A Bayesian Nonparametric Approach to Inference for Quantile Regression

Matthew A. TADDY
Booth School of Business, The University of Chicago, Chicago, IL 60637 (*matt.taddy@chicagobooth.edu*)

Athanasios KOTTAS
Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064
(*thanos@ams.ucsc.edu*)

We develop a Bayesian method for nonparametric model–based quantile regression. The approach involves flexible Dirichlet process mixture models for the joint distribution of the response and the covariates, with posterior inference for different quantile curves emerging from the conditional response distribution given the covariates. An extension to allow for partially observed responses leads to a novel Tobit quantile regression framework. We use simulated data sets and two data examples from the literature to illustrate the capacity of the model to uncover nonlinearities in quantile regression curves, as well as nonstandard features in the response distribution.

KEY WORDS: Dirichlet process mixture model; Markov chain Monte Carlo; Multivariate normal mixture; Tobit quantile regression.

1. INTRODUCTION

Quantile regression can be used for inference about the relationship between quantiles of the response distribution and available covariates. It offers a practically important alternative to traditional mean regression, because in general, a set of quantiles provides a more complete description of the response distribution than the mean. In many regression examples (in, e.g., 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 estimation for the standard pth quantile regression model, $y_i = \mathbf{x}_i^T\boldsymbol{\beta} + \epsilon_i$, where y_i denotes the response observations, $\mathbf{x}_i$ denotes the corresponding covariate vectors, and ϵ_i denotes the errors, which are typically assumed to be independent from a distribution [with density, say, $f_p(\cdot)$] that has pth quantile equal to 0 (see, e.g., Koenker 2005). This literature is dominated by *semiparametric* techniques, where the error density $f_p(\cdot)$ is left unspecified [apart from the restriction $\int_{-\infty}^{0} f_p(\epsilon)\,\mathrm{d}\epsilon = p$]. Thus, because 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 $\boldsymbol{\beta}$ minimize $\sum \rho_p(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$, where $\rho_p(u) = up - u1_{(-\infty,0)}(u)$; this form yields the least absolute deviations criterion for $p = 0.5$, that is, for the special case of median regression. Any inference beyond point estimation is based on asymptotic arguments or resampling methods. The classical literature includes also work that relaxes the parametric (linear) regression form for the quantile regression function (see, e.g., Horowitz and Lee 2005).

Compared 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 by Walker and Mallick (1999), Kottas and Gelfand (2001), and Hanson and Johnson (2002). This work is based on a parametric form for the median regression function and nonparametric modeling for the error distribution, using either Pólya tree or Dirichlet process (DP) priors (see, e.g., Müller and Quintana 2004 for a review of these nonparametric prior models). Regarding quantile regression, Yu and Moyeed (2001) and Tsionas (2003) discussed parametric inference based on linear regression functions and the asymmetric Laplace distribution for the errors, whereas Kottas and Krnjajić (2009) developed Bayesian semiparametric models using DP mixtures for the error distribution. Moreover, Hjort and Petrone (2007) and Hjort and Walker (2009) studied nonparametric inference for the quantile function based on DP priors and quantile pyramid priors, and briefly considered extension to semiparametric quantile regression. Finally, Chamberlain and Imbens (2003) and Dunson and Taylor (2005) proposed *semi-Bayesian* inference methods for linear quantile regression, which, in contrast to the work discussed earlier, do not involve probabilistic modeling for the response distribution.

A practical limitation of the Bayesian semiparametric models developed by Walker and Mallick (1999), Kottas and Gelfand (2001), Hanson and Johnson (2002), and Kottas and Krnjajić (2009) is that although they provide flexible shapes for the error distribution, they are based on linear quantile regression functions. Regarding inference for nonlinear quantile regression functions, Scaccia and Green (2003) modeled the conditional distribution of the response given a single continuous covariate with a discrete normal mixture with covariate-dependent weights. Moreover, Yu (2002) discussed a semi-Bayesian estimation method based on a piecewise polynomial representation for the quantile regression function again corresponding to

Journal of Business & Economic Statistics
July 2010, Vol. 28, No. 3
DOI: 10.1198/jbes.2009.07331

a single continuous covariate, but without a probability model for the error distribution. Finally, Kottas, Krnjajić, and Taddy (2007) presented an approach that combines the nonparametric prior model for the errors from Kottas and Krnjajić (2009) with a Gaussian process prior for the quantile regression function. We note that these approaches involve relatively complex Markov chain Monte Carlo (MCMC) methods for inference, and, most importantly, their extension to handle problems with more than one covariate appears to be nontrivial.

To the best of our knowledge, this article presents the first attempt at developing a model-based, fully inferential framework for Bayesian nonparametric quantile regression. We argue for the utility of Bayesian modeling, because it allows for exact and full inference for the quantile regression function, as well as for any functional of the response distribution that may be of interest. But then the flexibility of such inference under nonparametric prior models becomes attractive. We propose an approach to inference for nonparametric quantile regression founded on probabilistic modeling for the underlying unknown (random) distributions. In particular, we model the joint distribution of the response and the covariates with a flexible nonparametric mixture, then develop inference for different quantile curves based on the induced conditional distribution of the response given the covariates. The modeling framework can readily incorporate partially observed responses and in particular can be used to provide flexible inference for Tobit quantile regression. We present a method for MCMC posterior simulation, and illustrate inferences with simulated data and two data sets from the econometrics literature.

The article is organized as follows. In Sections 2 and 3 we formulate the probability model and the approach to inference for quantile regression (with technical details given in the Appendix). In Section 4 we provide applications of the modeling approach to simulated data sets and data on moral hazards from industrial chemical firms listed on the Tokyo stock exchange. In Section 5 we develop a nonparametric modeling approach for Tobit quantile regression, illustrating it with an example involving data on the labor supply of married women. We conclude with a summary in Section 6.

2. BAYESIAN MIXTURE MODELING FOR FULLY NONPARAMETRIC REGRESSION

In Section 2.1 we present the mixture modeling framework that forms the basis of our proposed approach for nonparametric quantile regression. We provide specific model formulations for categorical and/or continuous covariates in Section 2.2, and provide details regarding the choice of priors in Section 2.3.

2.1 Modeling Framework

The starting point for most existing approaches to quantile regression is the traditional additive regression framework, $y = h(\mathbf{x}) + \epsilon$, where the errors ϵ are assumed to be independent from a distribution with pth quantile 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, then the particular model must 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. This attribute of the additive formulation is shared by any approach that uses a probability model for the error distribution, regardless of the estimation method (likelihood or Bayesian). This limitation of the standard additive quantile regression framework provides the impetus for our methodology. We develop an alternative approach to inference for quantile regression that does not build on a structured regression model formulation and that yields flexible, fully nonparametric inference for quantile regression. In particular, it enables simultaneous inference for any set of quantile curves, resulting in estimates that satisfy the explicit ordering of percentiles of the response distribution.

The starting point for this approach is to consider a model for the joint distribution of the response, y, and the vector of covariates, $\mathbf{x}$, which in general comprises both continuous covariates, $\mathbf{x}_c$, and categorical covariates, $\mathbf{x}_d$, and thus $\mathbf{x} = (\mathbf{x}_c, \mathbf{x}_d)$. (We use lower-case letters for random variables as well as for their values; throughout the article, the distinction is clear from the context.) Based on the joint model for $\mathbf{z} = (y, \mathbf{x})$, inference for any set of quantile curves can be obtained from 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 $\mathbf{z}$. We use a nonparametric mixture model,

$$f(\mathbf{z}; G) = \int k(\mathbf{z}; \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta}) \tag{1}$$

for the density of $\mathbf{z}$, with a parametric kernel density, $k(\mathbf{z}; \boldsymbol{\theta})$, and a random mixing distribution G that is modeled nonparametrically. In this context, a flexible choice for the nonparametric prior for G is given by the DP, resulting in a DP mixture model for $f(\mathbf{z}; G)$.

Recall that the DP was developed by Ferguson (1973) as a prior probability model for random distributions (equivalently, distribution functions) G. A DP(α, 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 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 uses mixtures of DPs (Antoniak 1974), that is, 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 1994), which characterizes DP realizations as countable mixtures of point masses (and thus as random discrete distributions). Specifically, a random distribution G generated from DP(α, G_0) is (almost surely) of the form

$$G(\cdot) = \sum_{\ell=1}^{\infty} \omega_\ell \delta_{\vartheta_\ell}(\cdot), \tag{2}$$

where $\delta_\vartheta(\cdot)$ denotes a point mass at ϑ. The locations of the point masses, ϑ_ℓ, arise iid from G_0; the weights, ω_ℓ, arise from a *stick-breaking* mechanism based on iid draws $\{\zeta_k : k = 1, 2, \ldots\}$ from a Beta$(1, \alpha)$ distribution. In particular, $\omega_1 = \zeta_1$,

and, for each $\ell = 2, 3, \ldots$, $\omega_\ell = \zeta_\ell \prod_{k=1}^{\ell-1}(1-\zeta_k)$. Moreover, the sequences $\{\vartheta_\ell, \ell = 1, 2, \ldots\}$ and $\{\zeta_k : k = 1, 2, \ldots\}$ are independent.

The hierarchical model for the data, $\{\mathbf{z}_i = (y_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, corresponding to the DP mixture in (1), involves latent mixing parameters, $\boldsymbol{\theta}_i$, associated with each vector of response/covariate observations, $\mathbf{z}_i$, and can be written as follows:

$$\begin{aligned} \mathbf{z}_i \mid \boldsymbol{\theta}_i &\overset{\text{ind}}{\sim} k(\mathbf{z}_i; \boldsymbol{\theta}_i), \qquad i = 1, \ldots, n, \\ \boldsymbol{\theta}_i \mid G &\overset{\text{iid}}{\sim} G, \qquad i = 1, \ldots, n, \\ G \mid \alpha, \boldsymbol{\psi} &\sim \text{DP}(\alpha, G_0(\boldsymbol{\psi})). \end{aligned} \tag{3}$$

We place a gamma(a_α, b_α) prior (with mean a_α/b_α) on the DP precision parameter α, and further hyperpriors on the parameters, $\boldsymbol{\psi}$, of the base distribution G_0. The form of G_0 and of the priors for $\boldsymbol{\psi}$ depends on the choice of the mixture kernel $k(\cdot; \boldsymbol{\theta})$, as discussed in Section 2.2. Specification of the model hyperpriors is addressed in Section 2.3.

