henglian@ntu.edu.sg

text, Gooijer and Zerom (2003), Horowitz and Lee (2005), Yu and Lu (2004) all developed methodologies of nonparametric estimation for additive quantile regression. In the Bayesian context, Yue and Rue (2011) proposes a Bayesian quantile regression approach for additive mixed models.

There are some works focusing on the selection of significant components using penalization approaches from the frequentist perspective, such as Ravikumar et al. (2009), Meier et al. (2009) and Huang et al. (2010). Besides, a number of existing papers are concerned with the selection of function components using Bayesian inferences recently (Panagiotelis and Smith 2008; Shively et al. 1999; Yau et al. 2003). They express each function as a linear combination of basis terms and assign priors on the coefficients of the basis functions. They all introduce indicator variables to enable variable selection. However none of the works mentioned above considers linear component identification. These works considered only least squares regression. Usually, pre-specification of linear components is required. Recently, several works consider performing variable selection, parametric component identification, and parameter estimation all at the same time (Zhang et al. 2011; Lian 2012) from a frequentist perspective.

The quantile regression approach we propose in this article has the ability of separating function components into those with nonlinear effects, those with linear effects, and those irrelevant to responses, in the context of quantile regression. In a Bayesian context, this separation of components is a soft decision based on the posterior probabilities of the components being selected as nonlinear, linear, or zero. We establish a hierarchical Bayesian model by adopting the asymmetric Laplace distribution for errors (Yu and Moyeed 2011). Then, extending the Bayesian variable selection approach, we introduce two sets of indicator variables in the spike-and-slab priors which make the separation of components possible. Scheipl et al. (2012) also conducted a similar study which can separate the components into nonlinear, linear and zero ones, for generalized additive models.

The remainder of the paper proceeds as follows. In Sect. 2, we describe our hierarchical Bayesian model for quantile regression based on the additive model structure. We also discuss our prior choices and introduce the posterior sampling algorithm focusing on an efficient partially collapsed sampler. The details of the algorithm are explained in the Appendix. In Sect. 3, we present numerical illustrations including simulation studies and two real data examples. In Sect. 4, we conclude the paper with a discussion.

2 Hierarchical Bayesian modeling

Quantile regression is typically achieved by solving a minimization problem based on the check loss function. With model (1), the specific problem is estimating μ_τ and the sequence of functions $f_{\tau,j}$, $j = 1, \ldots, p$, by minimizing the following objective function,

$$L(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{n} \rho_\tau \left(y_i - \mu_\tau - \sum_{j=1}^{p} f_{\tau,j}(x_{ij}) \right), \tag{2}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, and $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is the so called *check function*. In a Bayesian setup, we assume $\epsilon_{\tau,i}$, $i = 1, \ldots, n$, are i.i.d. random variables from an asymmetric Laplace distribution with density

$$p(\epsilon_{\tau,i}) = \frac{\tau(1-\tau)}{\delta_0} \exp\left\{ -\frac{1}{\delta_0} \rho_\tau(\epsilon_{\tau,i}) \right\},$$

where δ_0 is the scale parameter. Then the conditional distribution of $\boldsymbol{y}$ is in the form of

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{\tau^n(1-\tau)^n}{\delta_0^n} \exp\left\{ -\frac{1}{\delta_0} \sum_{i=1}^{n} \rho_\tau \left(y_i - \mu_\tau - \sum_{j=1}^{p} f_{\tau,j}(x_{ij}) \right) \right\}. \tag{3}$$

Hence, maximizing the likelihood (3) is equivalent to minimizing (2), giving (2) a likelihood-based interpretation. By introducing the location-scale mixture representation of the asymmetric Laplace distribution (Kozumi and Kobayashi 2011), (3) can be equivalently written as

$$y_i = \mu_\tau + \sum_{j=1}^{p} f_{\tau,j}(x_{ij}) + k_1 e_i + \sqrt{k_2 \delta_0 e_i} z_i, \tag{4}$$

where $e_i \sim \exp(1/\delta_0)$ follows an exponential distribution with mean δ_0, z_i follows the standard normal distribution and is independent of e_i, $k_1 = \frac{1-2\tau}{\tau(1-\tau)}$, and $k_2 = \frac{2}{\tau(1-\tau)}$. For ease of notation, we will omit τ in the expressions in the following.

We assume the distribution of x_j, $j = 1, \ldots, p$, is supported on $[0, 1]$ and also impose the condition $Ef_j(x_j) = 0$ for identifiability. To model each unknown function f_j flexibly, we use the truncated power splines to approximate the functions in this article. Let $t_0 = 0 < t_1 < \cdots < t_k < 1 = t_{k+1}$ partition $[0, 1]$ into subintervals $[t_i, t_{i+1})$, $i = 0, \ldots, k$ with k internal knots. Here we focus on equally spaced knots although more complicated data-driven choice can be considered. For a given degree q, $B_0(x), \ldots, B_K(x)$ are used to denote $K + 1 = q + k$ truncated power spline basis $x, x^2, \ldots, x^q, (x - t_1)^q I(x > t_1), \ldots, (x - t_k)^q I(x > t_k)$. Because of the centering constraint $Ef_j(x_j) = 0$, we use the centered basis $\{B_{jk}(x) = B_k(x) - \sum_{i=1}^{n} B_k(x_{ij})/n, k = 0, \ldots, K\}$ with $K = q + k - 1$. We separate linear basis (a single linear function) and nonlinear basis to enable

the identification of linear and nonlinear components. Thus, with splines approximation, we have

$$y_i = \mu + \sum_{j=1}^{p} \alpha_j B_{j0}(x_{ij}) + \sum_{j=1}^{p}\sum_{k=1}^{K} \beta_{jk} B_{jk}(x_{ij}) + k_1 e_i + \sqrt{k_2 \delta_0 e_i} z_i, \quad i = 1, \ldots, n.$$

We view $e_i,\ i = 1, \ldots, n$ as latent variables and denote $\boldsymbol{e} = (e_1, \ldots, e_n)^T$. By defining $\boldsymbol{E} = k_2\delta_0 \mathrm{diag}(\mathrm{e}_1, \ldots, \mathrm{e}_n)$, $\boldsymbol{B}_0 = (\boldsymbol{B}_{10}, \ldots, \boldsymbol{B}_{p0})$ with $\boldsymbol{B}_{j0} = (B_{j0}(x_{1j}), \ldots, B_{j0}(x_{nj}))^T$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^T$, $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jK})^T$ and

$$\boldsymbol{B}_j = \begin{pmatrix} B_{j1}(x_{1j}) & B_{j2}(x_{1j}) & \ldots & B_{jK}(x_{1j}) \\ B_{j1}(x_{2j}) & B_{j2}(x_{2j}) & \ldots & B_{jK}(x_{2j}) \\ \vdots & \vdots & \vdots & \vdots \\ B_{j1}(x_{nj}) & B_{j2}(x_{nj}) & \ldots & B_{jK}(x_{nj}) \end{pmatrix},$$

the full conditional distribution of $\boldsymbol{y}$ can be expressed as,

$$p(\boldsymbol{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \boldsymbol{e}, \delta_0, \boldsymbol{x}, \mu) \propto \exp\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{f} - k_1\boldsymbol{e})^T \boldsymbol{E}^{-1}(\boldsymbol{y} - \boldsymbol{f} - k_1\boldsymbol{e})\}(\det[\boldsymbol{E}])^{-1/2}. \quad (5)$$

where

$$\boldsymbol{f} = \mu \mathbf{1}_n + \boldsymbol{B}_0\boldsymbol{\alpha} + \sum_{j=1}^{p} \boldsymbol{B}_j\boldsymbol{\beta}_j,$$

and $\mathbf{1}_n = (1, \ldots, 1)^T$ is a vector of dimension n with all components 1.

We choose spike-and-slab priors for α_j and $\boldsymbol{\beta}_j$ following George and McCulloch (1993), Panagiotelis and Smith (2008), and many others to enable variable selection and linear component selection. We introduce indicator variables $\boldsymbol{\gamma}^{(\nu)} = (\gamma_1^{(\nu)}, \ldots, \gamma_p^{(\nu)})^T$, such that $\nu_j = 0$ if and only if $\gamma_j^{(\nu)} = 0$, with $\nu = \alpha, \boldsymbol{\beta}$. In other words, f_j is regarded as a nonlinear function, a linear function or a zero function under the situation that $(\gamma_j^{(\boldsymbol{\beta})} = 1)$, $(\gamma_j^{(\alpha)} = 1, \gamma_j^{(\boldsymbol{\beta})} = 0)$, or $(\gamma_j^{(\alpha)} = 0, \gamma_j^{(\boldsymbol{\beta})} = 0)$, respectively.

For $\boldsymbol{\alpha}$, we choose an independent Gaussian prior on each component, which is also called *ridge prior* (Goldstein and Smith 1974),

$$p(\boldsymbol{\alpha}|\boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma}) \sim N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha),$$
$$\boldsymbol{\Sigma}_\alpha = \mathrm{diag}(\gamma_1^{(\alpha)}\sigma_1^2, \gamma_2^{(\alpha)}\sigma_2^2, \ldots, \gamma_p^{(\alpha)}\sigma_p^2),$$

where $\mathbf{0}_p$ is a zero vector of dimension p and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_p)$.

For $\boldsymbol{\beta}_j$, we take the conjugate prior usually undertaken in a Bayesian context $p(\boldsymbol{\beta}_j) \propto \exp\{-\frac{1}{2\tau_j^2}\boldsymbol{\beta}_j^T \boldsymbol{\Omega}_j \boldsymbol{\beta}_j\}$ when $\gamma_j^{(\boldsymbol{\beta})} = 1$, such as Smith and Kohn (1996) and Chib and Jeliazkov (2006). Here the (k, k') entry of $\boldsymbol{\Omega}_j$ is $\int_0^1 B_{jk}''(x) B_{jk'}''(x)dx$ (B_{jk}'' is the second derivative of B_{jk}). Since $\boldsymbol{\beta}_j = \mathbf{0}_K$ when $\gamma_j^{(\boldsymbol{\beta})} = 0$, the prior of $\boldsymbol{\beta}_j$ can be described as,

$$p\left(\boldsymbol{\beta}_j|\gamma_j^{(\boldsymbol{\beta})}, \tau_j\right) \sim N\left(\mathbf{0}_K, \gamma_j^{(\boldsymbol{\beta})}\tau_j^2\boldsymbol{\Omega}_j^{-1}\right).$$

We use the same prior as in Cripps et al. (2005) on $\boldsymbol{\gamma}^{(\nu)}$, $\nu = \alpha, \boldsymbol{\beta}$,

$$p(\boldsymbol{\gamma}^{(\nu)}) = \frac{1}{p+1}\begin{pmatrix} p \\ q_{\gamma^{(\nu)}} \end{pmatrix}^{-1},$$

where $q_{\gamma^{(\nu)}}$ is the number of non-zero ν_j. Under this prior, equal weights are placed on $\boldsymbol{\gamma}^{(\nu)}$ with different numbers of non-zero ν_j. Alternatively, we can use $p(\boldsymbol{\gamma}^{(\nu)}) = \pi^{q_{\gamma^{(\nu)}}}(1-\pi)^{p-q_{\gamma^{(\nu)}}}$ with π a constant. When $\pi = 1/2$, we have the uniform prior on $\boldsymbol{\gamma}^{(\nu)}$. We find the results are not sensitive to this choice of prior.

