A Nonparametric Model-based Approach to Inference for Quantile Regression

Matthew Taddy and **Athanasios Kottas**

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

In several regression applications, a different structural relationship might be anticipated for the higher or lower responses than the average responses. In such cases, quantile regression analysis can uncover important features that would likely be overlooked by traditional mean regression. We develop a Bayesian method for fully 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 distribution of the response given the covariates. Inference is implemented using a combination of posterior simulation methods for Dirichlet process mixtures. Partially observed responses can also be handled within the proposed modeling framework leading to a novel nonparametric method for Tobit quantile regression. We use two data examples from the literature to illustrate the utility of the model, in particular, its capacity to uncover non-linearities in quantile regression curves as well as non-standard features in the response distribution.

KEY WORDS: Bayesian nonparametrics; Dirichlet process priors; Markov chain Monte Carlo; Multivariate normal mixtures; Tobit quantile regression.

1. INTRODUCTION

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

There is a fairly extensive literature on classical estimation for the standard p-th quantile regression model, $y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i$, where y_i denotes the response observations, $\boldsymbol{x}_i$ the corresponding covariate vectors, and ϵ_i the errors, which are typically assumed independent from a distribution (with density, say, $f_p(\cdot)$) that has p-th quantile equal to 0 (see, e.g., Koenker, 2005). This literature is dominated by *semiparametric* techniques where the error density $f_p(\cdot)$ is left unspecified (apart from the restriction $\int_{-\infty}^{0} f_p(\epsilon)\mathrm{d}\epsilon = p$). Hence, since there is no probability model for the response distribution, point estimation for $\boldsymbol{\beta}$ proceeds by optimization of some *loss* function. For instance, under the standard setting with independent and uncensored responses, the point estimates for $\boldsymbol{\beta}$ minimize $\sum \rho_p(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})$, where $\rho_p(u) = up - u 1_{(-\infty,0)}(u)$; this form yields the least absolute deviations criterion for $p = 0.5$, i.e., 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., He, Ng and Portnoy, 1998; Horowitz and Lee, 2005).

By comparison with the existing volume of classical work, the Bayesian literature on quantile regression is relatively limited. The special case of median regression has been considered in Walker and Mallick (1999), 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, and Hanson, Branscum and Johnson, 2005, for reviews of these nonparametric prior models.) Regarding quantile regression, Yu and Moyeed (2001) and Tsionas (2003) discuss parametric inference based on linear regression functions and the asymmetric Laplace distribution for the errors; Kottas and Krnjajić (2005) develop Bayesian semiparametric models using DP mixtures for the error distribution; and Hjort and Petrone (2005) study nonparametric inference for the quantile function based on DP priors, including brief discussion of the semiparametric extension to quantile regression. Moreover, Chamberlain and Imbens (2003) and Dunson and Taylor (2005) propose *semi-Bayesian* inference methods for linear quantile regression, which, in contrast to the work discussed above, do not involve probabilistic modeling for the response distribution.

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

To our knowledge, this paper presents the first attempt to develop a model-based, fully inferential framework for Bayesian nonparametric quantile regression. We argue for the utility of Bayesian modeling, since it enables 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, which is 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, and 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 utilized to provide flexible inference for Tobit quantile regression. We present a method for MCMC posterior simulation, and illustrate inferences with two data sets that have been previously considered in the econometrics literature.

The outline of the paper is as follows. In Sections 2 and 3 we formulate the probability model and the approach to inference for quantile regression with continuous covariates. Section 4 illustrates the modeling approach with data on moral hazard from industrial chemical firms listed on the Tokyo stock exchange. In Section 5 we discuss the practically important extension of the methodology to data settings with both continuous and categorical covariates. Moreover, we develop a nonparametric modeling approach for Tobit quantile regression. An example with data on the labor supply of married women illustrates both of these extensions. Finally, Section 6 concludes with a summary.

2. BAYESIAN MIXTURE MODELING FOR FULLY NONPARAMETRIC REGRESSION

Section 2.1 presents the nonparametric mixture model that forms the basis of the proposed approach for quantile regression. Details on the choice of priors are given in Section 2.2.