2.2 Choice of the Mixture Kernel

When the covariate vector consists of continuous covariates (as in the data example of Sec. 4.2), a natural choice for the kernel of the DP mixture model in (1) is the $(L+1)$-variate normal distribution (perhaps after transformation for the values of some of the components of $\mathbf{z}$). In this case, L is the dimension of $\mathbf{x}_c \equiv \mathbf{x}$. Thus we model the joint density for $\mathbf{z}$ through a DP mixture of multivariate normals,

$$\begin{aligned} f(\mathbf{z}; G) &= \int \text{N}_{L+1}(y, \mathbf{x}_c; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \text{d}G(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\ G \mid \alpha, \boldsymbol{\psi} &\sim \text{DP}(\alpha, G_0(\boldsymbol{\psi})) \end{aligned} \tag{4}$$

with G_0 built from independent $\text{N}_{L+1}(\mathbf{m}, \mathbf{V})$ and $\text{IWish}(\nu, \mathbf{S})$ components for the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ of the normal mixture kernel. Here $\text{IWish}(\nu, \mathbf{S})$ denotes the inverse Wishart distribution for the $(L+1) \times (L+1)$ (positive definite) matrix $\boldsymbol{\Sigma}$ with density proportional to $|\boldsymbol{\Sigma}|^{-(\nu+L+2)/2} \exp\{-0.5\,\text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\}$. We work with fixed ν and place hyperpriors on $\boldsymbol{\psi} = (\mathbf{m}, \mathbf{V}, \mathbf{S})$. In particular, we use a $\text{N}_{L+1}(\mathbf{a}_\mathbf{m}, \mathbf{B}_\mathbf{m})$ prior for $\mathbf{m}$, an $\text{IWish}(a_\mathbf{V}, \mathbf{B}_\mathbf{V})$ prior for $\mathbf{V}$, and a $\text{Wish}(a_\mathbf{S}, \mathbf{B}_\mathbf{S})$ prior for the $(L+1) \times (L+1)$ positive definite matrix $\mathbf{S}$, with density proportional to $|\mathbf{S}|^{(a_\mathbf{S}-L-2)/2} \exp\{-0.5\,\text{tr}(\mathbf{S}\mathbf{B}_\mathbf{S}^{-1})\}$ (provided that $a_\mathbf{S} \geq L+1$).

Model (4) has been applied in various settings following the work of Müller, Erkanli, and West (1996); however, the scope of inference typically has been limited to posterior point estimates, obtained through posterior predictive densities, $p(\mathbf{z}_0 \mid \text{data}) \equiv \text{E}(f(\mathbf{z}_0; G) \mid \text{data})$, where $\mathbf{z}_0$ is a specified support point. Our application to quantile regression requires the entire posterior of $f(\mathbf{z}_0; G)$ at any $\mathbf{z}_0$; thus we use a more general approach to MCMC inference (discussed in Sec. 3) that includes sampling from the posterior of G.

The DP mixture model in (4) can be extended to incorporate both continuous and categorical covariates through replacement of the multivariate normal distribution with a mixed continuous/discrete specification for the mixture kernel $k(y, \mathbf{x}_c, \mathbf{x}_d; \boldsymbol{\theta})$. One possible specification emerges from independent components for $(y, \mathbf{x}_c)$ and $\mathbf{x}_d$. The former can be a multivariate normal distribution, as in (4), whereas the latter would be assigned an appropriate discrete distribution. For instance, with a single binary covariate x_d (as in the simulated data set of Sec. 4.1), the DP mixture model is based on a mixed normal/Bernoulli kernel,

$$\begin{aligned} f(\mathbf{z}; G) &= \int \text{N}_{L+1}(y, \mathbf{x}_c; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &\qquad \times \pi^{x_d}(1-\pi)^{1-x_d}\, \text{d}G(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi), \\ G \mid \alpha, \boldsymbol{\psi} &\sim \text{DP}(\alpha, G_0(\boldsymbol{\psi})). \end{aligned} \tag{5}$$

Here G_0 comprises independent components for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and π, the former as in model (4), with the corresponding hyperpriors for $(\mathbf{m}, \mathbf{V}, \mathbf{S})$, and the latter given by a beta(a_π, b_π) distribution (with fixed shape parameters). This approach is easily extended to general categorical x_d by replacing the Bernoulli kernel with a multinomial component and a Dirichlet base distribution for the multinomial mean vector.

As a further example, consider again a single categorical covariate x_d, involving in this case counts (as in the Tobit quantile regression data example of Sec. 5.2). Then a possible form for the DP mixture arises from a mixed normal/Poisson kernel,

$$\begin{aligned} f(\mathbf{z}; G) &= \int \text{N}_{L+1}(y, \mathbf{x}_c; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \text{Po}(x_d; \lambda)\, \text{d}G(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \lambda), \\ G \mid \alpha, \boldsymbol{\psi} &\sim \text{DP}(\alpha, G_0(\boldsymbol{\psi})), \end{aligned} \tag{6}$$

where $\text{Po}(\cdot; \lambda)$ denotes the probability mass function of the Poisson distribution with mean λ. (A similar model replacing the Poisson component with a negative binomial could be considered as a robust alternative.) Again, G_0 can be built from independent components for $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and λ, where now the latter could be a gamma distribution with fixed shape parameter and random scale parameter, which is assigned a gamma hyperprior.

In general, with a vector of categorical covariates, we would need a multivariate discrete distribution for the kernel component corresponding to $\mathbf{x}_d$. In its simplest form, this discrete distribution would comprise independent components for the individual elements of $\mathbf{x}_d$. More structured versions for $k(y, \mathbf{x}_c, \mathbf{x}_d; \boldsymbol{\theta})$ can be built from a conditional distribution for either the categorical or continuous part given the other variables. Dropping the kernel parameters from the notation, in the former case, $k(y, \mathbf{x}_c, \mathbf{x}_d) = \Pr(\mathbf{x}_d \mid y, \mathbf{x}_c) k(y, \mathbf{x}_c)$, where, for example, with one binary covariate x_d, a (linear) logistic form could be used for $\Pr(x_d = 1 \mid y, \mathbf{x}_c)$. The latter setting perhaps would be more appropriate given the direction of conditioning involving the response variable. In this case, we could have $k(y, \mathbf{x}_c, \mathbf{x}_d) = k(y, \mathbf{x}_c \mid \mathbf{x}_d) \Pr(\mathbf{x}_d)$, and use a multivariate normal density for $k(y, \mathbf{x}_c \mid \mathbf{x}_d)$ with parameters that are functions of $\mathbf{x}_d$. A simpler formulation would be $k(y, \mathbf{x}_c, \mathbf{x}_d) = k(y \mid \mathbf{x}_c, \mathbf{x}_d) k(\mathbf{x}_c) \Pr(\mathbf{x}_d)$, using, say, a normal density for $k(y \mid \mathbf{x}_c, \mathbf{x}_d)$ with a mean that is a function of $\mathbf{x}_c$ and $\mathbf{x}_d$.

We note that there is nothing ad hoc about our choices of kernel or base distributions in these DP mixture models. Each independent mixture kernel component is a member of a parametric family of densities that forms a standard Bayesian model for the respective type of data. Efficient posterior simulation is aided by the analytical convenience and intuition made possible through these conditionally conjugate component choices. This modeling approach allows for flexibility through nonparametric mixing despite the choice of such convenient kernel densities and the possible assumption of independence between kernel components.

2.3 Prior Specification

Here we discuss the choice of hyperpriors for the DP mixture models of Section 2.2. We propose an approach that requires a small amount of prior information, in particular only rough prior guesses at the center of the response and covariate variables, say h_y and h_{x_l}, $l = 1, \ldots, L$, as well as at their corresponding ranges, say r_y and r_{x_l}, $l = 1, \ldots, L$. Let $\mathbf{h} = (h_y, h_{x_1}, \ldots, h_{x_L})$ and let $\mathbf{H}$ denote the $(L+1) \times (L+1)$ diagonal matrix with diagonal elements $(r_y/4)^2$ and $(r_{x_l}/4)^2$, $l = 1, \ldots, L$, which are prior estimates for the variability of the response and covariates. Then, for a default prior specification for model (4), we consider a single component in the mixture, $\mathrm{N}_{L+1}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, that is, the limiting case of the model with $\alpha \to 0^+$. Thus we effectively seek to roughly center and scale the mixture model, using prior information that identifies the subset of R^{L+1} where the data are expected to be supported. Next, based on the form of G_0 and the hyperpriors for its parameters $\boldsymbol{\psi}$, we can obtain marginal prior moments for $\boldsymbol{\mu}$, that is, $\mathrm{E}(\boldsymbol{\mu}) = \mathbf{a_m}$, and $\mathrm{Cov}(\boldsymbol{\mu}) = (a_{\mathbf{V}} - L - 2)^{-1}\mathbf{B_V} + \mathbf{B_m}$, which are matched with $\mathbf{h}$ and $\mathbf{H}$. Specifically, we take $\mathbf{a_m} = \mathbf{h}$, and, using a variance inflation factor of 2, set $\mathbf{B_m} = \mathbf{H}$ and $(a_{\mathbf{V}} - L - 2)^{-1}\mathbf{B_V} = \mathbf{H}$. We use $\mathbf{H}$ to also specify the prior for $\mathbf{S}$ through $\mathbf{H} = \mathrm{E}(\boldsymbol{\Sigma}) = (\nu - L - 2)^{-1} a_{\mathbf{S}} \mathbf{B_S}$. Finally, we choose ν, $a_{\mathbf{V}}$, and $a_{\mathbf{S}}$ to appropriately scale the hyperpriors. Note, for example, that smaller values of $(\nu - L - 2)^{-1} a_{\mathbf{S}}$ yield more dispersed priors for $\mathbf{S}$, and that $a_{\mathbf{V}} = L + 3$ is the (integer) value that yields the largest possible dispersion while ensuring finite prior expectation for $\mathbf{V}$. For the data analyses presented in Section 4, we used $\nu = a_{\mathbf{V}} = a_{\mathbf{S}} = 2(L+2)$; we also have empirically observed that this choice works well for other data sets that we have studied with model (4).

This general approach to default prior specification—placing a hyperprior on the base distribution that would be an appropriate prior for the single-component model—is applicable to other kernel forms as well. In the normal/Bernoulli mixture model of (5), the expectation of π with respect to G_0 will be a prior guess for the marginal probability of $x_d = 1$. In the case of model (6), where x_d consists of count data, the base distribution mean for λ is set to a prior guess of the mean for marginal counts.

Regarding the prior choice for the DP precision α, guidelines are available based on the role that this parameter plays with regard to the number of distinct components in the DP mixture model. Note that when marginalizing G over its DP prior, the second and third stages of model (3) collapse into a joint prior distribution for the mixing parameters $\Theta = \{\boldsymbol{\theta}_i : i = 1, \ldots, n\}$, which arises according to a particular Pólya urn scheme. Specifically, as shown by Blackwell and MacQueen (1973), conditional on the DP hyperparameters,

$$p(\Theta \mid \alpha, \boldsymbol{\psi}) = g_0(\boldsymbol{\theta}_1; \boldsymbol{\psi}) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} g_0(\boldsymbol{\theta}_i; \boldsymbol{\psi}) + \frac{1}{\alpha + i - 1} \sum_{\ell=1}^{i-1} \delta_{\boldsymbol{\theta}_\ell}(\boldsymbol{\theta}_i) \right\}, \quad (7)$$

where g_0 is the density of G_0. This expression indicates the DP-induced clustering of the mixing parameters. In particular, Θ is partitioned into n^* ($\leq n$) distinct components, where the prior distribution for n^* is controlled by α (see, e.g., Antoniak 1974; Escobar and West 1995). In practice, larger values of α yield higher prior probabilities for larger n^*; for instance, for moderately large n, $\mathrm{E}(n^* \mid \alpha) \approx \alpha \log\{(\alpha + n)/\alpha\}$, which can be averaged over the gamma prior for α to obtain an approximation to $\mathrm{E}(n^*)$.