To summarize, the Bayesian hierarchical formulation is given by,

$$\boldsymbol{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \boldsymbol{e}, \delta_0, \boldsymbol{x}, \mu \sim N(\boldsymbol{f} + k_1\boldsymbol{e}, \boldsymbol{E}),$$
$$\boldsymbol{\alpha}|\boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma} \sim N(\mathbf{0}_p, \boldsymbol{\Sigma}_\alpha),\ \boldsymbol{\gamma}^{(\alpha)} \sim p\left(\boldsymbol{\gamma}^{(\alpha)}\right),$$
$$\boldsymbol{\beta}_j|\gamma_j^{(\boldsymbol{\beta})}, \tau_j \sim N\left(\mathbf{0}_K, \gamma_j^{(\boldsymbol{\beta})}\tau_j^2\boldsymbol{\Omega}_j^{-1}\right),\ \boldsymbol{\gamma}^{(\boldsymbol{\beta})} \sim p\left(\boldsymbol{\gamma}^{(\boldsymbol{\beta})}\right),$$
$$e_i \overset{i.i.d.}{\sim} \exp(1/\delta_0),\ \delta_0 \sim p(\delta_0),$$
$$\sigma_j^2 \sim p\left(\sigma_j^2\right),\ \tau_j^2 \sim p\left(\tau_j^2\right),\ \mu \sim p(\mu),$$

where $p(\delta_0)$, $p(\sigma_j^2)$ and $p(\tau_j^2)$ represent the hyperpriors of δ_0, σ_j^2 and τ_j^2. They are set to be $IG(a_1, a_2)$, where IG denotes the inverse Gamma distribution, and a_1 and a_2 are set to be 0.5 in all our numerical experiments as an uninformative choice. Sensitivity analysis reveals that our results are not sensitive to these choices. Finally, we use an uninformative prior on μ, $p(\mu) \propto 1$.

We use the Metropolis-within-Gibbs algorithm to sample from the posterior distributions. To improve mixing of the Markov chains, we integrate out ν_j, $\nu = \alpha, \boldsymbol{\beta}$ in some of the sampling steps, resulting in a partially collapsed sampler (Dyk and Park 2008). The readers are referred to the Appendix for details.

3 Numerical illustrations

We demonstrate the performance of our proposed quantile regression approach (denoted by BQPLAM) in terms of its estimation accuracy and model selection accuracy. The MCMC algorithm is implemented in R, and available upon

request. For comparison, we consider the following 4 additional methods.

Method 1: The hierarchical Bayesian model in Sect. 2 can be easily revised to deal with partially linear additive mean regression model. Consider the following model,

$$y_i = \mu + \sum_{j=1}^{p} f_j(x_{ij}) + \epsilon_i, \quad i = 1, \ldots, n,$$

where $\epsilon_i, i = 1, \ldots, n$ are i.i.d. normally distributed with mean zero and variance δ_0^2. The Bayesian hierarchical formulation will then be as follows,

$$\begin{aligned}
&\boldsymbol{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \delta_0^2, \boldsymbol{x}, \mu \sim N(\boldsymbol{f}, \delta_0^2 \boldsymbol{I}_{n\times n}),\\
&\boldsymbol{\alpha}|\boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma} \sim N(\boldsymbol{0}_p, \boldsymbol{\Sigma}_\alpha),\ \boldsymbol{\gamma}^{(\alpha)} \sim p\left(\boldsymbol{\gamma}^{(\alpha)}\right),\\
&\boldsymbol{\beta}_j, \tau_j \sim N\left(\boldsymbol{0}_K, \gamma_j^{(\boldsymbol{\beta})}\tau_j^2 \boldsymbol{\Omega}_j^{-1}\right),\ \boldsymbol{\gamma}^{(\boldsymbol{\beta})} \sim p\left(\boldsymbol{\gamma}^{(\boldsymbol{\beta})}\right),\\
&\delta_0^2 \sim p\left(\delta_0^2\right),\ \tau_j^2 \sim p\left(\tau_j^2\right),\ \mu \sim p(\mu),
\end{aligned}$$

where $\boldsymbol{I}_{n\times n}$ represents the $n \times n$ identity matrix. We denote this mean regression method to be BPLAM.

The other three are quantile regression approaches.

Method 2: We consider a Bayesian additive model that combines nonlinear and linear terms together. This only performs variable selection but not linear component identification. To be specific, the Bayesian additive regression is based on model (5), with

$$\boldsymbol{f} = \mu \boldsymbol{1}_n + \sum_{j=1}^{p} \boldsymbol{D}_j \boldsymbol{\beta}_j,$$

where $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jK})^T$ and

$$\boldsymbol{D}_j = \begin{pmatrix} B_{j0}(x_{1j}) & B_{j1}(x_{1j}) & \ldots & B_{jK}(x_{1j}) \\ B_{j0}(x_{2j}) & B_{j1}(x_{2j}) & \ldots & B_{jK}(x_{2j}) \\ \vdots & \vdots & \vdots & \vdots \\ B_{j0}(x_{nj}) & B_{j1}(x_{nj}) & \ldots & B_{jK}(x_{nj}) \end{pmatrix}.$$

With similar prior distributions and hyper-parameters as those described in Sect. 2, we obtain the following Bayesian hierarchical formulation,

$$\begin{aligned}
&\boldsymbol{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \boldsymbol{e}, \delta_0, \boldsymbol{x}, \mu \sim N(\boldsymbol{f} + k_1\boldsymbol{e}, \boldsymbol{E}),\\
&\boldsymbol{\beta}_j|\gamma_j^{(\boldsymbol{\beta})}, \tau_j \sim N\left(\boldsymbol{0}_{K+1}, \gamma_j^{(\boldsymbol{\beta})}\tau_j^2 \boldsymbol{\Omega}_j^{-1}\right),\ \boldsymbol{\gamma}^{(\boldsymbol{\beta})} \sim p\left(\boldsymbol{\gamma}^{(\boldsymbol{\beta})}\right),\\
&e_i \overset{i.i.d.}{\sim} \exp(1/\delta_0),\ \delta_0 \sim p(\delta_0),\\
&\tau_j^2 \sim p\left(\tau_j^2\right),\ \mu \sim p(\mu).
\end{aligned}$$

This method is denoted by BQAM_V.

Method 3: We consider Bayesian quantile linear regression based on model (5), with

$$\boldsymbol{f} = \mu \boldsymbol{1}_n + \boldsymbol{B}_0 \boldsymbol{\alpha},$$

in which all components are assumed to be linear. The Bayesian hierarchical structure is as follows,

$$\begin{aligned}
&\boldsymbol{y}|\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \boldsymbol{e}, \delta_0, \boldsymbol{x}, \mu \sim N(\boldsymbol{f} + k_1\boldsymbol{e}, \boldsymbol{E}),\\
&\boldsymbol{\alpha}|\boldsymbol{\gamma}^{(\alpha)}, \boldsymbol{\sigma} \sim N(\boldsymbol{0}_p, \boldsymbol{\Sigma}_\alpha),\ \boldsymbol{\gamma}^{(\alpha)} \sim p\left(\boldsymbol{\gamma}^{(\alpha)}\right),\\
&e_i \overset{i.i.d.}{\sim} \exp(1/\delta_0),\ \delta_0 \sim p(\delta_0),\\
&\sigma_j^2 \sim p\left(\sigma_j^2\right),\ \mu \sim p(\mu).
\end{aligned}$$

The indicator variables $\boldsymbol{\gamma}^{(\alpha)}$ enable component selection. This method is denoted by BQLM_V.

Method 4: The third quantile regression method for comparison is still based on the model (5), except that we fix all indicator variables $\gamma_j^{(\alpha)},\ j = 1, \ldots, p$, and $\gamma_j^{(\boldsymbol{\beta})},\ j = 1, \ldots, p$, to be 1. This method can only estimate component functions but cannot select variables. We denote it as BQAM_NV.

3.1 Simulation examples

We generate $n = 100$ observations $(\boldsymbol{x}_i, y_i),\ i = 1, \ldots, 100$, from the following heteroscedastic additive model

$$y_i = \sum_{j=1}^{p} f_j(x_{ij}) + (0.5 + x_{i2})\epsilon_i, \tag{6}$$

with $f_1(x) = \sin(2\pi \text{x})/(2 - \sin(2\pi \text{x}))$, $f_2(x) = 5x(1-x)$, $f_3(x) = 2x$, $f_4(x) = x$, $f_5(x) = -x$ and $p = 10$. Thus, the first 2 components are nonlinear components, followed by 3 linear components and 5 zero components. The covariates x_{ij} are generated from the standard normal distribution with correlations given by $Cov(x_{ij_1}, x_{ij_2}) = (1/2)^{|j_1 - j_2|}$, and then transformed to be marginally uniform on $[0, 1]$ by applying the cdf of the standard normal distribution. We consider two distributions of ϵ_i, a normal distribution with mean 0 and standard deviation 0.5, and a Student's t distribution with scale parameter 1/3 and degrees of freedom 2. We fit the mean regression model and the four quantile regression models at five different quantile levels $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. For each scenario, 100 data sets are generated and fitted. For each replicate, the MCMC algorithm is run for 20,000 iterations with a burn-in of 10,000.

The performance was evaluated by the integrated squared error (*ISE*), for each component function, which is approx-

Table 1 Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is normal and $p = 10$

		f_1	f_2	f_3	f_4	f_5	f_6	f
	BPLAM	0.165	0.109	0.069	0.067	0.061	0.016	0.219
		(0.049)	(0.043)	(0.051)	(0.057)	(0.061)	(0.032)	(0.047)
$\tau = 0.5$	BQPLAM	0.162	0.112	0.068	0.073	0.068	0.010	0.222
		(0.048)	(0.049)	(0.050)	(0.076)	(0.066)	(0.029)	(0.048)
	BQAM_v	0.156	0.161	0.156	0.190	0.176	0.006	0.224
		(0.046)	(0.055)	(0.054)	(0.076)	(0.080)	(0.011)	(0.072)
	BQLM_v	0.320	0.364	0.075	0.112	0.108	0.015	0.515
		(0.026)	(0.015)	(0.064)	(0.091)	(0.092)	(0.034)	(0.041)
	BQAM_nv	0.157	0.127	0.101	0.104	0.123	0.114	0.320
		(0.053)	(0.051)	(0.058)	(0.059)	(0.071)	(0.073)	(0.114)
$\tau = 0.1$	BQPLAM	0.195	0.184	0.105	0.106	0.110	0.023	0.287
		(0.066)	(0.074)	(0.068)	(0.091)	(0.092)	(0.059)	(0.084)
	BQAM_v	0.199	0.222	0.189	0.233	0.236	0.007	0.310
		(0.074)	(0.073)	(0.063)	(0.064)	(0.060)	(0.016)	(0.096)
	BQLM_v	0.359	0.377	0.107	0.133	0.147	0.046	0.526
		(0.063)	(0.036)	(0.093)	(0.114)	(0.099)	(0.083)	(0.053)
	BQAM_nv	0.148	0.165	0.158	0.150	0.146	0.151	0.396
		(0.045)	(0.060)	(0.062)	(0.052)	(0.054)	(0.059)	(0.139)
$\tau = 0.3$	BQPLAM	0.170	0.136	0.070	0.072	0.072	0.014	0.233
		(0.050)	(0.054)	(0.048)	(0.066)	(0.065)	(0.039)	(0.055)
	BQAM_v	0.164	0.167	0.157	0.233	0.221	0.008	0.262
		(0.050)	(0.060)	(0.050)	(0.055)	(0.061)	(0.013)	(0.089)
	BQLM_v	0.335	0.362	0.092	0.134	0.131	0.020	0.522
		(0.044)	(0.048)	(0.072)	(0.098)	(0.101)	(0.046)	(0.046)
	BQAM_nv	0.142	0.158	0.160	0.145	0.146	0.147	0.383
		(0.042)	(0.064)	(0.068)	(0.053)	(0.053)	(0.076)	(0.136)
$\tau = 0.7$	BQPLAM	0.166	0.132	0.070	0.084	0.073	0.010	0.239
		(0.045)	(0.055)	(0.052)	(0.076)	(0.071)	(0.030)	(0.064)
	BQAM_v	0.160	0.179	0.156	0.240	0.223	0.007	0.278
		(0.049)	(0.065)	(0.058)	(0.055)	(0.062)	(0.010)	(0.096)
	BQLM_v	0.359	0.375	0.077	0.114	0.114	0.014	0.517
		(0.063)	(0.033)	(0.068)	(0.086)	(0.090)	(0.029)	(0.043)
	BQAM_nv	0.150	0.166	0.153	0.145	0.144	0.155	0.379
		(0.056)	(0.045)	(0.056)	(0.056)	(0.049)	(0.055)	(0.126)
$\tau = 0.9$	BQPLAM	0.196	0.204	0.101	0.115	0.117	0.026	0.311
		(0.053)	(0.078)	(0.066)	(0.101)	(0.093)	(0.058)	(0.092)
	BQAM_v	0.181	0.226	0.179	0.246	0.234	0.003	0.357
		(0.055)	(0.081)	(0.069)	(0.056)	(0.057)	(0.004)	(0.108)
	BQLM_v	0.316	0.405	0.101	0.119	0.121	0.028	0.534
		(0.025)	(0.075)	(0.081)	(0.092)	(0.095)	(0.064)	(0.063)
	BQAM_nv	0.159	0.181	0.161	0.155	0.160	0.164	0.424
		(0.052)	(0.060)	(0.057)	(0.059)	(0.051)	(0.059)	(0.141)