2.1 The Modeling Approach

The starting point for most existing approaches to quantile regression is the traditional additive regression framework, $y = h(\boldsymbol{x}) + \epsilon$, where again the errors ϵ are assumed independent from a distribution with p-th 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, the particular model needs to be fitted separately for each corresponding p. In particular, note that estimated quantile regression functions for nearby values of p might not satisfy the explicit ordering of the corresponding percentiles, especially with small sample sizes and/or for extreme percentiles. And this attribute of the additive formulation is shared by any approach that utilizes a probability model for the error distribution, regardless of the estimation method (likelihood or Bayesian). Hence, the additive quantile regression framework seems suitable mainly for applications where interest lies in explaining one percentile of the response distribution in terms of available covariates; for instance, it offers a natural setting for median regression.

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 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 set of covariates, $\boldsymbol{x} = (x_1, ..., x_L)$. (We use lowercase letters for random variables as well as for their values, since, throughout the paper, the distinction is clear from the context.) Here, we consider covariate information on continuous variables. Section 5.1 discusses extensions of the methodology to handle applications where the covariate vector comprises continuous and/or categorical variables. Based on the joint model for $\boldsymbol{z} = (y, \boldsymbol{x})$, 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 $\boldsymbol{z}$. We employ a nonparametric mixture model, $f(\boldsymbol{z}; G) = \int k(\boldsymbol{z}; \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$, for the density of $\boldsymbol{z}$, with a parametric kernel density, $k(\boldsymbol{z}; \boldsymbol{\theta})$, 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 $f(\boldsymbol{z}; G)$.

Recall that the DP was developed by Ferguson (1973) as a prior probability model for random distributions (equivalently, distribution functions) G. A $\mathrm{DP}(\alpha, G_0)$ prior for G is defined in terms of two parameters, a parametric base distribution G_0 (the mean of the process) and a positive scalar parameter α, which can be interpreted as a precision parameter; larger values of α result in realizations G that are closer to G_0. We will write $G \sim \mathrm{DP}(\alpha, G_0)$ to indicate that a DP prior is used for the random distribution G. In fact, DP-based modeling typically utilizes mixtures of DPs (Antoniak, 1974), i.e., a more general version of the DP prior that involves hyperpriors for α and/or the parameters of G_0. The most commonly used DP definition is its constructive definition (Sethuraman, 1994), which characterizes DP realizations as countable mixtures of point masses (and thus as random discrete distributions). Specifically, a random distribution G generated from $\mathrm{DP}(\alpha, G_0)$ is (almost surely) of the form

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

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

Working with continuous covariates, a natural choice for the kernel of the DP mixture model $f(\boldsymbol{z}; G)$ is the $(L+1)$-variate normal distribution (perhaps, after transformation for the values of some of the components of $\boldsymbol{z}$). Therefore, we model the joint density for $\boldsymbol{z}$ through a DP mixture of multivariate normals,

$$f(\boldsymbol{z}; G) = \int \mathrm{N}_{L+1}(\boldsymbol{z}; \boldsymbol{\mu}, \Sigma) \mathrm{d}G(\boldsymbol{\mu}, \Sigma), \quad G \sim \mathrm{DP}(\alpha, G_0) \tag{1}$$

with G_0 built from independent $\mathrm{N}_{L+1}(\boldsymbol{m}, V)$ and $\mathrm{IWish}(\nu, S)$ components for the mean vector

$\boldsymbol{\mu}$ and the covariance matrix Σ of the normal mixture kernel. We work with random $\boldsymbol{m}$, V and S and fixed ν. Here, $\mathrm{IWish}(\nu, S)$ denotes the inverse Wishart distribution for the $(L+1)\times(L+1)$ (positive definite) matrix Σ with density proportional to $|\Sigma|^{-(\nu+L+2)/2}\exp\{-0.5\mathrm{tr}(S\Sigma^{-1})\}$.

Model (1) has been applied in various settings following the work of Müller, Erkanli and West (1996) on multivariate density estimation and curve fitting. However, the scope of inference has been typically limited to posterior point estimates, obtained through posterior predictive densities, $p(\boldsymbol{z}_0|\text{data}) \equiv \mathrm{E}(f(\boldsymbol{z}_0;G)|\text{data})$, where $\boldsymbol{z}_0$ is a specified support point and data comprises $\{\boldsymbol{z}_i = (y_i, \boldsymbol{x}_i) : i = 1, ..., n\}$. Our application to quantile regression requires the entire posterior of $f(\boldsymbol{z}_0;G)$ at any $\boldsymbol{z}_0$, and we thus employ a more general approach to MCMC inference (discussed in Section 3) that includes sampling from the posterior of G.