3. POSTERIOR INFERENCE FOR QUANTILE REGRESSION

Here we develop the general approach to estimating quantile curves based on the posterior for the conditional response density implied by the mixture modeling framework of Section 2. The full posterior corresponding to the generic DP mixture model in (3) comprises the mixing distribution G, the vector of mixing parameters $\Theta = \{\boldsymbol{\theta}_i : i = 1, \ldots, n\}$, and the DP hyperparameters α and $\boldsymbol{\psi}$. Recall from Section 2.3 that the DP induces a partition of Θ into n^* distinct components, say $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, n^*$. The $\boldsymbol{\theta}_j^*$, along with configuration indicators $\mathbf{w} = (w_1, \ldots, w_n)$, defined such that $w_i = j$ if and only if $\boldsymbol{\theta}_i = \boldsymbol{\theta}_j^*$, determine Θ. Thus an equivalent representation for Θ is given by $(n^*, \{\boldsymbol{\theta}_j^* : j = 1, \ldots, n^*\}, \mathbf{w})$.

Based on the work of Antoniak (1974), the full posterior can be expressed as

$$p(G, \Theta, \alpha, \boldsymbol{\psi} \mid \text{data}) = p(G \mid \Theta, \alpha, \boldsymbol{\psi}) p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data}). \quad (8)$$

Here the distribution for $G \mid \Theta, \alpha, \boldsymbol{\psi}$ corresponds to a DP with precision parameter $\alpha + n$ and mean $\tilde{G}_0(\cdot; \Theta, \alpha, \boldsymbol{\psi})$, which is a mixed distribution with continuous mass $\alpha(\alpha + n)^{-1}$ on $G_0(\boldsymbol{\psi})$ and point masses $n_j(\alpha + n)^{-1}$ at $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, n^*$, where $n_j = |\{i : w_i = j\}|$ is the size of the jth distinct component. Moreover,

$$p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data}) \propto p(\alpha) p(\boldsymbol{\psi}) p(\Theta \mid \alpha, \boldsymbol{\psi}) \prod_{i=1}^{n} k(\mathbf{z}_i; \boldsymbol{\theta}_i)$$

is the posterior of the finite-dimensional parameter vector that results by marginalizing G over its DP prior; in particular, $p(\Theta \mid \alpha, \boldsymbol{\psi})$ is given by (7), and $p(\alpha)$ and $p(\boldsymbol{\psi})$ are the (independent) hyperpriors for α and $\boldsymbol{\psi}$.

Thus sampling from (8) involves MCMC posterior simulation from $p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data})$, followed by direct sampling from $p(G \mid \Theta, \alpha, \boldsymbol{\psi})$. A general outline of the MCMC algorithm to sample from $p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data})$ is as follows:

(a) Update each $\boldsymbol{\theta}_i$, $i = 1, \ldots, n$, by drawing from its posterior full conditional,

$$p(\boldsymbol{\theta}_i \mid \{\boldsymbol{\theta}_\ell : \ell \neq i\}, \alpha, \boldsymbol{\psi}, \text{data}) \propto p(\boldsymbol{\theta}_i \mid \{\boldsymbol{\theta}_\ell : \ell \neq i\}, \alpha, \boldsymbol{\psi}) k(\mathbf{z}_i; \boldsymbol{\theta}_i),$$

where the prior full conditional, $p(\boldsymbol{\theta}_i \mid \{\boldsymbol{\theta}_\ell : \ell \neq i\}, \alpha, \boldsymbol{\psi})$, corresponding to the joint prior in (7), is a mixed distribution with point masses $(\alpha + n - 1)^{-1}$ at the $\boldsymbol{\theta}_\ell$, for $\ell \neq i$ and continuous mass $\alpha(\alpha + n - 1)^{-1}$ on $G_0(\boldsymbol{\psi})$. Note that updating all of the $\boldsymbol{\theta}_i$ provides implicitly posterior samples for n^*, for the $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, n^*$, and for $\mathbf{w}$.

(b) To improve mixing of the MCMC algorithm (Bush and MacEachern 1996), resample each $\boldsymbol{\theta}_j^*$, $j = 1, \ldots, n^*$, from its posterior full conditional

$$p(\boldsymbol{\theta}_j^* \mid n^*, \mathbf{w}, \boldsymbol{\psi}, \text{data}) \propto g_0(\boldsymbol{\theta}_j^*; \boldsymbol{\psi}) \prod_{\{i:w_i=j\}} k(\mathbf{z}_i; \boldsymbol{\theta}_j^*).$$

(c) Update hyperparameters $\boldsymbol{\psi}$ based on their posterior full conditional

$$p(\boldsymbol{\psi} \mid n^*, \{\boldsymbol{\theta}_j^* : j = 1, \ldots, n^*\}) \propto p(\boldsymbol{\psi}) \prod_{j=1}^{n^*} g_0(\boldsymbol{\theta}_j^*; \boldsymbol{\psi}).$$

(d) Update α using, for instance, the auxiliary variable method from Escobar and West (1995).

In all but (d), the details of the MCMC algorithm depend on the choice of DP mixture kernel and the form of G_0. The Appendix provides further details for the specific DP mixture models in (4)–(6) used for the data examples of Sections 4 and 5.2. Sampling from the conditional posterior $p(G \mid \Theta, \alpha, \boldsymbol{\psi}) = \text{DP}(\alpha + n, \tilde{G}_0)$ is more generic, and we can present an approach that does not depend upon specifics of the DP mixture model choice.

Given each posterior sample $\{\Theta_b, \alpha_b, \boldsymbol{\psi}_b : b = 1, \ldots, B\}$ from $p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data})$, it is possible to sample a posterior realization G_b from $p(G \mid \Theta, \alpha, \boldsymbol{\psi})$ using the DP constructive definition in (2) with a truncation approximation (e.g., Kottas 2006). Each G_b is a discrete distribution with point masses $\{\tilde{\boldsymbol{\theta}}_{rb} : r = 1, \ldots, R_b\}$, drawn iid from $\tilde{G}_0(\cdot; \Theta_b, \alpha_b, \boldsymbol{\psi}_b)$ as defined following equation (8), and corresponding weights $\{\omega_{rb} : r = 1, \ldots, R_b\}$, generated using the stick-breaking construction based on iid Beta$(1, \alpha_b + n)$ draws, and normalized so that $\sum_{r=1}^{R_b} \omega_{rb} = 1$. 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 may depend on the particular posterior realization, although in practice it suffices to consider a common value, R. Regardless, the approximation can be specified up to any desired accuracy; for instance, it can be shown that $\text{E}(\sum_{r=1}^{R_b-1} \omega_{rb} \mid \alpha_b) = 1 - \{(\alpha_b + n)/(\alpha_b + n + 1)\}^{R_b-1}$, and thus an approach to choosing a common truncation level R would be to make $\{(n + \max_b \alpha_b)/(n + 1 + \max_b \alpha_b)\}^{R-1}$ small to any desired accuracy.

For any specific combination of response and covariate values, say, $\mathbf{z}_0 = (y_0, \mathbf{x}_0)$,

$$f(y_0, \mathbf{x}_0; G_b) = \int k(y_0, \mathbf{x}_0; \boldsymbol{\theta}) \, \text{d}G_b(\boldsymbol{\theta}) = \sum_{r=1}^{R} \omega_{rb} k(y_0, \mathbf{x}_0; \tilde{\boldsymbol{\theta}}_{rb})$$

is a realization from the posterior of the random mixture density $f(\mathbf{z}; G)$ in (1) at point $\mathbf{z} = (y_0, \mathbf{x}_0)$. Analogously, we can obtain the draw from the posterior of the marginal density $f(\mathbf{x}; G)$ at point $\mathbf{x} = \mathbf{x}_0$ by computing $f(\mathbf{x}_0; G_b) = \int k_{\mathbf{x}}(\mathbf{x}_0; \boldsymbol{\theta}) \, \text{d}G_b(\boldsymbol{\theta})$, where $k_{\mathbf{x}}(\cdot; \boldsymbol{\theta})$ denotes the marginal density for $\mathbf{x}$ corresponding to the joint kernel density $k(y, \mathbf{x}; \boldsymbol{\theta})$. For instance, under model (4), $f(\mathbf{x}_0; G_b) = \int \text{N}_L(\mathbf{x}_0; \boldsymbol{\mu}^{\mathbf{x}}, \boldsymbol{\Sigma}^{\mathbf{x}}) \, \text{d}G_b(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $(\boldsymbol{\mu}^{\mathbf{x}}, \boldsymbol{\Sigma}^{\mathbf{x}})$ are the parameters of the marginal for $\mathbf{x}$ induced by the joint $\text{N}_{L+1}(y, \mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.