Standard errors based on simulations are shown in the *parentheses*

imated over an equally spaced 1,000 points $(t_1, \ldots, t_T)$, $T = 1,000$ on [0, 1] by,

$$\widehat{ISE} = \frac{1}{T}\sum_{i=1}^{T}(\hat{f}_j(t_i) - f_j(t_i))^2, \tag{7}$$

where $f_j(t_i)$ is the true value of function f_j at t_i, and $\hat{f}_j(t_i)$ is the posterior mean of f_j at t_i based on the 10,000 iterations after burn-in. Tables 1 and 2 summarize the average and standard deviation of $\sqrt{\widehat{ISE}}$ over 100 replicates for the first six

Table 2 Summary of $\sqrt{\widehat{ISE}}$ for each univariate function and $f = \sum_{j=1}^{10} f_j$ over 100 replicates, when the distribution of error is Student's t and $p = 10$

		f_1	f_2	f_3	f_4	f_5	f_6	f
	BPLAM	0.278	0.232	0.129	0.134	0.135	0.032	0.411
		(0.081)	(0.104)	(0.106)	(0.095)	(0.084)	(0.043)	(0.161)
$\tau = 0.5$	BQPLAM	0.176	0.126	0.072	0.088	0.087	0.007	0.244
		(0.055)	(0.067)	(0.050)	(0.089)	(0.089)	(0.023)	(0.073)
	$BQAM_v$	0.145	0.154	0.137	0.193	0.186	0.032	0.269
		(0.051)	(0.068)	(0.058)	(0.074)	(0.069)	(0.053)	(0.098)
	$BQLM_v$	0.319	0.361	0.075	0.136	0.121	0.015	0.517
		(0.029)	(0.029)	(0.064)	(0.098)	(0.099)	(0.041)	(0.041)
	$BQAM_{nv}$	0.160	0.161	0.158	0.165	0.161	0.147	0.439
		(0.057)	(0.060)	(0.065)	(0.062)	(0.058)	(0.059)	(0.160)
$\tau = 0.1$	BQPLAM	0.295	0.274	0.157	0.176	0.165	0.027	0.405
		(0.104)	(0.129)	(0.129)	(0.099)	(0.098)	(0.059)	(0.144)
	$BQAM_v$	0.230	0.283	0.244	0.238	0.241	0.073	0.492
		(0.089)	(0.157)	(0.155)	(0.105)	(0.124)	(0.176)	(0.301)
	$BQLM_v$	0.373	0.370	0.144	0.191	0.186	0.048	0.529
		(0.063)	(0.027)	(0.123)	(0.105)	(0.096)	(0.088)	(0.055)
	$BQAM_{nv}$	0.259	0.251	0.258	0.261	0.247	0.231	0.610
		(0.239)	(0.178)	(0.296)	(0.244)	(0.265)	(0.235)	(0.392)
$\tau = 0.3$	BQPLAM	0.201	0.142	0.076	0.097	0.093	0.008	0.268
		(0.080)	(0.065)	(0.064)	(0.092)	(0.086)	(0.029)	(0.087)
	$BQAM_v$	0.153	0.166	0.154	0.193	0.186	0.037	0.274
		(0.062)	(0.072)	(0.061)	(0.076)	(0.071)	(0.058)	(0.108)
	$BQLM_v$	0.339	0.359	0.098	0.136	0.120	0.020	0.523
		(0.049)	(0.022)	(0.077)	(0.100)	(0.103)	(0.050)	(0.041)
	$BQAM_{nv}$	0.183	0.185	0.178	0.185	0.170	0.164	0.464
		(0.082)	(0.071)	(0.081)	(0.079)	(0.072)	(0.065)	(0.173)
$\tau = 0.7$	BQPLAM	0.182	0.146	0.078	0.098	0.096	0.006	0.266
		(0.057)	(0.085)	(0.061)	(0.094)	(0.094)	(0.021)	(0.089)
	$BQAM_v$	0.163	0.167	0.154	0.205	0.195	0.032	0.308
		(0.054)	(0.076)	(0.073)	(0.069)	(0.067)	(0.050)	(0.129)
	$BQLM_v$	0.317	0.369	0.083	0.130	0.130	0.013	0.519
		(0.028)	(0.027)	(0.077)	(0.093)	(0.099)	(0.037)	(0.048)
	$BQAM_{nv}$	0.191	0.205	0.189	0.175	0.168	0.178	0.505
		(0.070)	(0.110)	(0.081)	(0.080)	(0.078)	(0.076)	(0.020)
$\tau = 0.9$	BQPLAM	0.279	0.293	0.146	0.160	0.156	0.031	0.408
		(0.100)	(0.134)	(0.112)	(1.090)	(0.116)	(0.080)	(0.155)
	$BQAM_v$	0.254	0.264	0.272	0.247	0.235	0.083	0.575
		(0.130)	(0.134)	(0.174)	(0.091)	(0.072)	(0.131)	(0.362)
	$BQLM_v$	0.345	0.396	0.130	0.172	0.175	0.048	0.576
		(0.054)	(0.084)	(0.107)	(0.100)	(0.105)	(0.085)	(0.111)
	$BQAM_{nv}$	0.253	0.258	0.259	0.268	0.220	0.217	0.679
		(0.130)	(0.139)	(0.165)	(0.189)	(0.097)	(0.123)	(0.387)

components and for the regression function $f = \sum_{j=1}^{10} f_j$. From the results, we can see that BQPLAM is obviously more efficient than $BQAM_V$ and $BQAM_{NV}$ for the parametric components due to the separated linear basis and its associated indicator variables. $BQLM_V$ performs poorly as expected, since it cannot capture nonlinear effects of the components. BQPLAM and $BQAM_V$ outperform $BQAM_{NV}$ for zero components (note that we only present results of f_6 among the five zero components). The results show that the two sets of indicators can help reduce errors besides select-

Table 3 Summary of testing errors for mean estimators and median estimators over 100 replicates, when $p = 10$. "Normal" and "Student's t" indicate the distribution of ϵ_i

	RMSE	AD
Normal		
BPLAM	0.586	0.457
	(0.030)	(0.026)
BQPLAM	0.587	0.459
	(0.031)	(0.026)
BQAM_{υ}	0.646	0.505
	(0.032)	(0.026)
BQLM_{υ}	0.762	0.604
	(0.030)	(0.025)
BQAM_{nv}	0.696	0.540
	(0.251)	(0.165)
Student's t		
BPLAM	1.224	0.648
	(0.079)	(0.101)
BQPLAM	1.160	0.551
	(0.032)	(0.054)
BQAM_{υ}	1.350	0.589
	(0.022)	(0.031)
BQLM_{υ}	1.354	0.679
	(0.015)	(0.022)
$\text{BQAM}_{n\upsilon}$	1.236	0.666
	(0.043)	(0.058)

Table 4 Summary of testing errors for quantile estimators at different quantile levels over 100 replicates, when $p = 10$. "Normal" and "Student's t" indicate the distribution of ϵ_i

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
Normal				
BQPLAM	0.634	0.311	0.322	0.660
	(0.100)	(0.041)	(0.046)	(0.128)
BQAM_{υ}	0.652	0.338	0.351	0.669
	(0.083)	(0.045)	(0.048)	(0.096)
BQLM_{υ}	0.868	0.415	0.424	0.846
	(0.119)	(0.058)	(0.064)	(0.113)
BQAM_{nv}	0.684	0.356	0.355	0.664
	(0.084)	(0.065)	(0.048)	(0.089)
Student's t				
BQPLAM	0.819	0.360	0.365	0.843
	(0.219)	(0.061)	(0.063)	(0.182)
BQAM_{υ}	0.831	0.393	0.391	0.853
	(0.261)	(0.063)	(0.058)	(0.228)
BQLM_{υ}	0.988	0.464	0.464	0.967
	(0.206)	(0.060)	(0.566)	(0.146)
BQAM_{nv}	0.835	0.445	0.456	0.850
	(0.514)	(0.077)	(0.075)	(0.309)

ing components. The results of mean regression are similar to those of median regression when the errors follow a normal distribution. However, the advantage of median regression is more obvious when the errors follow a Student's t distribution.

To measure the prediction accuracy of our method, we generate $n' = 100{,}000$ independent test samples from the same generating model (6) and present the test errors in Tables 3 and 4. For mean and median regression, we consider two error measures including the root mean squared errors (RMSE) and absolute deviation errors (AD). For quantile regressions at quantiles $\{0.1, 0.3, 0.7, 0.9\}$, the prediction error refers to the average check loss (ACL). More sepcifically,

$$RMSE = \sqrt{\frac{1}{n'}\sum_{i=1}^{n'}(\hat{y}_i - y_i)^2},$$

$$AD = \frac{1}{n'}\sum_{i=1}^{n'}|\hat{y}_i - y_i|$$

and

$$ACL = \frac{1}{n'}\sum_{i=1}^{n'}\rho_\tau(\hat{y}_i - y_i),$$

where $y_i,\ i = 1, \ldots, n'$, are the responses of the testing data and $\hat{y}_i, i = 1, \ldots, n'$, are the predicted values estimated with

$$\hat{y}_i = \hat{\mu} + \sum_{j=1}^{p}\hat{f}_j(x_{ij}),$$

where $\hat{\mu}$ and $\hat{f}_j(x_{ij})$ are posterior mean based on the 10,000 sampled values after burn-in. Our method BQPLAM results in smaller test errors and outperforms the other three quantile regression methods at all quantile levels. The median regression performs similarly as mean regression when the errors are Gaussian, while it outperforms mean regression when errors are Student's t as expected.

The results of zero and linear component selection are based on the estimation of posterior distribution of $\boldsymbol{\gamma}^{(\alpha)}$ and $\boldsymbol{\gamma}^{(\boldsymbol{\beta})}$. For Bayesian approaches, many may agree that a "hard decision" for variable selection is not necessary and one can report the posterior probabilities which already take into account all sources of uncertainties. However, in some situations a hard decision is still desirable. We will use marginal posterior probabilities similar to George and McCulloch (1993); Barbieri and Berger (2004). The approach based

Table 5 Summary of the component selection results for mean estimators and median estimators over 100 replicates, when $p = 10$

	Normal	Student's t
BPLAM		
#of nonzero variables	5.21	4.69
	(0.67)	(1.89)
#of correct nonzero variables	4.91	4.00
	(0.32)	(1.21)
#of linear variables	3.19	3.22
	(0.82)	(1.99)
#of correct linear variables	2.81	2.16
	(0.46)	(1.03)
BQPLAM		
#of nonzero variables	5.14	4.69
	(0.69)	(1.01)
#of correct nonzero variables	4.88	4.54
	(0.39)	(0.93)
#of linear variables	2.90	2.58
	(0.94)	(0.98)
#of correct linear variables	2.56	2.35
	(0.69)	(0.92)
BQAM_v		
#of nonzero variables	4.51	5.66
	(0.93)	(2.23)
#of correct nonzero variables	4.43	4.31
	(0.87)	(0.85)
BQLM_v		
#of nonzero variables	4.39	3.82
	(1.18)	(1.26)
#of correct nonzero variables	3.90	3.49
	(0.89)	(0.97)

"Normal" and "Student's t" indicate the distribution of ϵ_i

on false discovery rate as in Müller et al. (2006) could also be used, which is however more suitable for problems with much higher dimensions.