The hierarchical model for the data, corresponding to the DP mixture in (1), involves latent mixing parameters, $(\boldsymbol{\mu}_i, \Sigma_i)$, associated with each vector of response/covariate observations, $\boldsymbol{z}_i$, and can be written as follows

$$
\begin{aligned}
\boldsymbol{z}_i|\boldsymbol{\mu}_i, \Sigma_i &\overset{ind}{\sim} \mathrm{N}_{L+1}(\boldsymbol{z}_i;\boldsymbol{\mu}_i,\Sigma_i), \quad i=1,...,n \\
(\boldsymbol{\mu}_i, \Sigma_i)|G &\overset{iid}{\sim} G, \quad i=1,...,n \\
G \mid \alpha, \boldsymbol{\psi} &\sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\psi})).
\end{aligned} \tag{2}
$$

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

Under the modeling framework defined by (1) and (2), the discreteness of G, induced by its DP prior, is a key feature as it enables flexible shapes for the joint distribution of the response and covariates through data-driven *clustering* of the mixing parameters $(\boldsymbol{\mu}_i, \Sigma_i)$. Note, however, that we employ the DP mixture setting to model random distributions (as it was originally intended) and not as a clustering mechanism (as used, to some extent, in the more recent literature). In this regard, although it may be of methodological interest to study some of the recent extensions of the DP (e.g., Ishwaran and James, 2001; Lijoi, Mena and Prünster, 2005) as alternative priors for G, these prior models would, arguably, not lead to practical advantages over the DP with regard to the resulting inference.

2.2 Prior Specification

Here, we discuss the choice of hyperpriors for the DP mixture model of Section 2.1. 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}, $l = 1, ..., L$, as well as at their corresponding ranges, say, r_y and r_{x_l}, $l = 1, ..., L$. Let $\boldsymbol{h} = (h_y, h_{x_1}, ..., h_{x_L})$ and denote by H the $(L+1) \times (L+1)$ diagonal matrix with diagonal elements $(r_y/4)^2$ and $(r_{x_l}/4)^2$, $l = 1, ..., L$, which are prior estimates for the variability of the response and covariates. For a default specification we consider a single component in the mixture, $\mathrm{N}_{L+1}(\cdot; \boldsymbol{\mu}, \Sigma)$, i.e., the limiting case of model (2) with $\alpha \to 0^+$. Therefore, 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 $\boldsymbol{\mu}$, i.e., $\mathrm{E}(\boldsymbol{\mu}) = a_{\boldsymbol{m}}$, and $\mathrm{Cov}(\boldsymbol{\mu}) = (a_V - L - 2)^{-1} B_V + B_{\boldsymbol{m}}$, which are matched with $\boldsymbol{h}$ and H. Specifically, we take $a_{\boldsymbol{m}} = \boldsymbol{h}$, and, using a variance inflation factor of 2, set $B_{\boldsymbol{m}} = H$ and $(a_V - L - 2)^{-1} B_V = H$. We use H to specify also the prior for S through $H = \mathrm{E}(\Sigma) =$

$(\nu - L - 2)^{-1} a_S B_S$. Finally, ν, a_V, and a_S are chosen to scale appropriately the hyperpriors, e.g., note that smaller values of $(\nu - L - 2)^{-1} a_S$ yield more dispersed priors for S, and that $a_V = L + 3$ is the (integer) value that yields the largest possible dispersion while ensuring finite prior expecation for V. For the data analysis presented in Section 4, we used $\nu = a_V = a_S = 2(L + 2)$; we have also empirically observed this choice to work well for other data sets that we have studied with model (2).

Regarding the prior choice for the DP precision α, guidelines are available based on the role this parameter plays with regard to the number of distinct components in the DP mixture model. Note that, marginalizing G over its DP prior, the second and third stages of model (2) collapse into a joint prior distribution for the mixing parameters $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_i, \Sigma_i) : i = 1, ..., n\}$, which arises according to a particular Pólya urn scheme. Specifically, as shown by Blackwell and MacQueen (1973), conditional on the DP hyperparameters,

$$p(\boldsymbol{\theta} \mid \alpha, \boldsymbol{\psi}) = g_0(\boldsymbol{\mu}_1, \Sigma_1; \boldsymbol{\psi}) \prod_{i=2}^{n} \left\{ \frac{\alpha}{\alpha + i - 1} g_0(\boldsymbol{\mu}_i, \Sigma_i; \boldsymbol{\psi}) + \frac{1}{\alpha + i - 1} \sum_{\ell=1}^{i-1} \delta_{(\boldsymbol{\mu}_\ell, \Sigma_\ell)}(\boldsymbol{\mu}_i, \Sigma_i) \right\} \quad (3)$$

where g_0 is the density of G_0. This expression indicates the DP-induced clustering of the mixing parameters. In particular, $\boldsymbol{\theta}$ 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, under a gamma(a_α, b_α) prior for α (with mean a_α / b_α), a useful approximation, for moderately large n, to the prior expectation for n^* is given by $(a_\alpha / b_\alpha) \log\{1 + (n b_\alpha / a_\alpha)\}$.