Thus we obtain $f(y_0 \mid \mathbf{x}_0; G_b) = f(y_0, \mathbf{x}_0; G_b)/f(\mathbf{x}_0; G_b)$, which is a realization from the posterior of the conditional density $f(y \mid \mathbf{x}; G)$, at point $(y, \mathbf{x}) = (y_0, \mathbf{x}_0)$. 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 $f(\cdot \mid \mathbf{x}_0; G)$ for the specific covariate values $\mathbf{x}_0$. For any $0 < p < 1$, the conditional quantile $q_p(\mathbf{x}_0) \equiv q_p(\mathbf{x}_0; G)$ satisfies $\int^{q_p(\mathbf{x}_0)} f(y \mid \mathbf{x}_0; G) \, \text{d}y = F(q_p(\mathbf{x}_0) \mid \mathbf{x}_0; G) = p$. Numerical integration of the posterior realizations from the conditional response density, $f(\cdot \mid \mathbf{x}_0; G)$, yields draws from the posterior of $q_p(\mathbf{x}_0)$ for any set of percentiles that might be of interest. Alternatively, certain kernel choices allow for direct calculation of the conditional response distribution function, $F(\cdot \mid \mathbf{x}_0; G)$, precluding the need for numerical integration. Consider, for example, model (4) with the partition of kernel parameters, $\tilde{\boldsymbol{\theta}}_{rb} = (\tilde{\boldsymbol{\mu}}_{rb}, \tilde{\boldsymbol{\Sigma}}_{rb})$, into components for y and $\mathbf{x}$. Specifically, $\tilde{\boldsymbol{\mu}}_{rb}$ comprises $L \times 1$ vector $\tilde{\boldsymbol{\mu}}_{rb}^{\mathbf{x}}$ and scalar $\tilde{\mu}_{rb}^{y}$, and $\tilde{\boldsymbol{\Sigma}}_{rb}$ is a square block matrix with diagonal elements given by $L \times L$ covariance matrix $\tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}}$ and scalar variance $\tilde{\boldsymbol{\Sigma}}_{rb}^{y}$, and above and below diagonal vectors $\tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}y}$ and $\tilde{\boldsymbol{\Sigma}}_{rb}^{y\mathbf{x}}$. Then a posterior realization for $F(y_0 \mid \mathbf{x}_0; G_b)$ is calculated as $[\sum_{r=1}^{R} \omega_{rb} \text{N}_L(\mathbf{x}_0; \tilde{\boldsymbol{\mu}}_{rb}^{\mathbf{x}}, \tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}}) \Phi((y_0 - m(\mathbf{x}_0))/s(\mathbf{x}_0))]/f(\mathbf{x}_0; G_b)$, where $\Phi(\cdot)$ is the standard normal distribution function, $m(\mathbf{x}_0) = \tilde{\mu}_{rb}^{y} + \tilde{\boldsymbol{\Sigma}}_{rb}^{y\mathbf{x}} (\tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}})^{-1} (\mathbf{x}_0 - \tilde{\boldsymbol{\mu}}_{rb}^{\mathbf{x}})$, and $s^2(\mathbf{x}_0) = \tilde{\boldsymbol{\Sigma}}_{rb}^{y} - \tilde{\boldsymbol{\Sigma}}_{rb}^{y\mathbf{x}} (\tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}})^{-1} \tilde{\boldsymbol{\Sigma}}_{rb}^{\mathbf{x}y}$.

Because of the need to obtain the posterior of $f(\cdot \mid \mathbf{x}_0; G)$ [or $F(\cdot \mid \mathbf{x}_0; G)$] over a sufficiently dense grid of $\mathbf{x}_0$ values, implementation of inference becomes computationally intensive for high-dimensional covariate spaces. However, it is only ever possible to plot estimates for quantile regression functions given one- or two-variable subsets of the covariate vector (see, e.g., Figures 2 and 3 in Section 4.2). In these cases, the input grid is over a lower-dimensional space, and the computational expense is reduced. Note that inference for a marginal $p(q_p(x_{0i}) \mid \text{data})$, where $x_{0i} \in \mathbf{x}_0$, is exactly the same as inference for the full conditional quantile except based on marginal kernel densities; for example, in the normal mixture model, the joint density kernel would be $\text{N}(\cdot; \boldsymbol{\mu}^{x_i y}, \boldsymbol{\Sigma}^{x_i y})$. Moreover, if interest were focused on the posterior of conditional response densities $f(y \mid \mathbf{x}_0; G)$ (see, e.g., Figure 4), or on corresponding conditional quantiles, for a small number of specified $\mathbf{x}_0$ values, then this approach would be feasible in higher dimensions.

We can thus obtain samples from $p(q_p(\mathbf{x}_0) \mid \text{data})$ for any $\mathbf{x}_0$ and for any $0 < p < 1$. For any set of p values, working with a grid over the covariate space, we can compute simultaneous point and interval estimates for the corresponding quantile curves $q_p(\cdot; G)$. Moreover, because inference for all quantiles is based on a single density function, these estimates necessarily satisfy the ordering of percentiles of the response distribution. Thus, while providing a very flexible framework for quantile regression inference, our model-based nonparametric approach also avoids any issues with crossing quantiles. Estimated crossing quantiles may arise under classical nonparametric methods for quantile regression, and the related literature includes various techniques for addressing this problem (see, e.g., Dette and Volgushev 2008 and references therein). Moreover, the model does not rely on the additive nonparametric regression formulation and thus can uncover interactions between covariates that might influence certain quantile regression curves.

4. DATA ILLUSTRATIONS

Section 4.1 presents results from a small simulation experiment, whereas Section 4.2 considers an example involving data on moral hazard from Japanese industrial chemical firms.

4.1 Simulation Experiment

We consider synthetic data to study empirically some key aspects of the performance of the modeling approach developed in Sections 2 and 3. Although an extensive simulation study is beyond the scope of this article, these examples serve to indicate the effect of the sample size and prior choice on the resulting posterior inference under a setting in which the true regression function and response distribution are known.

Two data sets, of size $n = 200$ and $n = 2{,}000$, consist of realizations of a continuous response y, a binary covariate x_d, and a continuous covariate x_c. The data were generated such that

$$x_c \sim \mathrm{N}(0, 1), \qquad x_d \mid x_c \sim \mathrm{Bernoulli}(\Phi(x_c)),$$
$$y \mid x_c, x_d \sim \mathrm{N}(h(x_c), \sigma^2(x_d)),$$

where $\mathrm{N}(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2; $\sigma(0) = 0.25$, $\sigma(1) = 0.5$; and $h(x_c) = 0.3 + 0.4x_c + 0.5\sin(2.7x_c) + 1.1(1 + x_c^2)^{-1}$. The marginal conditional distribution for y given x_c is defined by heteroscedastic normal errors around a mean/median function (taken from Neal 1997) that is nonlinear within the likely range for x_c. Note that the data are generated through an additive error mechanism, as assumed by the majority of quantile regression models, even though this is not the setting under which our model was developed.

The model specification follows the outline of Section 2. In particular, the mixed normal/Bernoulli kernel of model (5) is assumed. The base distribution is the product of normal and inverse Wishart components for the kernel parameters related to (y, x_c), and a uniform component for π, such that $G_0(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \pi; \mathbf{m}, \mathbf{V}, \mathbf{S}) = \mathrm{N}_2(\boldsymbol{\mu}; \mathbf{m}, \mathbf{V})\,\mathrm{IWish}(\boldsymbol{\Sigma}; \nu, \mathbf{S})\,\mathrm{beta}(\pi; 1, 1)$. To study prior sensitivity, we considered two dramatically different prior specifications. Both have the same mean for $\mathbf{m}$ at $(0, 1.5)$, corresponding to the approximate mean for (x_c, y), and in both cases the required priors for variance matrixes are specified in terms of a single matrix $\mathbf{H}$ and the appropriate number of degrees of freedom, following the procedure described in Section 2.3. In the first specification, referred to as the *default* prior, $\mathbf{H}$ is set to $\mathbf{I}$, the identity matrix, such that (with variance of about 1 for x_c and y) the prior variance matrixes are the approximate expectation of the data-dependent hyperparameters proposed in Section 2.3. Moreover, under the default specification, $\pi(\alpha) = \mathrm{gamma}(2, 0.2)$. The second (*alternative*) prior specification is based on $\mathbf{H} = 10\mathbf{I}$ and $\pi(\alpha) = \mathrm{gamma}(2, 2)$. Thus the alternative specification inflates the prior expectation for variance matrixes by a factor of 10 and deflates the prior expectation for α by a factor of 10 (thus, e.g., under $n = 200$, decreasing the approximate prior expectation for n^* from about 28 to 5). These two priors represent very different (but still plausible) representations of uncertainty about the DP mixture prior parameters.

All of the results are based on MCMC samples of 80,000 draws, recorded on every tenth iteration, following a burn-in period of 20,000 iterations. Inference for the median and 90th percentile regression functions is shown in the two left columns of Figure 1, and the results are encouraging. There is a striking similarity between posterior mean and interval estimates corresponding to the two very different prior specifications. Compared with inference under the small sample, posterior means informed by the larger sample are closer to the truth and the 90% intervals are tighter, such that an increase in information leads to a corresponding increase in posterior precision. Because inference for extreme quantile functions is notoriously challenging, it is notable that accurate estimation and quantification of uncertainty holds in the case of the 90th percentile as well as for the median.

Figure 1 (far right column) also plots posterior estimates for the conditional response density $f(y|x_c = 0, x_d = 1; G)$. Again, there is a desirable uniformity among results corresponding to the different prior choices. Moreover, as the sample size increases to 2,000, the posterior mean estimates approach the true conditional density function. For the 200-point sample, inference relies heavily on a small number of local observations (i.e., response observations associated with x_c around 0) and shows posterior mean density functions that are shifted to the right of the true density. The wide posterior 90% interval reflects this uncertainty, although the default prior analysis appears to provide a better quantification of uncertainty than that based on the alternative prior. Results for the larger sample show a substantial improvement with the posterior distribution effectively capturing the truth under both prior choices. This is achieved despite the still limited amount of information provided about the entire response density corresponding to any specific conditioning.

4.2 Moral Hazard Data

Here we illustrate the methodology with real data used by Yafeh and Yoshua (2003) to investigate the relationship between shareholder concentration and several indexes for managerial moral hazard in the form of expenditure with scope for private benefit. The data set includes a variety of variables describing 185 Japanese industrial chemical firms listed on the Tokyo stock exchange. (The data set is available online through the *Economic Journal* at *http://www.res.org.uk*.) A subset of these data was also considered by Horowitz and Lee (2005) in application of their classical nonparametric estimation technique for an additive quantile regression model. As those authors did, we consider a single model proposed by Yafeh and Yoshua (2003) in which index *MH*5, consisting of general sales and administrative expenses deflated by sales, is the response y related to a four-dimensional covariate vector $\mathbf{x}$, which includes *Leverage* (ratio of debt to total assets), $\log(Assets)$, the *Age* of the firm, and *TOPTEN*, the percentage of ownership held by the 10 largest shareholders. The response and all 4 covariates are continuous and, although *Leverage* and *TOPTEN* occur over subsets of the real line, the data lies far enough from support boundaries to render the DP mixture of multivariate normals in (4) a suitable choice for the analysis.