More specifically, we compare the estimated values of the posterior probabilities $p(\gamma_j^{(\boldsymbol{\beta})} = 1|\boldsymbol{y}, \boldsymbol{x})$, $p(\gamma_j^{(\alpha)} = 1, \gamma_j^{(\boldsymbol{\beta})} = 0|\boldsymbol{y}, \boldsymbol{x})$ and $p(\gamma_j^{(\alpha)} = 0, \gamma_j^{(\boldsymbol{\beta})} = 0|\boldsymbol{y}, \boldsymbol{x})$. The component will be identified as nonlinear, linear or zero according to which of the three probabilities is the largest (this rule is used only for illustration and Bayesians usually do not perform this step of making hard decision). These marginal probabilities are estimated by the proportion of the respective events in the posterior samples after burn-in. In Tables 5, 6, and 7, we report the number of nonzero components selected, number of nonzero components selected that are nonzero in the true model, number of linear components selected, and the number of linear components selected that are linear in the true model for BQPLAM and BPLAM. We report the number of

Table 6 Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Gaussian and $p = 10$

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
BQPLAM				
#of nonzero variables	5.00	5.01	4.92	4.92
	(1.00)	(0.77)	(0.73)	(1.05)
#of correct nonzero variables	4.56	4.77	4.73	4.53
	(0.75)	(0.52)	(0.61)	(0.79)
#of linear variables	2.40	2.67	2.58	2.45
	(1.39)	(1.05)	(1.05)	(1.39)
#of correct linear variables	1.97	2.40	2.35	2.00
	(1.04)	(0.84)	(0.88)	(1.00)
BQAM_v				
#of nonzero variables	4.17	4.37	4.28	4.21
	(0.91)	(0.91)	(0.81)	(0.83)
#of correct nonzero variables	4.12	4.24	4.26	4.17
	(0.90)	(0.89)	(0.81)	(0.83)
BQLM_v				
#of nonzero variables	4.44	3.88	4.19	4.88
	(1.92)	(1.91)	(1.18)	(1.54)
#of correct nonzero variables	3.45	3.49	3.75	3.52
	(1.07)	(1.05)	(0.90)	(0.96)

Table 7 Summary of the component selection results for quantile estimators at different quantile levels over 100 replicates, when the distribution of errors is Student's t and $p = 10$

	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.7$	$\tau = 0.9$
BQPLAM				
#of nonzero variables	4.15	4.58	4.57	4.31
	(1.53)	(0.96)	(0.94)	(1.66)
#of correct nonzero variables	3.72	4.51	4.49	3.76
	(1.36)	(0.93)	(0.87)	(1.07)
#of linear variables	2.31	2.49	2.49	2.63
	(1.38)	(1.11)	(1.12)	(1.96)
#of correct linear variables	1.76	2.28	2.29	1.65
	(1.03)	(0.92)	(1.10)	(1.02)
BQAM_v				
#of nonzero variables	4.20	5.45	5.18	4.22
	(2.07)	(2.23)	(2.07)	(1.97)
#of correct nonzero variables	3.69	4.21	4.13	3.66
	(0.93)	(0.82)	(0.83)	(0.85)
BQLM_v				
#of nonzero variables	3.68	3.78	3.94	4.07
	(1.93)	(1.56)	(1.39)	(2.08)
#of correct nonzero variables	2.81	3.35	3.58	3.02
	(1.26)	(1.08)	(1.05)	(1.29)

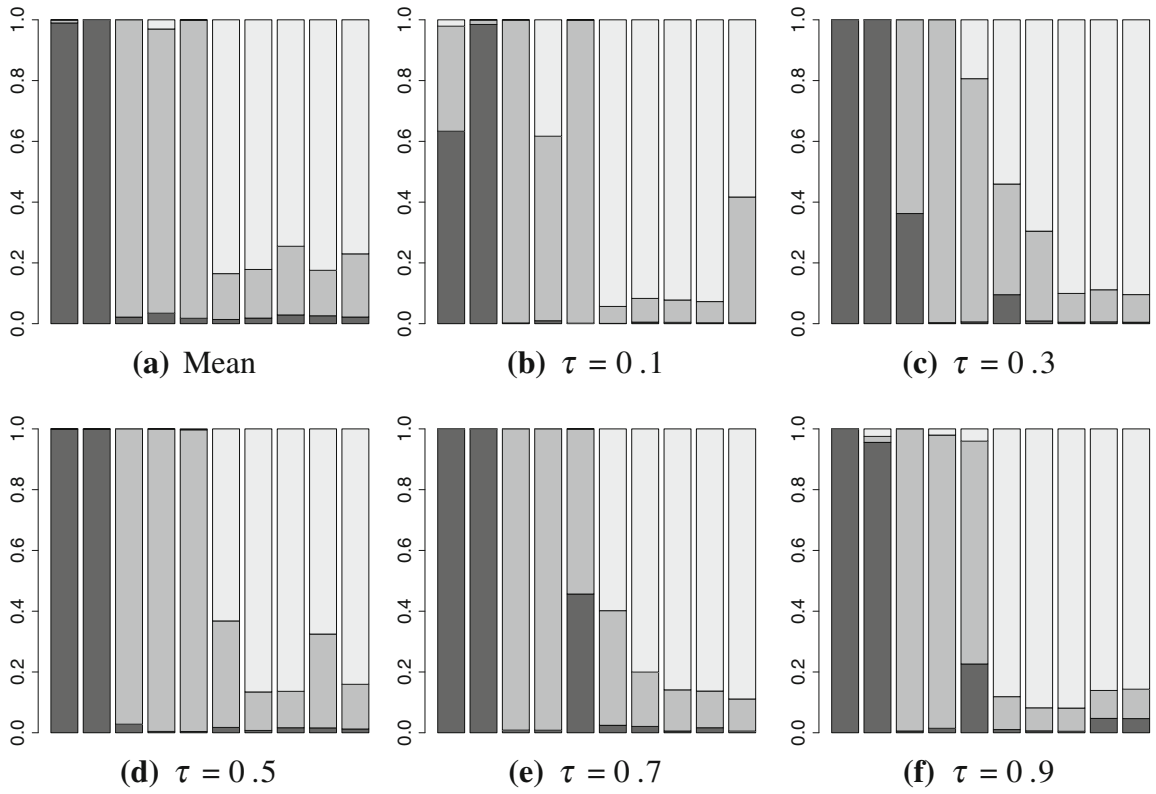

Fig. 1 Component selection results for one randomly selected replicate, when the distribution of errors is Gaussian and $p = 10$. The *black*, *dark grey*, and *light grey* areas represent the posterior probabilities of the component being nonlinear, linear and zero, respectively

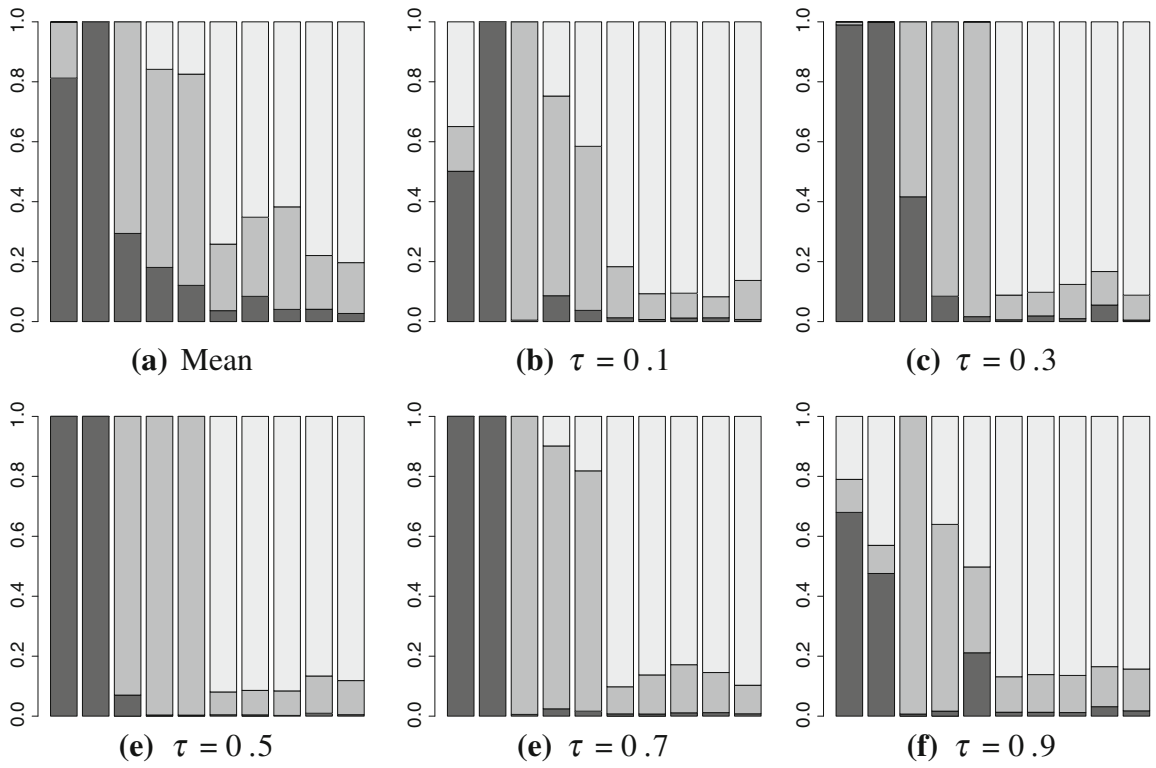

Fig. 2 Component selection results for one randomly selected replicate, when the distribution of errors is Student's t and $p = 10$

Table 8 Summary of testing errors for mean and median estimators over 100 replicates, when $p = 50$ and 100

	$p = 50$		$p = 100$	
	RMSE	AD	RMSE	AD
Normal				
BPLAM	0.624	0.482	0.638	0.497
	(0.055)	(0.046)	(0.060)	(0.050)
BQPLAM	0.625	0.483	0.643	0.506
	(0.057)	(0.048)	(0.054)	(0.045)
Student's t				
BPLAM	1.251	0.685	1.563	0.709
	(0.069)	(0.088)	(0.657)	(0.108)
BQPLAM	1.176	0.569	1.185	0.588
	(0.039)	(0.065)	(0.052)	(0.081)

"Normal" and "Student's t" indicate the distribution of ϵ_i

Table 9 Summary of the component selection results for mean and median estimators over 100 replicates, when $p = 50$ and 100

	$p = 50$		$p = 100$	
	Normal	Student's t	Normal	Student's t
BPLAM				
#of nonzero variable	4.32	2.73	4.82	2.42
	(1.12)	(1.45)	(1.11)	(1.73)
#of correct nonzero variable	4.19	2.57	4.10	2.21
	(0.99)	(1.38)	(1.11)	(1.49)
#of linear variable	2.38	1.92	2.26	1.61
	(1.30)	(1.21)	(1.04)	(1.40)
#of correct linear variable	2.10	1.45	1.96	1.31
	(1.07)	(0.93)	(0.99)	(1.32)
BQPLAM				
#of nonzero variable	4.41	4.05	4.29	3.72
	(0.17)	(1.17)	(1.07)	(1.41)
#of correct nonzero variable	4.30	4.02	4.20	3.72
	(1.05)	(1.16)	(1.04)	(1.43)
#of linear variable	2.24	2.13	2.16	2.06
	(1.26)	(0.97)	(1.07)	(1.10)
#of correct linear variable	2.06	1.91	1.98	1.81
	(1.03)	(1.05)	(1.06)	(1.01)