3. POSTERIOR INFERENCE FOR QUANTILE REGRESSION

We describe here the approach to estimate quantile curves based on the posterior for the conditional response density $f(y|\boldsymbol{x};G)$ implied by DP mixture model (1).

The first step involves MCMC sampling from the posterior of model (2) with G marginalized over its DP prior. As discussed in Section 2.2, this marginalization yields a model with a finite-dimensional parameter vector consisting of the mixing parameters $\boldsymbol{\theta} = \{(\boldsymbol{\mu}_i, \Sigma_i) : i = 1,...,n\}$ and the DP hyperparameters α and $\boldsymbol{\psi}$.

We update each $(\boldsymbol{\mu}_i, \Sigma_i)$ using algorithm 5 from Neal (2000), which is based on Metropolis-Hastings steps with proposal distribution given by the prior full conditional of $(\boldsymbol{\mu}_i, \Sigma_i)$ implied by (3). Updating all the $(\boldsymbol{\mu}_i, \Sigma_i)$, $i = 1,...,n$, generates a posterior realization for the partition of $\boldsymbol{\theta}$ comprising n^* distinct components $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, $j = 1,...,n^*$. The $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, along with configuration indicators $\boldsymbol{w} = (w_1,...,w_n)$ defined such that $w_i = j$ if and only if $(\boldsymbol{\mu}_i, \Sigma_i) = (\boldsymbol{\mu}_j^*, \Sigma_j^*)$, determine $\boldsymbol{\theta}$. Hence, an equivalent representation for $\boldsymbol{\theta}$ is given by $(n^*, \{(\boldsymbol{\mu}_j^*, \Sigma_j^*) : j = 1,...,n^*\}, \boldsymbol{w})$. The Metropolis-Hastings approach to update the $(\boldsymbol{\mu}_i, \Sigma_i)$ can potentially lead to poor mixing. However, it is straightforward to implement and, combined with the technique from Bush and MacEachern (1996) to resample the $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$, yields an efficient MCMC method. For each $j = 1,...,n^*$, the posterior full conditional for $(\boldsymbol{\mu}_j^*, \Sigma_j^*)$ is proportional to $g_0(\boldsymbol{\mu}_j^*, \Sigma_j^*; \boldsymbol{\psi}) \prod_{\{i:w_i=j\}} \mathrm{N}_{L+1}(\boldsymbol{z}_i; \boldsymbol{\mu}_j^*, \Sigma_j^*)$, and is sampled by drawing from the full conditionals for $\boldsymbol{\mu}_j^*$ and Σ_j^*. The former is $(L+1)$-variate normal with mean vector $(V^{-1} + n_j \Sigma_j^{*-1})^{-1}(V^{-1}\boldsymbol{m} + n_j \Sigma_j^{*-1} \tilde{\boldsymbol{z}}_j)$ and covariance matrix $(V^{-1} + n_j \Sigma_j^{*-1})^{-1}$, where $n_j = |\{i : w_i = j\}|$ and $\tilde{\boldsymbol{z}}_j = n_j^{-1} \sum_{\{i:w_i=j\}} \boldsymbol{z}_i$. The latter is inverse Wishart with scalar parameter $\nu + n_j$ and matrix parameter $S + \sum_{\{i:w_i=j\}} (\boldsymbol{z}_i - \boldsymbol{\mu}_j^*)(\boldsymbol{z}_i - \boldsymbol{\mu}_j^*)^T$.