The model is implemented using the prior specification approach outlined in Section 2.3. In the absence of genuine prior

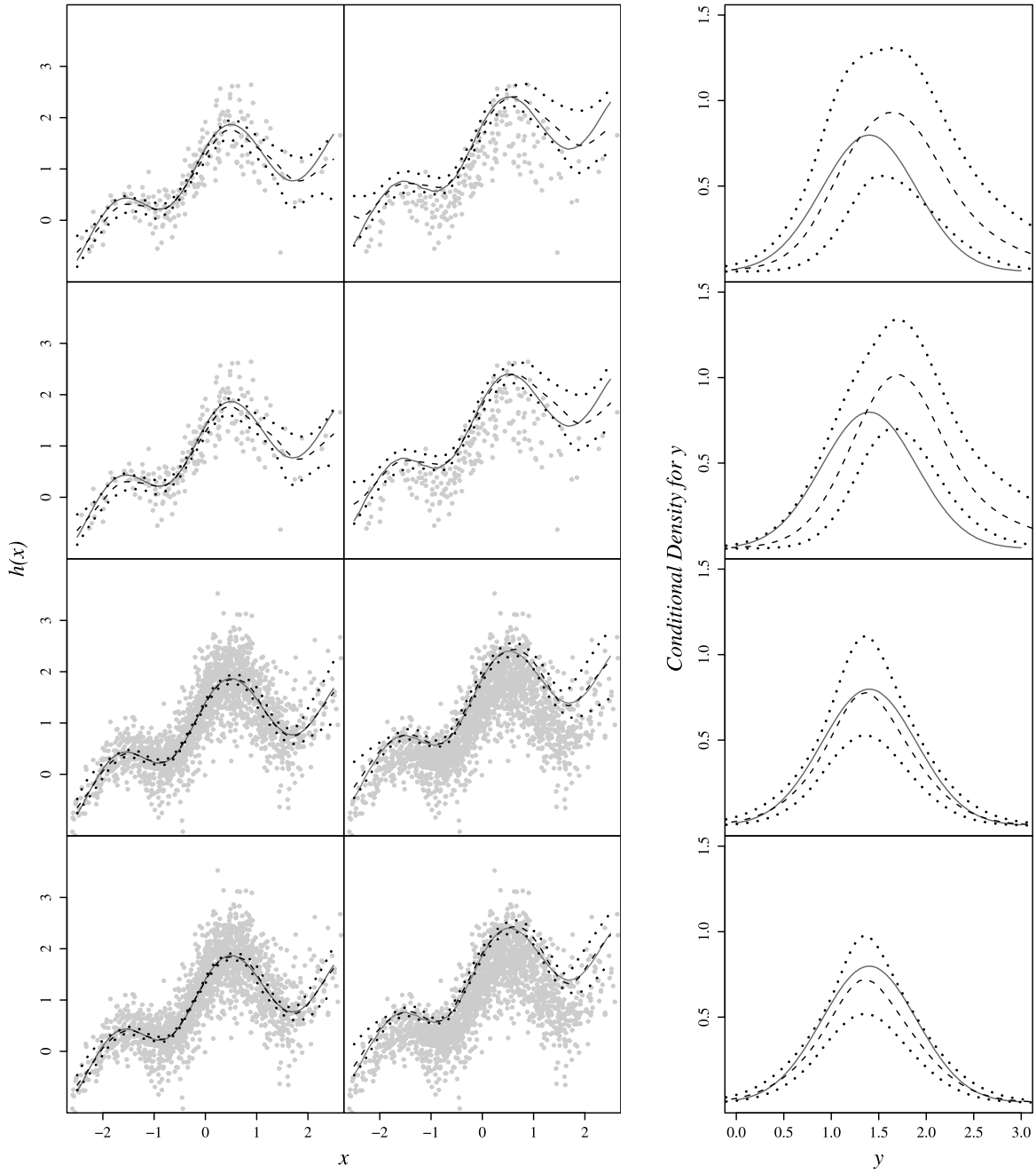

Figure 1. Simulated data. Each row shows posterior estimates for (from left to right) the marginal conditional median and 90th percentile for y given x_c ($x_c \equiv x$ in the plot labels), and the conditional density $f(y \mid x_c = 0, x_d = 1; G)$. The solid lines are truth, dashed lines are posterior mean estimates, and dotted lines contain a 90% interval. The rows correspond to the sample with $n = 200$ with the default prior (top) and with the alternative prior (second from top), and to the sample with $n = 2{,}000$ under the default prior (third from top) and the alternative prior (bottom).

information in our illustrative analysis, we take values from the data for the *prior* guesses of the center and range for the response and four covariates. The results were insensitive to reasonable changes in the prior specification; for example, doubling the observed data range for the response and covariates did not affect the posterior estimates in Figures 2–4. A gamma$(1, 0.2)$ prior is placed on the DP precision parameter α, implying $\mathrm{E}(n^*) \approx 16$. Experimentation with alternative gamma priors, yielding smaller prior estimates for the number of distinct mixture components, has resulted in essentially identical posterior inference. Results are based on an MCMC sample of 150,000 parameter draws, recorded on every tenth iteration, following a (conservative) burn-in of 50,000 iterations.

Although it is not possible to show the response quantile functions over all four variables, as discussed in Section 3, it is straightforward to obtain quantile curves for the response given any one-dimensional or two-dimensional subset of the covariates. Figure 2, plots posterior point and 90% interval estimates for the response median and 90th percentile as a function of each individual covariate. In addition, Figure 3 provides inference for the response median and 90th percentile surfaces over the two-dimensional covariate space defined by *Leverage*

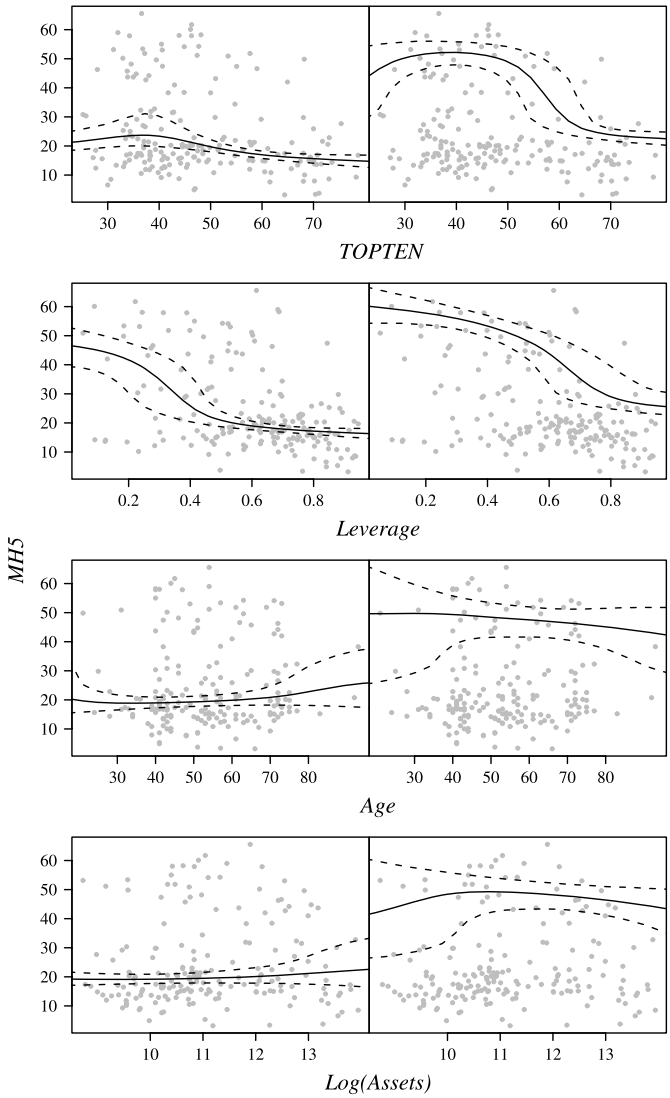

Figure 2. Moral hazard data. Posterior estimates for median regression (left column) and 90th percentile regression (right column) for *MH5* conditional on each individual covariate. The solid lines are posterior mean estimates, and dashed lines contain a 90% posterior interval. Data scatterplots are shown in gray.

and *TOPTEN*. [Note that Yafeh and Yoshua (2003) found these two covariates to be the most significant.] In particular, shown are point estimates, through the posterior mean, and a measure of the related uncertainty, through the posterior interquartile range.

These two figures indicate the capacity of the model to capture nonlinear shapes in the estimated quantile curves as well as to quantify the associated uncertainty. Figure 2 shows that the marginal relationship between each covariate and *MH5* may differ significantly depending on the quantile of interest; this is particularly clear in the contrast between median and 90th percentile curves for *MH5* conditional on *TOPTEN*. The inference results displayed in Figure 3 show an interaction between the effects of *Leverage* and *TOPTEN* in both the median and 90th percentile surfaces, suggesting that it is useful to relax the assumption of additivity over the covariate space (which forms the basis of the method in Horowitz and Lee 2005). Moreover, Figure 3 indicates that posterior uncertainty about the quantile functions is highly variable throughout the covariate space; for each quantile, regions of steep change in the quantile function correspond to significantly higher uncertainty around the function estimate. Figure 4 illustrates inference for the conditional response density $f(y \mid \mathbf{x}_0; G)$. Included are results for four values, $\mathbf{x}_0$, of the covariate vector $\mathbf{x} = (TOPTEN, Leverage, Age, \log(Assets))$, specifically, clockwise from top left, $\mathbf{x}_0 = (40, 0.3, 55, 11)$, $(35, 0.6, 55, 11)$, $(40, 0.3, 70, 13)$, and $(70, 0.8, 55, 11)$. This type of inference highlights the ability of the model to capture nonstandard distributional features such as heavy tails, skewness, and multimodality. The posterior estimates in Figure 4 clearly indicate that the response distribution changes significantly throughout the covariate space in ways that cannot be modeled with standard parametric forms. Inspection of the data scatterplots in Figure 2 makes it clear that the nonstandard features captured in the posterior estimates from the DP mixture model are driven by the data and are not simply an artifact of the flexible nonparametric prior mixture model. In this regard, note also that for the simulated data of Section 4.1, arising from a normal response distribution, the DP mixture model yields unimodal, roughly symmetric estimates for conditional response densities (refer again to Figure 1 for results under a specific combination of covariate values).

Finally, given the results of this section, it is worth drawing some comparisons between the proposed modeling approach with existing methods for quantile regression discussed in Section 1. First, given the nonlinearities in regression relationships (Figure 2) and nonstandard response density shapes (Figure 4), it is evident that the standard linear quantile regression model would be outperformed by the DP mixture model. To a greater or lesser extent, this would be the case regardless of the estimation approach—classical semiparametric, Bayesian parametric (e.g., Yu and Moyeed 2001), or Bayesian semiparametric (e.g., Hjort and Petrone 2007; Kottas and Krnjajić 2009). Classical nonparametric estimation methods likely would fare better with regard to capturing nonlinear quantile regression relationships; however, such estimation techniques are limited with respect to inference for the response distribution; for example, the results reported in Figure 4 would not be possible under these approaches. Although comparison with Bayesian nonparametric methods for nonlinear quantile regression is more relevant, there has been very little work in this direction. Moreover, as discussed in Section 1, extensions of existing work (e.g., Scaccia and Green 2003; Kottas, Krnjajić, and Taddy 2007) to incorporate more than one covariate is challenging.

5. NONPARAMETRIC TOBIT QUANTILE REGRESSION

In Section 5.1 we develop the extension to nonparametric Tobit quantile regression, and in Section 5.2 we give a data example that illustrates this extension.