"Normal" and "Student's t" indicate the distribution of ϵ_i

variables selected and the number of variables selected that are truly nonzero for BQAM_v and BLAM_v, since these two methods can only select nonzero components. Figures 1 and 2 display estimates of the three posterior probabilities for each component based on one randomly selected replicate among the 100. The black, dark grey, and light grey areas represent the percentage of times the component is selected as nonlinear, linear and zero, respectively. It is seen from the figures

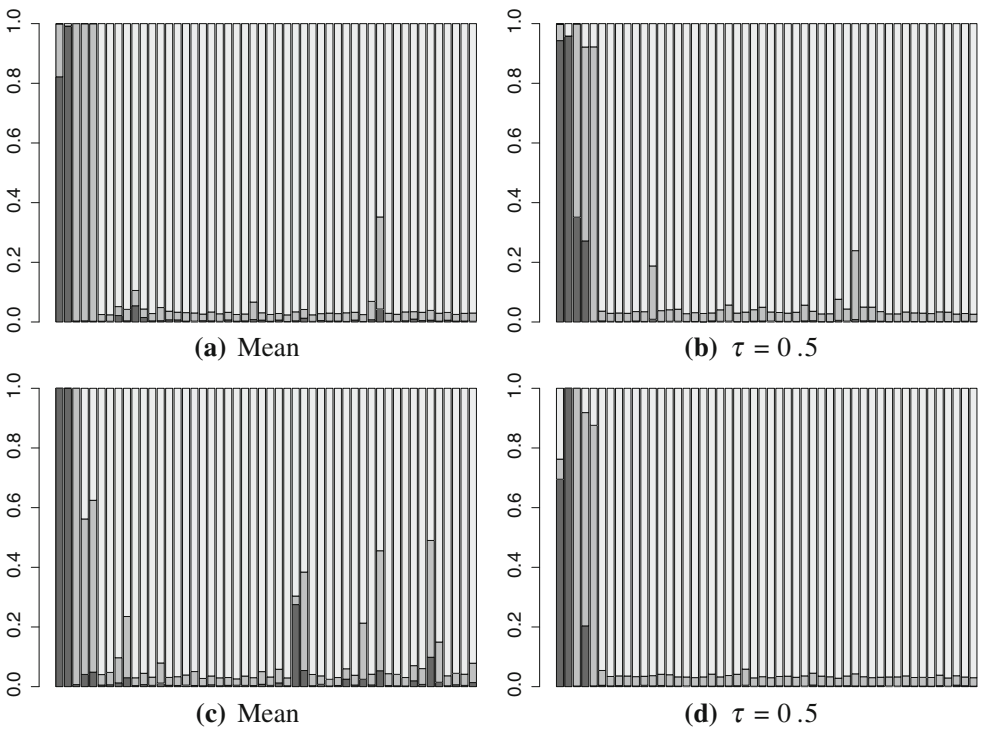

Fig. 3 Component selection results for one randomly selected replicate with $p = 50$. The *upper panels* are results when error distribution is Gaussian, and the *lower panels* are results when the error distribution is Student's t

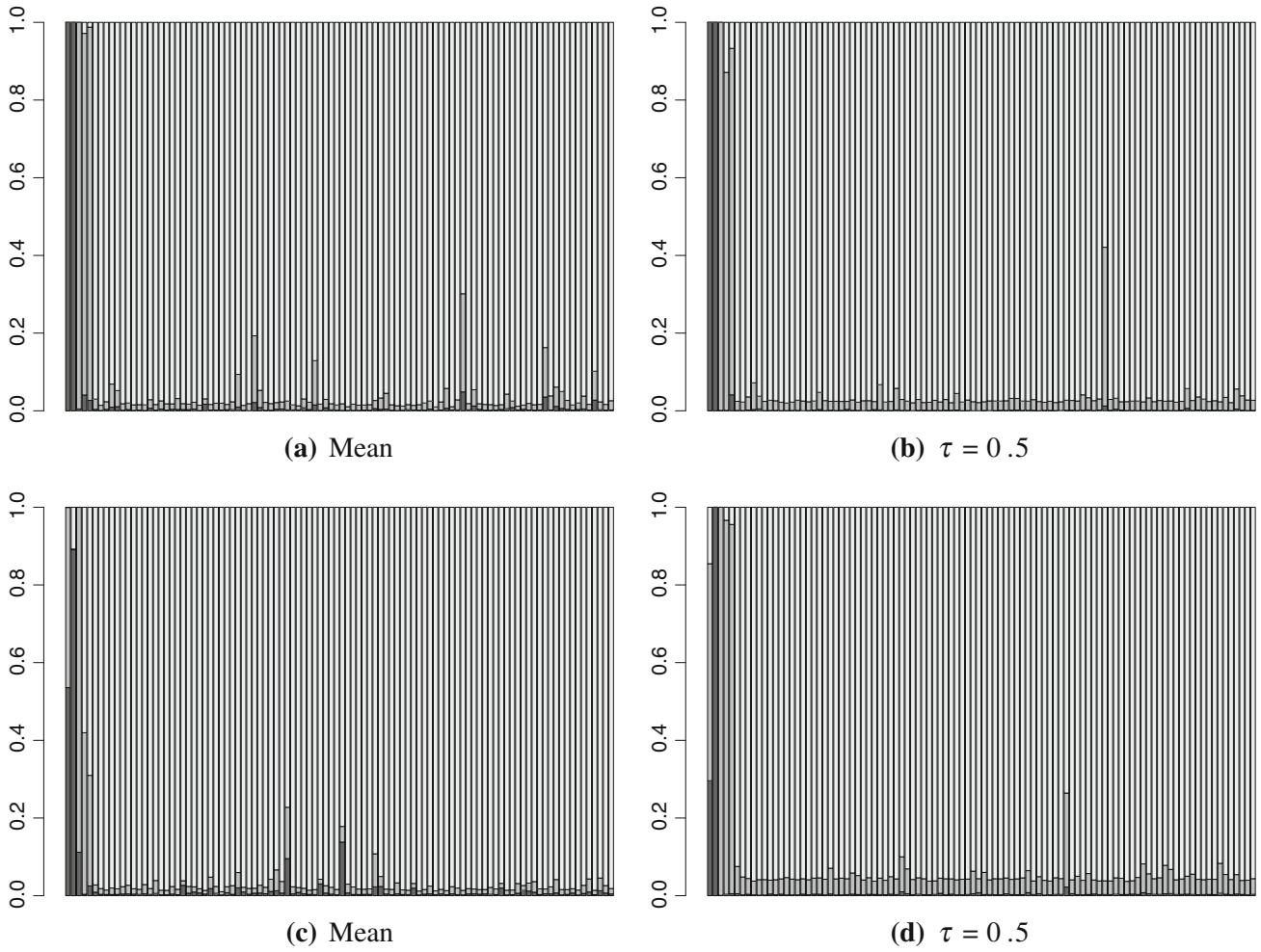

Fig. 4 Component selection results for one randomly selected replicate with $p = 100$. The *upper panels* are results when error distribution is Gaussian, and the *lower panels* are results when the error distribution is Student's t

that BQPLAM can detect linear and nonlinear components quite accurately.

Next, we increase the dimension in the example to examine the performance of our proposed method in higher dimensions. We consider the dimension $p = 50$ and 100. To save space, we only demonstrate the results of BQPLAM at $\tau = 0.5$ and BPLAM here. Test errors and variable selection results are presented in Tables 8 and 9. The estimates of three posterior probabilities for each component (for one data set) are displayed in Figs. 3 and 4. The results show that our proposed method still performs well.

We have also conducted other simulations which are put in the supplementary material. These include simulations with $p = 50$ for quantile regression with τ other than 0.5, simulations with 10 nonzero components, and simulations with skewed error distribution. These extra simulations will hopefully give a more complete picture of the performance of the proposed approach.

3.2 Real data examples

Now we illustrate the methodology with two data sets. In the following examples, we fix $\boldsymbol{\gamma}^{(\boldsymbol{\beta})}$ associated with the discrete dummy variables to be zero all the time, which means that dummy variables can only be modeled as linear or zero based on the posterior estimate of $\boldsymbol{\gamma}^{(\alpha)}$. All covariates are standardized by a linear transformation to lie in [0, 1] before analysis.

3.2.1 GDP growth data

We apply our method to a GDP growth data. This data set includes data for 74 countries over two 20-year time intervals: 1960–1979 and 1980–1999, with sample size $n = 147$. This data set is used by Tan (2010) to uncover the interplay between geography, institutions, and fractionalization in economic development by employing a regression tree analysis. Here we use our method to detect the determinants

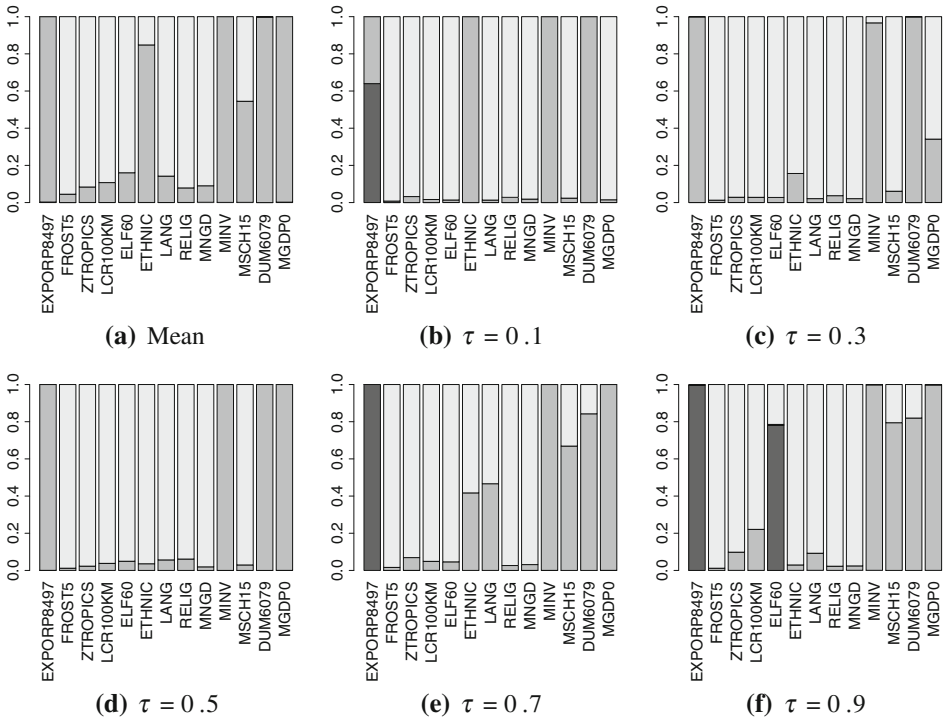

Fig. 5 Component selection results for the GDP growth data

of economic growth. The dependent variable is the difference between the logarithmic values of real per capita GDP for the start and end years of each 20-year time interval. We consider four categories of independent variables, a total of 13. The first covariate we consider is a variable measuring the quality of expropriation risk of government across the years 1984–1997, which is denoted by EXPROP8497. Tan (2010) also considers another variable ICRG8497 to measure the quality of institution. Since the correlation between the two variables is very high, at over 0.8, we drop the second one and only keep EXPROP8497. The second category of independent variables includes: proportion of a country's land area that experiences more than 5 frost days per month in winter (FROST5), percentage of a country's land area that is classified as a tropical eco-zone (ZTROPICS), and percentage of a country's land area within 100 km of an ice-free coast (LCR100KM). This category is used as proxy for climate. The third category measures degree of ethnic fractionalization, including ELF60, ETHNIC, LANG and RELIG. ELF60 measures the probability in 1960 that two randomly selected people from a given country will not belong to the same ethno-linguistic subgroup. ETHNIC combines racial and linguistic characteristics, LANG is based on data for shares of languages spoken as 'mother tongues', and RELIG describes differences in religion. The last category of independent variables includes several familiar neoclassical determinants: log net depreciation rate (MNGD), log investment share (MINV), log schooling (MSCH15), a dummy variable for the period 1960–1979 (DUM6079), and log initial per capita income (MGDP0).