5.1 The Modeling Approach

There are several regression applications that involve constrained observations for the response variable, and possibly

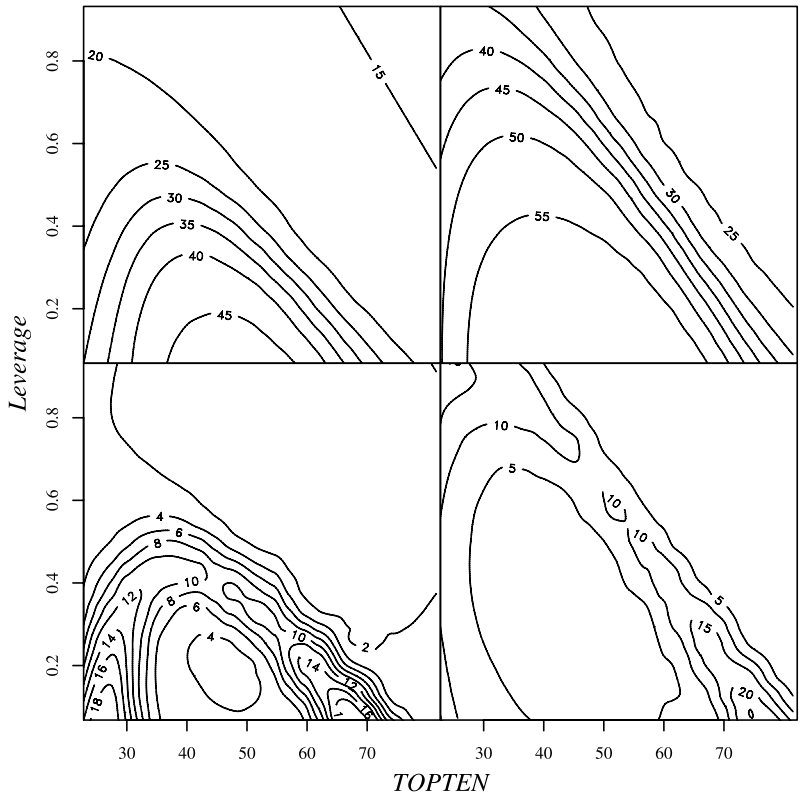

Figure 3. Moral hazard data. Posterior estimates of median surfaces (left column) and 90th percentile surfaces (right column) for *MH*5 conditional on *Leverage* and *TOPTEN*. The posterior mean is shown on the top row; the posterior interquartile range, on the bottom.

also for the covariates. For instance, different types of censoring or truncation are commonly present in survival analysis data. In econometrics applications, a standard scenario involves certain forms of partially observed responses, leading to what is typically referred to as Tobit regression models, after the work by Tobin (1958) (See, e.g., Amemiya 1984 for a thorough review of various types of Tobit models.)

The standard Tobit model is formulated through latent random variables y_i^*, which are assumed to be independent and normally distributed with mean $\mathbf{x}_i^T\boldsymbol{\beta}$ and variance σ^2. Tobit quantile regression arises by modeling a specific quantile of the latent response distribution as a function of the covariates. The covariate vectors $\mathbf{x}_i$ are observed for all subjects in the data; however, the observed responses, y_i, are constrained according to $y_i = \max\{y_i^0, y_i^*\}$, where the y_i^0 are fixed threshold points. In applications, the threshold value is typically the same for all data subjects, and thus we can set, without loss of generality, $y_i^0 = 0$ (as in our data example of Sec. 5.2). Formally, this data structure corresponds to (fixed) left censoring; however, there is a subtle difference with more traditional survival analysis applications, because in economics settings the latent variable y^* may exist only conceptually, for example, as a particular *util-*

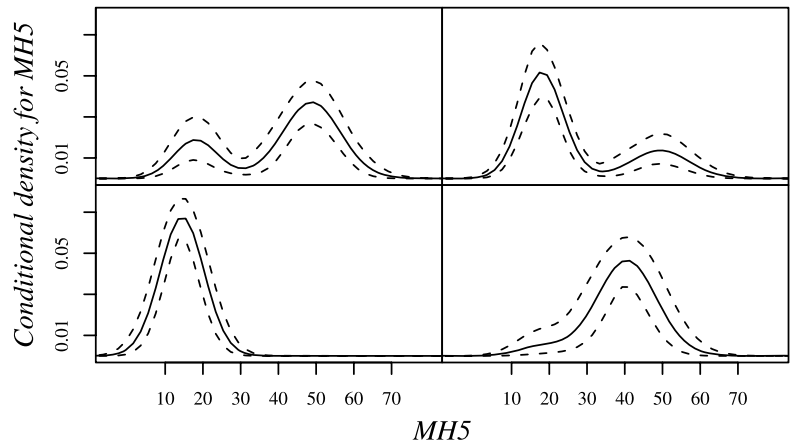

Figure 4. Moral hazard data. Posterior mean estimates (solid lines) and 90% interval estimates (dashed lines) for four conditional densities $f(y \mid \mathbf{x}_0; G)$ (see Sec. 4 for the values of $\mathbf{x}_0$).

ity functional formulated based on empirical and/or theoretical studies.

The classical semiparametric literature includes several estimation techniques for both mean and quantile regression Tobit models (see, e.g., Buchinsky and Hahn 1998 and references therein). Again, these approaches do not include probabilistic modeling for the latent response distribution and thus are limited in terms of the range of inferences that they can provide. Bayesian approaches to Tobit regression for econometrics applications appear to have focused on parametric modeling with linear regression functions. For instance, the early work of Chib (1992) developed Bayesian inference for linear Tobit regression with normal errors whereas, and, more recently, Yu and Stander (2007) have studied linear Tobit quantile regression with asymmetric Laplace errors.

The modeling framework developed in Sections 2 and 3 can be used to provide a flexible nonparametric approach to inference for Tobit quantile regression. Again, we start with a DP mixture model, $f(y^*, \mathbf{x}; G) = \int k(y^*, \mathbf{x}; \boldsymbol{\theta})\, \mathrm{d}G(\boldsymbol{\theta})$, $G \mid \alpha, \boldsymbol{\psi} \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\psi}))$, for the joint distribution of the latent response variable y^* and the vector of covariates $\mathbf{x}$. The mixture kernel can be specified following the approach outlined in Section 2.2. The first stage of the hierarchical model for the data, $(y_i, \mathbf{x}_i)$, $i = 1, \ldots, n$, is built again from conditional independence given the mixing parameters $\boldsymbol{\theta}_i$, $i = 1, \ldots, n$, but is modified with respect to (3) by replacing the (conditional) response kernel density with its corresponding distribution function for all i's with $y_i = 0$.

The analogous modifications to the MCMC method of Section 3 yield the full posterior for G, α, $\boldsymbol{\psi}$ and the $\boldsymbol{\theta}_i$, $i = 1, \ldots, n$. (Specific details for the DP mixture model used in Sec. 5.2 are provided in Section A.3.) In particular, full and exact inference for any set of quantile regression curves emerges from the posterior realizations for the conditional response density $f(\cdot \mid \mathbf{x}_0; G)$ over grid values $\mathbf{x}_0$ in the covariate space. Note that here, for any specified point $y_0 > 0$ associated with fully observed responses, $f(y_0 \mid \mathbf{x}_0; G)$ in the notation of Section 3 is given through $f(y_0 \mid y^* = y_0 > 0, \mathbf{x}_0; G)$. Thus inference for Tobit quantile regression is based on the conditional response density, given $\mathbf{x}$, arising from the underlying DP mixture $f(y^*, \mathbf{x}; G)$, conditionally also on $y^* > 0$. Moreover, using the posterior realizations for $f(y^* \mid \mathbf{x}; G)$, we can obtain the posterior for $\Pr(y^* \le 0 \mid \mathbf{x}_0; G)$. A collection of these posteriors for a set of specified $\mathbf{x}_0$ provides information on the relationship between the covariates and the censoring mechanism for the response. Because of the flexibility of the mixture model for the joint distribution of y^* and $\mathbf{x}$, the proposed modeling approach enables potentially different structures for the relationship between the response and the covariates across different quantile regression curves, as well as for the relationship between the covariates and the underlying mechanism that constrains the response. This is a practically important advantage over parametric formulations (as in, e.g., Yu and Stander 2007) that postulate a linear regression form for all of the aforementioned relationships.

5.2 Data Example

Here we consider a subset of the data on female labor supply corresponding to the University of Michigan Panel Study of Income Dynamics for year 1975. Using this data set, Mroz (1987) presented a systematic analysis of theoretical and statistical assumptions used in empirical models of the female labor supply. The sample considered by Mroz (1987) comprises 753 married white women age 30–60, 428 of whom worked at some time during 1975. The 428 fully observed responses, y_i, are given by the wife's *work* (in 100 hours) during 1975. For the remaining 325 women, the observed *work* of $y_i = 0$ corresponds to negative values for the latent *labor supply* response, y^*. The response variable can be treated as continuous (nonzero observed responses ranging from 12 to 4,950 hours). The data set includes covariate information on family income, wife's wage, education, age, number of children of different age groups, and mother's and father's educational attainment, as well as on husband's age, education, wage, and hours of work. For our illustrative analysis, we consider *number of children* as the single categorical covariate, $x_d \equiv x$. This covariate combines observations from two variables in the data set, "number of children less than 6 years old in household" and "number of children between ages 6 and 18 in household"; the observed values range from 0 to 8 children.

To model the joint distribution of the covariate and the latent labor supply response, we work with the special case of DP mixture (6) given by

$$f(y^*, x; G) = \int \mathrm{N}(y^*; \mu, \sigma^2)\, \mathrm{Po}(x; \lambda)\, \mathrm{d}G(\mu, \sigma^2, \lambda), \qquad G \mid \alpha, \boldsymbol{\psi} \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\psi})). \tag{9}$$

Here G_0 is built from independent components, specifically $\mathrm{N}(\psi_1, \psi_2)$ for μ, $\mathrm{gamma}(c, \psi_3)$ for σ^{-2}, and $\mathrm{gamma}(d, \psi_4)$ for λ, with hyperpriors placed on $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4)$.

Posterior inference under model (9) is implemented using the MCMC method detailed in Section A.3. The results reported here are based on a $\mathrm{gamma}(1, 0.2)$ prior for α, and $\mathrm{N}(10, 40)$, $\mathrm{gamma}(2, 40)$, $\mathrm{gamma}(2, 0.2)$, and $\mathrm{gamma}(3, 3)$ priors for ψ_1, ψ_2^{-1}, ψ_3, and ψ_4. The remaining parameters of G_0 are set to $c = 2$ and $d = 1$. We have experimented with increasing and decreasing the variability around α and ψ_1 and the prior expectations for ψ_2 and ψ_3, as well as with alternative specifications for ψ_4, and have not found this to affect the analysis. Results are based on an MCMC sample of 100,000 parameter draws, recorded on every fifth iteration, following a (conservative) burn-in period of 50,000 iterations.

The posterior samples for G can be used to obtain the posterior of the conditional distribution for the latent labor supply response given a specific value for the covariate number of children. Posterior estimates for the conditional densities $f(y^* \mid x; G)$, corresponding to $x = 0, \ldots, 5$ children, are shown in Figure 5. The estimated latent response densities have nonstandard shapes that change with the covariate value in a fashion that is difficult to describe with a parametric regression relationship. The peak at around 2,000 hours of work, which is seen in conditional response densities for lower numbers of children, corresponds to a traditional full-time job (50 weeks of 40 hours). The nonparametric DP mixture model exposes a density structure that would have been missed under standard parametric assumptions for the latent response distribution, such as the models developed by Chib (1992) and Yu and

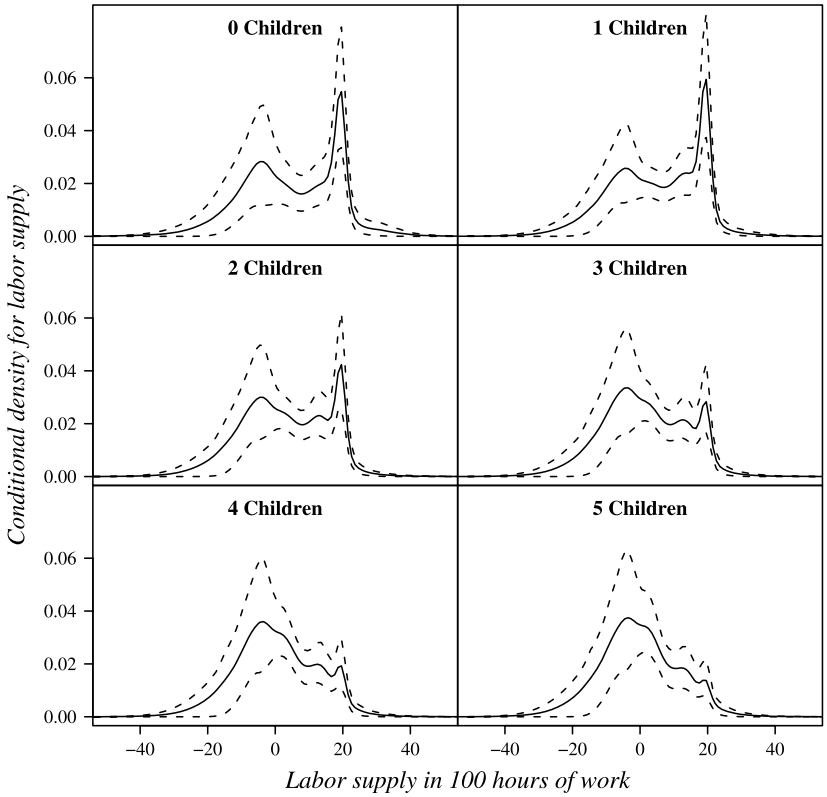

Figure 5. Female labor supply data. Posterior estimates for $f(y^* \mid x; G)$ given $x = 0, \ldots, 5$ children. The solid lines correspond to posterior mean estimates; the dashed lines, to 90% posterior interval estimates.

Stander (2007) based on normal and asymmetric Laplace distributions. In particular, the density mode corresponding to full-time labor decreases in magnitude as the number of children increases and the probability mass is redistributed in the region with $y^* < 2{,}000$ hours of work. From an economic perspective, this suggests that the main effect of an increase in offspring on labor supply is to reduce the proportion of women working full time.

Inference about conditional quantiles $q_p(x)$ for positive observed work proceeds based on posterior realizations for $\Pr(y^* < u \mid y^* > 0, x; G) = \Pr(0 < y^* < u, x; G)/\Pr(y^* > 0, x; G)$, that is, the conditional distribution function at $u > 0$, given positive observed work and given x. In particular, for any specified p and any value x for the number of children, the posterior samples $\{q_{pb}(x) : b = 1, \ldots, B = 10{,}000\}$ are obtained (with interpolation) from $p = \Pr(y^* < q_{pb}(x) \mid y^* > 0, x; G_b) = [\sum_{r=1}^{R} \omega_{rb} \operatorname{Po}(x; \tilde{\lambda}_{rb})(\Phi((q_{pb}(x) - \tilde{\mu}_{rb})/\tilde{\sigma}_{rb}) - \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb}))]/[\sum_{r=1}^{R} \omega_{rb} \operatorname{Po}(x; \tilde{\lambda}_{rb})(1 - \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb}))]$, where, following the notation of Section 3, $G_b = \{\omega_{rb}, (\tilde{\mu}_{rb}, \tilde{\sigma}^2_{rb}, \tilde{\lambda}_{rb}) : r = 1, \ldots, R\}$ is the bth posterior realization for G. As an illustration, Figure 6 shows boxplots of the posterior samples for $q_{0.5}(x)$ and $q_{0.9}(x)$. (Boxplots are constructed such that the boxes contain the interquartile sample range and the whiskers extend to the most extreme sample point that is no more than 1.5 times the interquartile range outside the central box.) Noteworthy is the different rates of decrease in the median and 90th percentile regression relationships between positive observed work and number of children. Also note that the posteriors for $q_{0.9}(x)$ at $x = 1, 2, 3, 4$ children are more concentrated than the posterior for $q_{0.9}(0)$, whereas such a difference is substantially less pronounced in the posteriors for $q_{0.5}(x)$.

Finally, as discussed in Section 5.1, also of interest might be inference for $\Pr(y^* \leq 0 \mid x; G)$, that is, the probability of 0 hours of observed work given the number of children. For any value of $x = 0, \ldots, 8$, posterior samples for this probability arise from $\Pr(y^* \leq 0 \mid x; G_b) = [\sum_{r=1}^{R} \omega_{rb} \operatorname{Po}(x; \tilde{\lambda}_{rb}) \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb})]/\sum_{r=1}^{R} \omega_{rb} \operatorname{Po}(x; \tilde{\lambda}_{rb})$, for $b = 1, \ldots, B$. Figure 7 shows boxplots

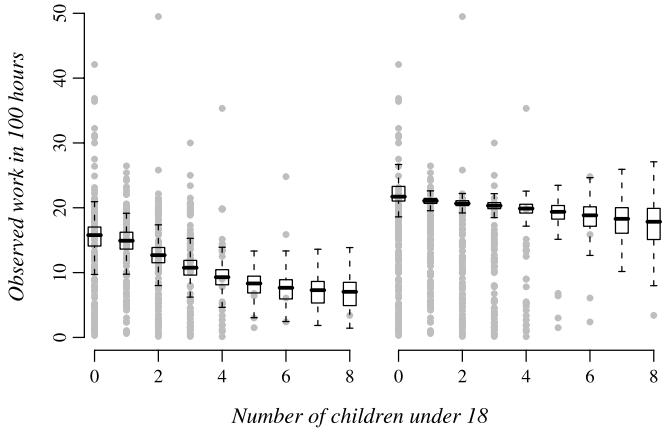

Figure 6. Female labor supply data. Posterior samples of positive observed work median (left) and 90th percentile (right) given the realized values of the covariate. The positive data observations are shown in gray.

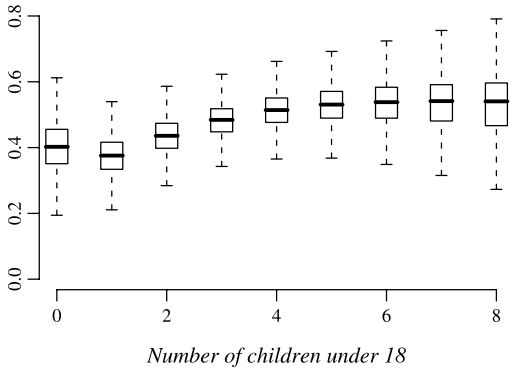

Figure 7. Female labor supply data. Posterior samples for $\Pr(y^* \le 0 \mid x; G)$.

of these posterior samples, indicating fairly similar relationships between the covariate and the censoring mechanism for the response when $x = 0, 1$ children; a noticeable increase in the probability of 0 hours of observed work with $x = 2, 3, 4$ children; and similar probabilities, albeit with increased posterior uncertainty, for $x = 5, 6, 7, 8$ children.

6. SUMMARY

We have developed a fully inferential Bayesian approach for quantile regression. The modeling approach uses flexible Dirichlet process mixtures for the joint distribution of the response and covariates, with inference for quantile curves emerging from the posterior of the induced conditional distribution of the response given the covariates. We have discussed MCMC posterior simulation methods for such inference. Our modeling framework allows incorporation of both categorical and continuous covariates, as well as partially observed responses. In particular, we have presented an approach to fully nonparametric Tobit quantile regression. Finally, we have provided illustrations with simulated and real data examples.

APPENDIX: MCMC POSTERIOR SIMULATION DETAILS

A.1 DP Mixture of Multivariate Normals Model

Here we provide details of the MCMC algorithm, outlined in Section 3, to sample from $p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data})$ under DP mixture model (4). Regarding step (a), we update each $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ using algorithm 5 from Neal (2000), which is based on Metropolis–Hastings steps with proposal distribution given by the prior full conditional of $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $p((\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \mid \{(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell) : \ell \ne i\}, \alpha, \boldsymbol{\psi})$, implied by (7). Updating all of the $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, \ldots, n$, generates a posterior realization for the partition of Θ comprising n^* distinct components $\boldsymbol{\theta}_j^* = (\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$, $j = 1, \ldots, n^*$. The Metropolis–Hastings approach to updating the $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ potentially can lead to poor mixing; however, it is straightforward to implement and, combined with step (b), which resamples the $(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$, yields an efficient MCMC method. For each $j = 1, \ldots, n^*$, the posterior full conditional for $(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$ is proportional to $g_0(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*; \boldsymbol{\psi}) \times \prod_{\{i:\, w_i = j\}} \mathrm{N}_{L+1}(\mathbf{z}_i; \boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$, and is sampled by drawing from the full conditionals for $\boldsymbol{\mu}_j^*$ and $\boldsymbol{\Sigma}_j^*$. The former is $(L+1)$-variate normal with mean vector $(\mathbf{V}^{-1} + n_j \boldsymbol{\Sigma}_j^{*-1})^{-1}(\mathbf{V}^{-1}\mathbf{m} + n_j \boldsymbol{\Sigma}_j^{*-1} \tilde{\mathbf{z}}_j)$ and covariance matrix $(\mathbf{V}^{-1} + n_j \boldsymbol{\Sigma}_j^{*-1})^{-1}$, where $n_j = |\{i : w_i = j\}|$ and $\tilde{\mathbf{z}}_j = n_j^{-1} \sum_{\{i:\, w_i = j\}} \mathbf{z}_i$. The latter is inverse Wishart with scalar parameter $\nu + n_j$ and matrix parameter $\mathbf{S} + \sum_{\{i:\, w_i = j\}} (\mathbf{z}_i - \boldsymbol{\mu}_j^*)(\mathbf{z}_i - \boldsymbol{\mu}_j^*)^T$.