In Fig. 5, we show the variable selection results of BPLAM and BQPLAM. For mean regression, we identified 6 nonzero components, all of which are linear components. For quantile regression, we identified 4, 3, 4, 5 and 6 nonzero components, of which 1, 0, 0, 1, 2 components are nonlinear at 5 quantile levels respectively. The barplots show that some covariates only have effects at lower or upper quantiles. For example, MSCH15 and ELF60 only have effects at upper quantiles, and ETHNIC only affects lower quantiles. In Fig. 6, we show the fitted curves of the estimated regression functions (posterior mean based on the 10,000 iterations after burn-in) as one covariate varies while others are fixed at 0.5. We performed some convergence diagnostic for this data set and the details are contained in the supplementary material.

To summarize, besides those neoclassicial determinants, which are fundamental determinants used in the production function, both institutions and ethnic fractionalization are found to have effects on economical development while geography is found to play no roles. These results are consistent with the results of Tan (2010), and lead to the same conclusion that more attention should be paid on the impact of quality of institutions and degree of ethnic fractionalization on the development of economics.

3.2.2 Housing price data

Another data set we use is a housing price data. This data set is used to investigate the relationship between house price and several features, including physical attributes of the residential houses, characteristics of the surrounding rural dis-

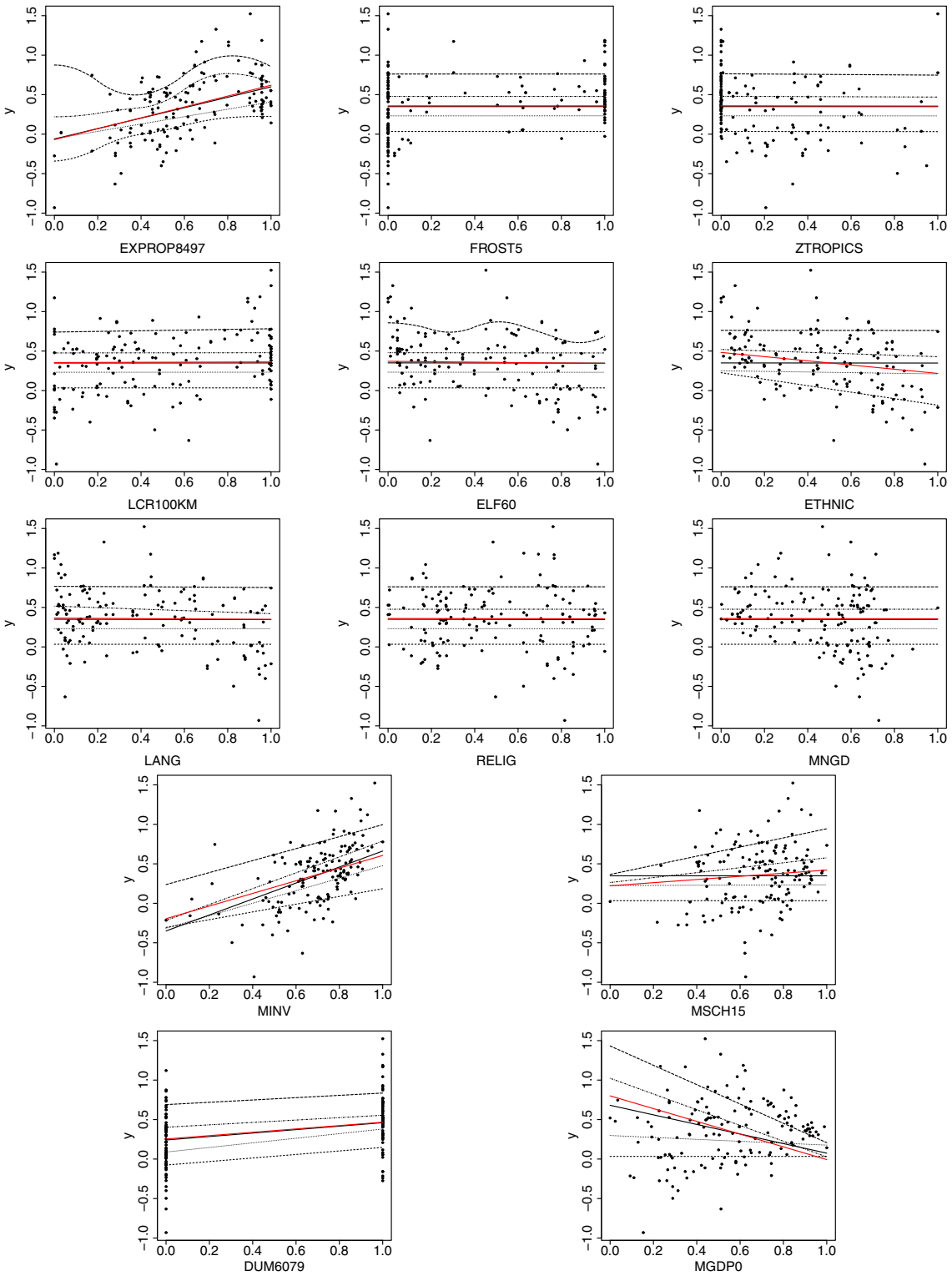

Fig. 6 The fitted regression functions for the GDP growth data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. The *red solid lines* are the fitted curves of mean regression. (Color figure online)

trict and environmental nuisances. This application concerns transaction prices ($n = 2053$) for the residential houses sold during 1996 and 1997 in Brittany. We consider four physical characteristics of a house and its adjacent lot: age of the house (AGE), state of repair (REPAIR)(dummy), number of rooms (ROOMS), and lot size (LOT). We use another

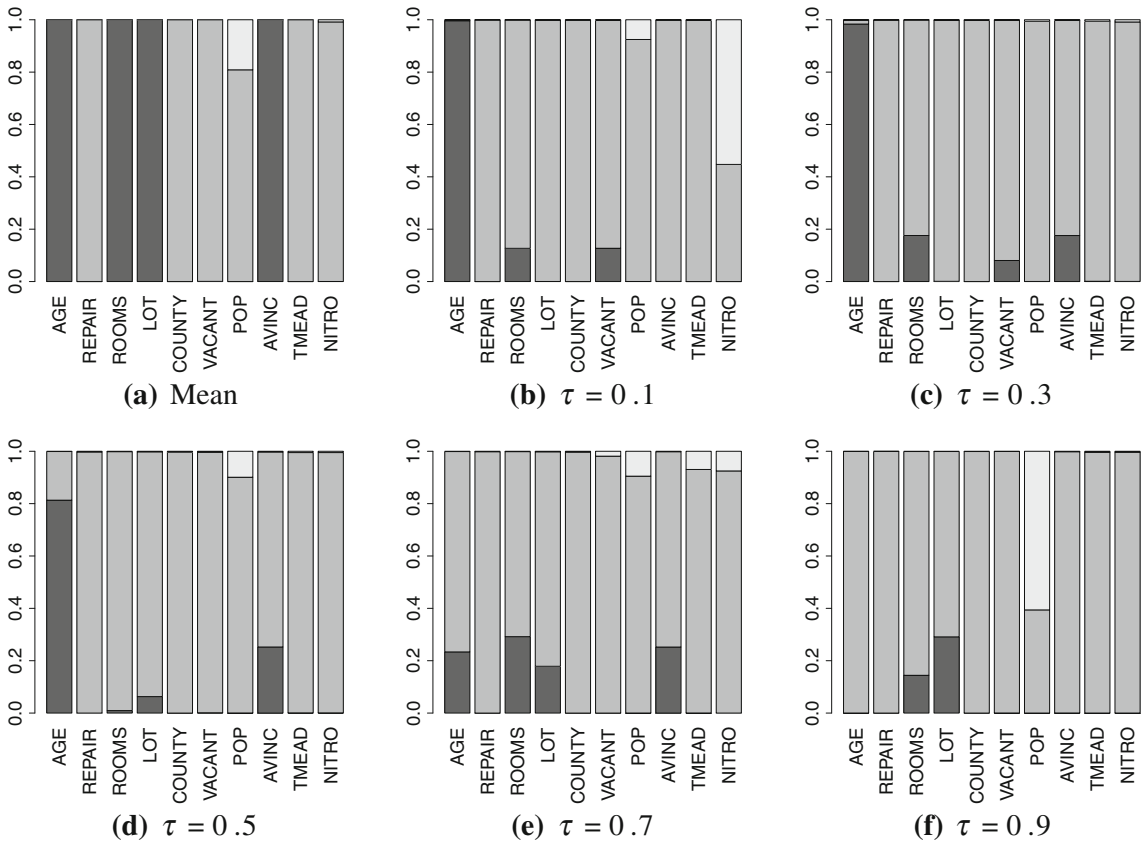

Fig. 7 Component selection results for the housing price data

four indicators on the characteristics of the surrounding rural district: the population of the district (POP), average taxable family income (AVINC), proportion of vacant houses (VACANT), and a dummy variable indicating whether or not the surrounding district is located in the department of Ille et Vilaine (COUNTY). The environmental nuisances are expressed by two indicators, amount of nitrogen emissions from livestock farming per hectare of arable land in the rural district where the residential house is located (NITRO), and proportion of permanent grassland converted into cultivated grassland (TMEAD). This data set was used by Bontemps et al. (2008) to investigate the effects of agricultural pollution on house prices. Bontemps et al. (2008) pre-specifes all the explanatory variables (AGE, REPAIR, ROOMS, LOT, COUNTY, VACANT, POP and AVINC) except the two environment indicators (TMEAD and NITRO) to be linear components in the semiparametric models, and incorporated the two environmental indicators in a nonparametric way.

We display the barplots in Fig. 7 and the fitted regression functions in Fig. 8. The results show that all the components are identified as nonzero components and 4 of them are nonlinear for mean regression. For quantile regression at the 5 quantile levels, we identified 9, 10, 10, 10 and 9 nonzero components, and 1, 1, 1, 0, 0 nonlinear components respectively. Examining the influence of the physical characteristics of houses on prices at different levels, we find that a larger number of rooms, a bigger lot size and the fact that a house has been repaired are factors contributing to an increase in the price of a house, while an older age has a negative impact on the price. Similarly, the characteristics of the surrounding districts show their effects on house prices that conform to our expectations. For example, the price of houses located in the districts of the most urbanized county of Brittany (Ille-et-Vilaine) is higher, and the price of residential houses located in districts with lower housing vacancy rates is higher. For the two environmental indicators, we reveal their negative effects on house prices. The effects of covariates on responses at different quantile levels show some degree of heterogeneity. For example, the effect of AGE is linear at quantile 0.9, but is nonlinear at lower quantiles; the negative effect of NITRO is not obvious at quantile 0.1, but more obvious at upper quantiles.

4 Conclusion and discussions

In this article, we have proposed a Bayesian quantile regression method for partially linear additive models, which

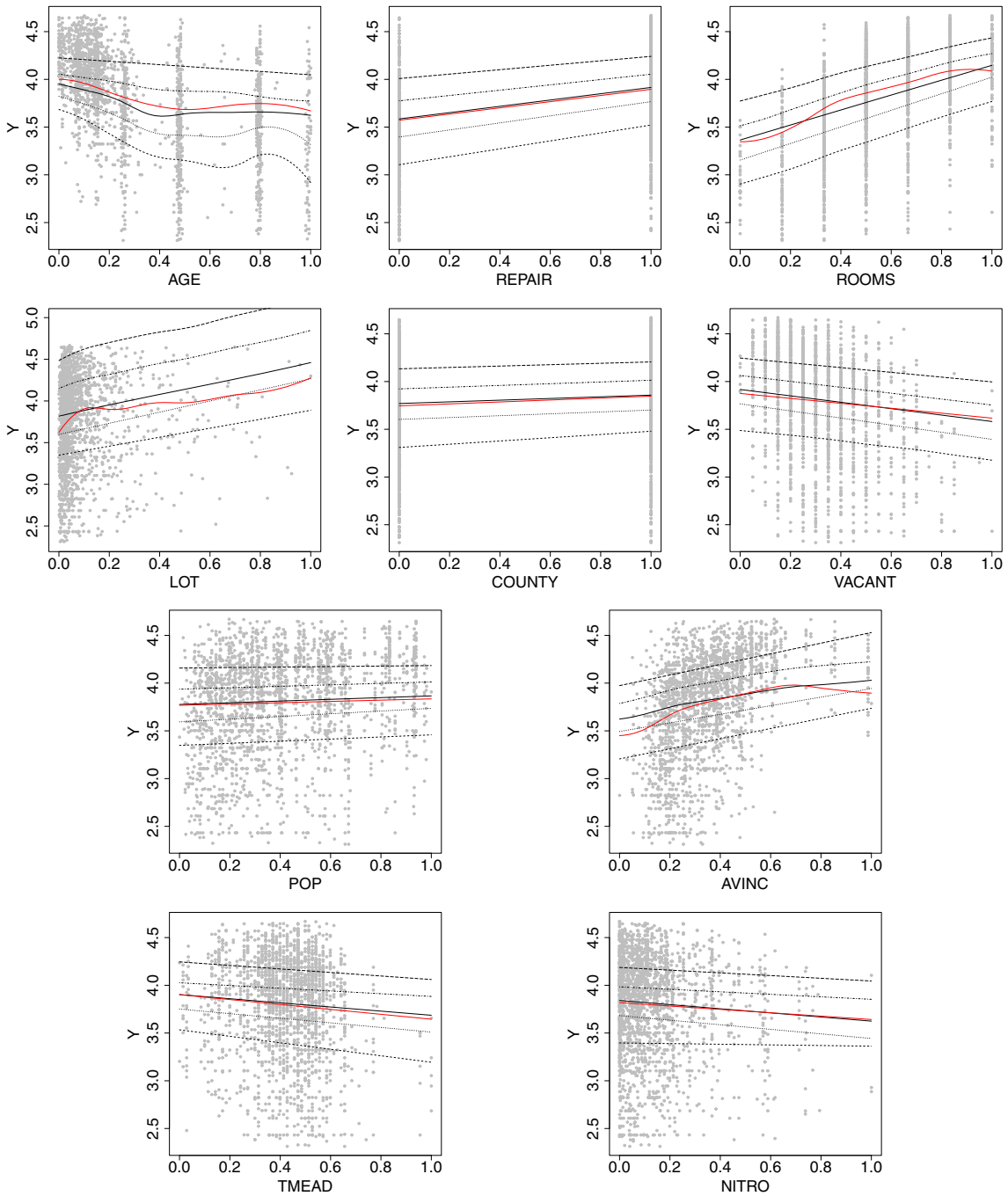

Fig. 8 The fitted regression functions for housing price data when one covariate varies and others are fixed at 0.5, at quantile levels $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. The *red solid lines* are the fitted curves of mean regression. (Color figure online)

explicitly models components that have linear and nonlinear effects. As detailed in the Appendix, we designed an efficient MCMC algorithm for posterior inferences. With simulation studies, we illustrated the empirical performances of our proposed quantile regression approach and demonstrated the efficacy of the two sets of indicator variables. The method can automatically determine the type of component effects. The performance of the proposed approach in our simulations is quite encouraging, even when p is large.

Note that the error distributions we use in the simulation examples are very different from the asymmetric Laplace distribution. Our simulation results show that our methods perform well even when the errors are not generated based on their assumed generating mechanism. Theoretically it has

been recently demonstrated that Bayesian quantile regression based on the asymmetric Laplace distribution is consistent even when the error distribution is misspecified Sriram et al. (2013). More specifically, Sriram et al. (2013) showed that if the τ-conditional quantile of y_i is $\alpha_0+\boldsymbol{\beta}_0^T\boldsymbol{x}_i$, without assuming the conditional distribution of y_i given $\boldsymbol{x}_i$ is ALD, the posterior distribution for Bayesian quantile regression using ALD will concentrate in any open neighborhood of $(\alpha_0, \boldsymbol{\beta}_0)$ in this parametric setting for both proper and improper priors that satisfy some mild assumptions. Although their results are valid only for the parametric setting, it is naturally conjectured that consistency also holds in our semiparametric models. Li et al. (2010) and Hu et al. (2013) have also discussed this problem and showed that Bayesian methods are not sensitive to this assumption. On the other hand, we note that consistency under misspecification does not necessarily mean that inferences (that is, credible intervals constructed) are also valid under misspecification, which deserves some studies in the future.

A problem with BQPLAM, as with most Bayesian approaches based on MCMC algorithm, is its computational burden. In terms of computational speed, on our desktop PC, it takes about 20 minutes to fit BQPLAM on a single generated dataset with $n = 100$ and $p = 10$. For this reason, we can only deal with datasets with sample size in the order of several hundreds. One possibility to deal with this problem is to use a variational Bayes approximation in place of MCMC. Another possibility is to parallelize the sampling of parameters associated with different predictors, which would be helpful if p is large.

The number of knots can also be regarded as a tuning parameter in our model. However, this is less critical in a Bayesian approach where overfitting is automatically controlled by the prior. We find that empirically the results change very little with 5–10 internal knots (we used 5 internal knots for all numerical examples). One could use adaptive knot selection to achieve possibly even better performance as in Kohn et al. (2001) by introducing more indicator variables at the cost of increased computational burden.

Finally, using asymmetric Laplace distribution is only one among a few possible alternatives for Bayesian quantile regression. A problem of the current approach is that regression functions at different quantile levels are separately estimated, which can cause the quantile curves at different levels to cross each other. Other approaches may be able to address this issue (He 1997; Reich et al. 2011; Tokdar and Kadane 2011), but we choose to use asymmetric Laplace distribution for its simplicity and computational convenience.

Acknowledgments The authors want to thank the AE and three anonymous reviewers for their insightful comments and suggestions that lead to significant improvements on the manuscript.

Appendix: MCMC algorithm details

The joint distribution of all the variables is

$$
\begin{aligned}
&p(\boldsymbol{\alpha}, \{\boldsymbol{\beta}_j\}, \boldsymbol{E}, \mu, \delta_0|\boldsymbol{y}, \boldsymbol{x}) \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\mu\mathbf{1}_n-\boldsymbol{B}_0\boldsymbol{\alpha}-\sum_{j=1}^{p}\boldsymbol{B}_j\boldsymbol{\beta}_j-k_1\boldsymbol{e}\right)^T\boldsymbol{E}^{-1}\right. \\
&\left.\left(\boldsymbol{y}-\mu\mathbf{1}_n-\boldsymbol{B}_0\boldsymbol{\alpha}-\sum_{j=1}^{p}\boldsymbol{B}_j\boldsymbol{\beta}_j-k_1\boldsymbol{e}\right)\right\}\times\det[\boldsymbol{E}]^{-1/2} \\
&\times p(\boldsymbol{\alpha})\times\prod_{j=1}^{p}p(\boldsymbol{\beta}_j)\times p(\delta_0)\times\prod_{i=1}^{n}p(e_i)\times p(\mu),
\end{aligned}
$$

where $p(\boldsymbol{\beta}_j)$, $p(\boldsymbol{\alpha})$, $p(e_i)$, $p(\delta_0)$ and $p(\mu)$ are the prior distributions of $\boldsymbol{\beta}_j$, $\boldsymbol{\alpha}$, e_i, δ_0, and μ respectively.

We use the Metropolis-within-Gibbs algorithm to sample from the posterior distribution. We integrate out α_j in step 3 and $\boldsymbol{\beta}_j$ in step 5 to improve mixing of the Markov Chain. The posterior distribution of each variable is as follows ($\sim$ denotes all variables except the one to be sampled):

1. Sample $p(\alpha_j|\sim) = p\left(\alpha_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)}\right)$, $j = 1, \ldots, p$, from the conditional distribution of α_j,

$$
\begin{aligned}
&p\left(\alpha_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)}=1\right)\sim N(\mu_j, \xi_j^2), \\
&p\left(\alpha_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \sigma_j^2, \gamma_j^{(\alpha)}=0\right)=0,
\end{aligned}
$$

where $\boldsymbol{y}^* = \boldsymbol{y}-\mu\mathbf{1}_n-\sum_{i\neq j}^{p}\alpha_i\boldsymbol{B}_{i0}-\sum_{i=1}^{p}\boldsymbol{B}_i\boldsymbol{\beta}_i-k_1\boldsymbol{e}$, $\xi_j^2 = \left(\boldsymbol{B}_{j0}^T\boldsymbol{E}^{-1}\boldsymbol{B}_{j0}+\frac{1}{\sigma_j^2}\right)^{-1}$ and $\mu_j = \xi_j^2\boldsymbol{B}_{j0}^T\boldsymbol{E}^{-1}\boldsymbol{y}^*$.

2. Sample $p(\mu|\sim) = p(\mu|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0)$, from the conditional distribution of μ,

$$
p(\mu|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0)\sim N(\mu_0, \xi_0^2),
$$

where $\xi_0^2 = k_2\delta_0(\sum_{i=1}^{n}e_i^{-1})^{-1}$, $\mu_0 = \xi_0^2\mathbf{1}_n^T\boldsymbol{E}^{-1}\boldsymbol{y}^*$, and $\boldsymbol{y}^* = \boldsymbol{y}-\sum_{i=1}^{p}\alpha_i\boldsymbol{B}_{i0}-\sum_{i=1}^{p}\boldsymbol{B}_i\boldsymbol{\beta}_i-k_1\boldsymbol{e}$.

3. Sample $p(\gamma_j^{(\alpha)}|\sim) = p(\gamma_j^{(\alpha)}|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \sigma_j^2)$, $j = 1, \ldots, p$, from its conditional posterior after integrating over α_j, $p(\gamma_j^{(\alpha)}=1|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \sigma_j^2) = \frac{1}{1+h}$, with

$$
\begin{aligned}
h &= h_1\frac{p\left(\gamma_j^{(\alpha)}=0|\gamma_{i\neq j}^{(\alpha)}\right)}{p\left(\gamma_j^{(\alpha)}=1|\gamma_{i\neq j}^{(\alpha)}\right)}, \\
h_1 &= \exp\left\{-\frac{1}{2}(\boldsymbol{B}_{j0}^T\boldsymbol{E}^{-1}\boldsymbol{y}^*)^2\left(\boldsymbol{B}_{j0}^T\boldsymbol{E}^{-1}\boldsymbol{B}_{j0}+\frac{1}{\sigma_j^2}\right)^{-1}\right\} \\
&\quad\times(\sigma_j^2\boldsymbol{B}_{j0}^T\boldsymbol{E}^{-1}\boldsymbol{B}_{j0}+1)^{\frac{1}{2}},
\end{aligned}
$$

where $p(\gamma_j^{(\alpha)} = 0|\gamma_{i\neq j}^{(\alpha)}) = (p - q_{\gamma_0^{(\alpha)}})/(p+1)$ and $p(\gamma_j^{(\alpha)} = 1|\gamma_{i\neq j}^{(\alpha)}) = (1 + q_{\gamma_0^{(\alpha)}})/(p+1)$ with $\gamma_0^{(\alpha)} = (\gamma_1^{(\alpha)}, \ldots, \gamma_{j-1}^{(\alpha)}, 0, \gamma_{j+1}^{(\alpha)}, \ldots, \gamma_p^{(\alpha)})^T$, and $\boldsymbol{y}^* = \boldsymbol{y} - \mu\mathbf{1}_n - \sum_{i\neq j}^p \alpha_i \boldsymbol{B}_{i0} - \sum_{i=1}^p \boldsymbol{B}_i\boldsymbol{\beta}_i - k_1\boldsymbol{e}$.