Regarding the hyperparameters $\boldsymbol{\psi} = (\mathbf{m}, \mathbf{V}, \mathbf{S})$ of G_0 [step (c)], the posterior full conditional for $\mathbf{m}$ is $(L+1)$-variate normal with mean vector $(\mathbf{B}_\mathbf{m}^{-1} + n^* \mathbf{V}^{-1})^{-1}(\mathbf{B}_\mathbf{m}^{-1}\mathbf{a}_\mathbf{m} + n^* \mathbf{V}^{-1} \tilde{\boldsymbol{\mu}}^*)$, with $\tilde{\boldsymbol{\mu}}^* = n^{*-1} \sum_{j=1}^{n^*} \boldsymbol{\mu}_j^*$, and covariance matrix $(\mathbf{B}_\mathbf{m}^{-1} + n^* \mathbf{V}^{-1})^{-1}$. The full conditional for $\mathbf{V}$ is inverse Wishart with scalar parameter $a_\mathbf{V} + n^*$ and matrix parameter $\mathbf{B}_\mathbf{V} + \sum_{j=1}^{n^*} (\boldsymbol{\mu}_j^* - \mathbf{m})(\boldsymbol{\mu}_j^* - \mathbf{m})^T$, and the full conditional for $\mathbf{S}$ is given by a Wishart distribution with scalar parameter $a_\mathbf{S} + \nu n^*$ and matrix parameter $(\mathbf{B}_\mathbf{S}^{-1} + \sum_{j=1}^{n^*} \boldsymbol{\Sigma}_j^{*-1})^{-1}$.

Finally, we update the DP precision parameter α [step (d)] using the augmentation method of Escobar and West (1995). Specifically, an auxiliary variable u is introduced such that the joint density of α and u has full conditionals $p(u \mid \alpha, \text{data}) = \text{Beta}(\alpha + 1, n)$ and $p(\alpha \mid u, n^*, \text{data}) = p\,\text{gamma}(a_\alpha + n^*, b_\alpha - \log(u)) + (1 - p)\,\text{gamma}(a_\alpha + n^* - 1, b_\alpha - \log(u))$, where $p = (a_\alpha + n^* - 1)/\{n(b_\alpha - \log(u)) + a_\alpha + n^* - 1\}$.

A.2 DP Mixture Model for the Simulation Example of Section 4.1

The MCMC posterior sampling algorithm for DP mixture model (5) differs only slightly from the approach described in Section A.1. Each $\boldsymbol{\theta}_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \pi_i)$ is again updated using algorithm 5 from Neal (2000), in this case sampling each proposed $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \pi_i)$ from $p((\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \pi_i) \mid \{(\boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell, \pi_\ell) : \ell \ne i\}, \alpha, \boldsymbol{\psi})$ as implied by (7). The algorithm is again augmented by a resampling step from the posterior full conditional for each $\boldsymbol{\theta}_j^* = (\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*, \pi_j^*)$, $j = 1, \ldots, n^*$. This step is aided by noting that, given the allocation of observations to each unique kernel component, $(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$ is conditionally independent of π_j^*. Thus resampling for $(\boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)$ proceeds exactly as described in Section A.1. The full conditional for π_j^* is proportional to $g_0(\pi_j^*) \prod_{\{i:\, w_i = j\}} \pi_j^{*x_{di}} (1 - \pi_j^*)^{(1 - x_{di})}$, resulting in a beta$(a_\pi + \sum_{\{i:\, w_i = j\}} x_{di}, b_\pi + \sum_{\{i:\, w_i = j\}} (1 - x_{di}))$ distribution. Sampling for α and $\boldsymbol{\psi}$ is the same as described in Section A.1.

A.3 DP Mixture Model for Tobit Quantile Regression Example of Section 5.2

Here we describe the MCMC approach to sampling from $p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data})$ for model (9), where $\Theta = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$, with $\boldsymbol{\theta}_i = (\mu_i, \sigma_i^2, \lambda_i)$. We have

$$
\begin{aligned}
&p(\Theta, \alpha, \boldsymbol{\psi} \mid \text{data}) \\
&\quad \propto p(\alpha) p(\boldsymbol{\psi}) p(\Theta \mid \alpha, \boldsymbol{\psi}) \\
&\qquad \times \prod_{i \in I_0} \Phi(-\mu_i/\sigma_i) \prod_{i \in I_1} \mathrm{N}(y_i; \mu_i, \sigma_i^2) \prod_{i=1}^{n} \mathrm{Po}(x_i; \lambda_i),
\end{aligned}
$$

where $I_0 = \{i : y_i = 0\}$, $I_1 = \{i : y_i > 0\}$, and $p(\Theta \mid \alpha, \boldsymbol{\psi})$ is given by (7). The structure of the Metropolis–Hastings steps for the $\boldsymbol{\theta}_i$ [step (a)] is the same with the models discussed in Sections A.1

and A.2; however, when resampling, for $j = 1, \ldots, n^*$, the distinct components [step (b)] from

$$g_0(\mu_j^*, \sigma_j^{2*}, \lambda_j^*; \boldsymbol{\psi}) \prod_{\{i:w_i=j\}} \mathrm{Po}(x_i; \lambda_j^*) \times \prod_{i\in I_0\cap\{i:w_i=j\}} \Phi(-\mu_j^*/\sigma_j^*) \prod_{i\in I_1\cap\{i:w_i=j\}} \mathrm{N}(y_i; \mu_j^*, \sigma_j^{2*}),$$

the posterior full conditionals for μ_j^* and σ_j^{2*} are no longer available in a form easily drawn from. Sampling proceeds through Metropolis–Hastings steps with normal proposals for μ_j^* and gamma proposals for σ_j^{2*}. The posterior full conditional for λ_j^* is a gamma distribution with shape parameter $d + \sum_{\{i:\, w_i=j\}} x_i$ and rate parameter $\psi_4 + n_j$. The posterior full conditionals for all four hyperparameters in $\boldsymbol{\psi}$ have standard forms; specifically, they are given by a normal distribution for ψ_1 and by gamma distributions for ψ_2^{-1}, ψ_3, and ψ_4.

ACKNOWLEDGMENTS

This research is part of the first author's Ph.D. dissertation, completed at University of California, Santa Cruz, and was supported in part by the National Science Foundation (grants DMS-0505085 and DEB-0727543). The authors thank an associate editor and two referees for helpful comments that improved the presentation of the methodology.

[Received December 2007. Revised February 2009.]

REFERENCES

Amemiya, T. (1984), "Tobit Models: A Survey," *Journal of Econometrics*, 24, 3–61. [365]

Antoniak, C. E. (1974), "Mixtures of Dirichlet Processes With Applications to Nonparametric Problems," *The Annals of Statistics*, 2, 1152–1174. [358, 360]

Blackwell, D., and MacQueen, J. B. (1973), "Ferguson Distributions via Pólya Urn Schemes," *The Annals of Statistics*, 1, 353–355. [360]

Buchinsky, M., and Hahn, J. (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671. [366]

Bush, C. A., and MacEachern, S. N. (1996), "A Semiparametric Bayesian Model for Randomised Block Designs," *Biometrika*, 83, 275–285. [361]

Chamberlain, G., and Imbens, G. W. (2003), "Nonparametric Applications of Bayesian Inference," *Journal of Business & Economic Statistics*, 21, 12–18. [357]

Chib, S. (1992), "Bayes Inference in the Tobit Censored Regression Model," *Journal of Econometrics*, 51, 79–99. [366]

Dette, H., and Volgushev, S. (2008), "Non-Crossing Non-Parametric Estimates of Quantile Curves," *Journal of the Royal Statistical Society, Ser. B*, 70, 609–627. [361]

Dunson, D. B., and Taylor, J. A. (2005), "Approximate Bayesian Inference for Quantiles," *Journal of Nonparametric Statistics*, 17, 385–400. [357]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588. [360,361,368]

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230. [358]

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. [357]

Hjort, N. L., and Petrone, S. (2007), "Nonparametric Quantile Inference Using Dirichlet Processes," in *Advances in Statistical Modeling and Inference: Essays in Honor of Kjell A. Doksum*, ed. V. Nair, Singapore: World Scientific, pp. 463–492. [357,364]

Hjort, N. L., and Walker, S. G. (2009), "Quantile Pyramids for Bayesian Nonparametrics," *The Annals of Statistics*, 37, 105–131. [357]

Horowitz, J. L., and Lee, S. (2005), "Nonparametric Estimation of an Additive Quantile Regression Model," *Journal of the American Statistical Association*, 100, 1238–1249. [357,362,364]

Koenker, R. (2005), *Quantile Regression*, New York: Cambridge University Press. [357]

Kottas, A. (2006), "Nonparametric Bayesian Survival Analysis Using Mixtures of Weibull Distributions," *Journal of Statistical Planning and Inference*, 136, 578–596. [361]

Kottas, A., and Gelfand, A. E. (2001), "Bayesian Semiparametric Median Regression Modeling," *Journal of the American Statistical Association*, 96, 1458–1468. [357]

Kottas, A., and Krnjajić, M. (2009), "Bayesian Semiparametric Modeling in Quantile Regression," *Scandinavian Journal of Statistics*, 36, 297–319. [357,358,364]

Kottas, A., Krnjajić, M., and Taddy, M. (2007), "Model-Based Approaches to Nonparametric Bayesian Quantile Regression," in *Proceedings of the 2007 Joint Statistical Meetings*, pp. 1137–1148. [358,364]

Mroz, T. A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799. [366]

Müller, P., and Quintana, F. A. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110. [357]

Müller, P., Erkanli, A., and West, M. (1996), "Bayesian Curve Fitting Using Multivariate Normal Mixtures," *Biometrika*, 83, 67–79. [359]

Neal, R. M. (1997), "Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification," Technical Report CRG-TR-97-2, University of Toronto, Dept. of Computer Science. [362]

——— (2000), "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9, 249–265. [368]

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. [357,364]

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650. [358]

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, 26, 24–36. [365]

Tsionas, E. G. (2003), "Bayesian Quantile Inference," *Journal of Statistical Computation and Simulation*, 73, 659–674. [357]

Walker, S. G., and Mallick, B. K. (1999), "A Bayesian Semiparametric Accelerated Failure Time Model," *Biometrics*, 55, 477–483. [357]

Yafeh, Y., and Yoshua, O. (2003), "Large Shareholders and Banks: Who Monitors and How?" *The Economic Journal*, 113, 128–146. [362,364]

Yu, K. (2002), "Quantile Regression Using RJMCMC Algorithm," *Computational Statistics & Data Analysis*, 40, 303–315. [357]

Yu, K., and Moyeed, R. A. (2001), "Bayesian Quantile Regression," *Statistics and Probability Letters*, 54, 437–447. [357,364]

Yu, K., and Stander, J. (2007), "Bayesian Analysis of a Tobit Quantile Regression Model," *Journal of Econometrics*, 137, 260–276. [366,367]