4. Sample $p(\boldsymbol{\beta}_j|\ \sim) = p(\boldsymbol{\beta}_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \tau_j^2, \gamma_j^{(\boldsymbol{\beta})})$, $j = 1, \ldots, p$,

$$p\left(\boldsymbol{\beta}_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \tau_j^2, \gamma_j^{(\boldsymbol{\beta})} = 1\right) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j),$$
$$p\left(\boldsymbol{\beta}_j|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \tau_j^2, \gamma_j^{(\boldsymbol{\beta})} = 0\right) = 0,$$

where $\boldsymbol{y}^* = \boldsymbol{y} - \mu\mathbf{1}_n - \sum_{i=1}^p \alpha_i \boldsymbol{B}_{i0} - \sum_{i\neq j}^p \boldsymbol{B}_i\boldsymbol{\beta}_i - k_1\boldsymbol{e}$, $\boldsymbol{\mu}_j = \boldsymbol{\Sigma}_j \boldsymbol{B}_j^T \boldsymbol{E}^{-1}\boldsymbol{y}^*$, $\boldsymbol{\Sigma}_j = (\boldsymbol{B}_j^T\boldsymbol{E}^{-1}\boldsymbol{B}_j + \frac{1}{\tau_j^2}\boldsymbol{\Omega}_j)^{-1}$.

5. Sample $p(\gamma_j^{(\boldsymbol{\beta})}|\ \sim) = p(\gamma_j^{(\boldsymbol{\beta})}|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \tau_j^2)$, $j = 1, \ldots, p$, from its conditional posterior after integrating over $\boldsymbol{\beta}_j$, $p(\gamma_j^{(\boldsymbol{\beta})} = 1|\boldsymbol{y}^*, \boldsymbol{e}, \delta_0, \tau_j^2) = \frac{1}{1+h}$, with

$$h = h_1 \frac{p\left(\gamma_j^{(\boldsymbol{\beta})} = 0|\ \gamma_{i\neq j}^{(\boldsymbol{\beta})}\right)}{p\left(\gamma_j^{(\boldsymbol{\beta})} = 1|\ \gamma_{i\neq j}^{(\boldsymbol{\beta})}\right)},$$
$$h_1 = \exp\left\{-\frac{1}{2}\left[\boldsymbol{y}^{*T}\boldsymbol{E}^{-1}\boldsymbol{B}_j\left(\boldsymbol{B}_j^T\boldsymbol{E}^{-1}\boldsymbol{B}_j + \frac{1}{\tau_j^2}\boldsymbol{\Omega}_j\right)^{-1}\boldsymbol{B}_j^T\boldsymbol{E}^{-1}\boldsymbol{y}^*\right]\right\} \times \det\left[\frac{1}{\tau_j^2}\boldsymbol{\Omega}_j\right]^{-\frac{1}{2}} \times \det\left[\boldsymbol{B}_j^T\boldsymbol{E}^{-1}\boldsymbol{B}_j + \frac{1}{\tau_j^2}\boldsymbol{\Omega}_j\right]^{\frac{1}{2}},$$

where $\boldsymbol{y}^* = \boldsymbol{y} - \mu\mathbf{1}_n - \sum_{i=1}^p \alpha_i \boldsymbol{B}_{i0} - \sum_{i\neq j}^p \boldsymbol{B}_i\boldsymbol{\beta}_i - k_1\boldsymbol{e}$, $p(\gamma_j^{(\boldsymbol{\beta})} = 0|\gamma_{i\neq j}^{(\boldsymbol{\beta})}) = (p - q_{\gamma_0^{(\boldsymbol{\beta})}})/(p+1)$ and $p(\gamma_j^{(\boldsymbol{\beta})} = 1|\gamma_{i\neq j}^{(\boldsymbol{\beta})}) = (1 + q_{\gamma_0^{(\boldsymbol{\beta})}})/(p+1)$ with $\gamma_0^{(\boldsymbol{\beta})} = (\gamma_1^{(\boldsymbol{\beta})}, \ldots, \gamma_{j-1}^{(\boldsymbol{\beta})}, 0, \gamma_{j+1}^{(\boldsymbol{\beta})}, \ldots, \gamma_p^{(\boldsymbol{\beta})})^T$.

6. Sample δ_0,

$$\delta_0 \sim IG(a_1 + 3n/2, \nu),$$
$$\nu = a_2 + \left((2k_2e_i)^{-1}\sum_{i=1}^n\left(y_i - \sum_{j=1}^p \alpha_j B_{j0}(x_{ij}) - \sum_{j=1}^p\sum_{k=1}^K \beta_{jk}B_{jk}(x_{ij}) - k_1e_i\right)^2 + e_i\right).$$

7. Sample σ_j^2, $j = 1, \ldots, p$, and τ_j^2, $j = 1, \ldots, p$, from their conditional posterior distributions if $\gamma_j^{(\alpha)}, \gamma_j^{(\boldsymbol{\beta})} \neq 0$,

$$\sigma_j^2 \sim IG(a_1 + 1/2, a_2 + (\alpha_j^2/2)),$$
$$\tau_j^2 \sim IG(a_1 + K/2, a_2 + (\boldsymbol{\beta}_j^T\boldsymbol{\Omega}_j\boldsymbol{\beta}_j/2)).$$

Otherwise they are generated from their priors.

8. The full conditional distribution of e_i, $i = 1, \ldots, n$ is a generalized inverse Gaussian distribution (*GIG*),

$$p(e_i|\delta_0, \nu_i) \sim GIG\left(\frac{1}{2}, \sqrt{\frac{(y_i - \nu_i)^2}{k_2\delta_0}}, \sqrt{\frac{k_1^2}{k_2\delta_0} + \frac{2}{\delta_0}}\right),$$
$$\nu_i = y_i - \mu - \sum_{j=1}^p \alpha_j B_{j0}(x_{ij}) - \sum_{j=1}^p\sum_{k=1}^K \beta_{jk}B_{jk}(x_{ij}),$$

where the probability density function of $GIG(\rho, m, n)$ is

$$f(x|\rho, m, n) = \frac{(n/m)^\rho}{2K_\rho(mn)}x^{\rho-1}\exp\left\{-\frac{1}{2}(m^2x^{-1} + n^2x)\right\},$$
$$x > 0,\ -\infty < \rho < \infty,\ m \geq 0,\ n \geq 0,$$

and K_ρ is the modified Bessel function of the third kind (Barndorff-Nielsen and Shephard 2001).

References

Abrevaya, J.: The effects of demographics and maternal behavior on the distribution of birth outcomes. Empiric. Econ. **26**(1), 247–257 (2001)

Barbieri, M.M., Berger, J.O.: Optimal predictive model selection. Ann. Stat. **32**(3), 870–897 (2004)

Barndorff-Nielsen, O., Shephard, N.: Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. J. R Stat. Soc. Ser. B **63**(2), 167–241 (2001)

Bontemps, C., Simioni, M., Surry, Y.: Semiparametric hedonic price models: assessing the effects of agricultural nonpoint source pollution. J. Appl. Econom. **23**(6), 825–842 (2008)

Buchinsky, M.: Changes in the US wage structure 1963–1987: application of quantile regression. Econometrica **62**(2), 405–458 (1994)

Cade, B.S., Noon, B.R.: A gentle introduction to quantile regression for ecologists. Frontiers Ecol. Environ. **1**(8), 412–420 (2003)

Chib, S., Jeliazkov, I.: Inference in semiparametric dynamic models for binary longitudinal data. J. Am. Stat. Assoc. **101**(474), 685–700 (2006)

Cripps, E., Carter, C., Kohn, R.: Variable selection and covariance selection in multivariate regression models. Handb. Stat. **25**(2), 519–552 (2005)

De Gooijer, J., Zerom, D.: On additive conditional quantiles with high-dimensional covariates. J. Am. Stat. Assoc. **98**(461), 135–146 (2003)

George, E., McCulloch, R.: Variable selection via Gibbs sampling. J. Am. Stat. Assoc. **88**(423), 881–889 (1993)

Goldstein, M., Smith, A.: Ridge-type estimators for regression analysis. J R Stat. Soc. Ser B **36**(2), 284–291 (1974)

He, X.: Quantile curves without crossing. Am. Stat. **51**(2), 186–192 (1997)

Horowitz, J., Lee, S.: Nonparametric estimation of an additive quantile regression model. J. Am. Stat. Assoc. **100**(472), 1238–1249 (2005)

Hu, Y., Gramacy, R., Lian, H.: Bayesian quantile regression for single-index models. Stat. Comput. **23**, 437–454 (2013)

Huang, J., Horowitz, J., Wei, F.: Variable selection in nonparametric additive models. Ann. Stat. **38**(4), 2282–2312 (2010)

Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**(1), 33–50 (1978)

Kohn, R., Smith, M., Chan, D.: Nonparametric regression using linear combinations of basis functions. Stat. Comput. **11**(4), 313–322 (2001)

Kozumi, H., Kobayashi, G.: Gibbs sampling methods for Bayesian quantile regression. J. Stat. Comput. Simul. **81**(11), 1565–1578 (2011)

Li, Q., Xi, R., Lin, N.: Bayesian regularized quantile regression. Bayesian Anal. **5**(3), 533–556 (2010)

Lian, H.: Identification of partially linear structure in additive models with an application to gene expression prediction from sequences. J. Bus. Econ. Stat. **30**, 337–350 (2012)

Liang, H., Thurston, S.W., Ruppert, D., Apanasovich, T., Hauser, R.: Additive partial linear models with measurement errors. Biometrika **95**(3), 667–678 (2008)

Meier, L., Van De Geer, S., Bühlmann, P.: High-dimensional additive modeling. Ann. Stat. **37**(6B), 3779–3821 (2009)

Müller, P., Parmigiani, G., Rice, K.: FDR and Bayesian multiple comparisons rules. In: Proceedings of Valencia / ISBA 8th World Meeting on Bayesian Statistics (2006)

Panagiotelis, A., Smith, M.: Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. J. Econom. **143**(2), 291–316 (2008)

Ravikumar, P., Lafferty, J., Liu, H., Wasserman, L.: Sparse additive models. J. R Stat. Soc. Ser. B **71**(5), 1009–1030 (2009)

Reich, B.J., Fuentes, M., Dunson, D.B.: Bayesian spatial quantile regression. J. Am. Stat. Assoc **106**(493), 6–20 (2011)

Scheipl, F., Fahrmeira, L., Kneib, T.: Spike-and-slab priors for function selection in structured additive regression models. J. Am. Stat. Assoc. **107**(500), 1518–1532 (2012)

Shively, T., Kohn, R., Wood, S.: Variable selection and function estimation in additive nonparametric regression using a data-based prior. J. Am. Stat. Assoc. **94**(447), 777–794 (1999)

Smith, M., Kohn, R.: Nonparametric regression using Bayesian variable selection. J. Econom. **75**(2), 317–343 (1996)

Sriram, K., Ramamoorthi, R., Ghosh, P.: Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. Bayesian Anal. **8**(2), 1–26 (2013)

Tan, C.: No one true path: uncovering the interplay between geography, institutions, and fractionalization in economic development. J. Appl. Econom. **25**(7), 1100–1127 (2010)

Tokdar, S., Kadane, J.: Simultaneous linear quantile regression: a semiparametric Bayesian approach. Bayesian Anal. **6**(4), 1–22 (2011)

van Dyk, D., Park, T.: Partially collapsed Gibbs samplers. J. Am. Stat. Assoc. **103**(482), 790–796 (2008)

Yau, P., Kohn, R., Wood, S.: Bayesian variable selection and model averaging in high-dimensional multinomial nonparametric regression. J. Comput. Graph. Stat. **12**(1), 23–54 (2003)

Yoshida, T.: Asymptotics for penalized spline estimators in quantile regression. Commun. Stat. Theory Methods **43**, 377 (2014)

Yu, K., Lu, Z.: Local linear additive quantile regression. Scand. J. Stat. **31**(3), 333–346 (2004)

Yu, K., Moyeed, R.A.: Bayesian quantile regression. Stat. Probab. Lett. **54**(4), 437–447 (2011)

Yue, Y., Rue, H.: Bayesian inference for additive mixed quantile regression models. Comput. Stat. Data Anal. **55**(1), 84–96 (2011)

Zhang, H., Cheng, G., Liu, Y.: Linear or nonlinear? Automatic structure discovery for partially linear models. J. Am. Stat. Assoc. **106**(495), 1099–1112 (2011)