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

Next, note that, based on Antoniak (1974), the full posterior of model (2) is given by

$$p(G, \boldsymbol{\theta}, \alpha, \boldsymbol{\psi} | \text{data}) = p(G | \boldsymbol{\theta}, \alpha, \boldsymbol{\psi}) p(\boldsymbol{\theta}, \alpha, \boldsymbol{\psi} | \text{data}). \tag{4}$$

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

Hence, we can draw from the full posterior in (4) by augmenting each posterior sample from $p(\boldsymbol{\theta}, \alpha, \boldsymbol{\psi}|\text{data})$ with a draw from $p(G|\boldsymbol{\theta}, \alpha, \boldsymbol{\psi})$. The latter requires simulation from the DP with parameters given above, which we implement using the DP constructive definition (discussed in Section 2.1) with a truncation approximation (Gelfand and Kottas, 2002; Kottas, 2006). Therefore, this approach yields samples $\{G_b, \boldsymbol{\theta}_b, \alpha_b, \boldsymbol{\psi}_b : b = 1, ..., B\}$ from the full posterior (4). Each posterior realization G_b is a discrete distribution with point masses at $\boldsymbol{\vartheta}_{rb} = (\tilde{\boldsymbol{\mu}}_{rb}, \tilde{\Sigma}_{rb})$, $r = 1, ..., R_b$, drawn i.i.d. from $\tilde{G}_0(\cdot; \boldsymbol{\theta}, \alpha, \boldsymbol{\psi})$, and corresponding weights ω_{rb}, $r = 1, ..., R_b$, generated using the stick-breaking construction based on i.i.d. Beta$(1, \alpha_b)$ 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, and the approximation can be specified up to any desired accuracy (see Kottas, 2006, for a specific rule to choose R_b).

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

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

is a realization from the posterior of the random mixture density $f(y, \boldsymbol{x}; G)$ in (1) at point $(y, \boldsymbol{x}) = (y_0, \boldsymbol{x}_0)$. Analogously, we can obtain the draw from the posterior of the marginal density $f(\boldsymbol{x}; G)$ at point $\boldsymbol{x} = \boldsymbol{x}_0$ by computing $f(\boldsymbol{x}_0; G_b) = \int \mathrm{N}_L(\boldsymbol{x}_0; \boldsymbol{\mu}_{\boldsymbol{x}}, \Sigma_{\boldsymbol{x}}) \mathrm{d}G_b(\boldsymbol{\mu}, \Sigma)$, where $(\boldsymbol{\mu}_{\boldsymbol{x}}, \Sigma_{\boldsymbol{x}})$ are the parameters of the marginal for $\boldsymbol{x}$ induced by the joint $\mathrm{N}_{L+1}(y, \boldsymbol{x}; \boldsymbol{\mu}, \Sigma)$ distribution. Therefore, we obtain $f(y_0 \mid \boldsymbol{x}_0; G_b) = f(y_0, \boldsymbol{x}_0; G_b)/f(\boldsymbol{x}_0; G_b)$, which is a realization from the posterior of the conditional density $f(y \mid \boldsymbol{x}; G)$, at point $(y, \boldsymbol{x}) = (y_0, \boldsymbol{x}_0)$. Repeating over a grid in y, that covers the range of response values of interest, we obtain a posterior realization from the random conditional density function $f(\cdot \mid \boldsymbol{x}_0; G)$ for the specific covariate values $\boldsymbol{x}_0$. Note that this is a posterior realization for the entire function, obtained, of course, up to the grid approximation. Now, for any $0 < p < 1$, the conditional quantile $q_p(\boldsymbol{x}_0) \equiv q_p(\boldsymbol{x}_0; G)$ satisfies $\int^{q_p(\boldsymbol{x}_0)} f(y \mid \boldsymbol{x}_0; G)\,\mathrm{d}y = p$. Hence, using numerical integration (with interpolation) of the posterior realizations from the conditional density $f(\cdot \mid \boldsymbol{x}_0; G)$, yields draws from the posterior of $q_p(\boldsymbol{x}_0)$ for any set of percentiles that might be of interest.

Therefore, for any $\boldsymbol{x}_0$, and for any $0 < p < 1$, we obtain samples from $p(q_p(\boldsymbol{x}_0) \mid \text{data})$ that can be used to summarize the information from these conditional quantiles in any desired form. In particular, for any set of p values, working with a grid over the covariate space, we can compute point and interval estimates for the corresponding quantile curves $q_p(\cdot; G)$. Evidently, graphical depiction of these estimates for the entire curve is not feasible for problems with more than two covariates. As shown in Section 4, for such applications, one can focus on illustrations involving the quantile regression function given subsets of the covariate vector including specific choices of one or two covariates.

Because of the need to obtain the posterior of $f(\cdot \mid \boldsymbol{x}_0; G)$ over a sufficiently dense grid of $\boldsymbol{x}_0$ values, implementation of inference becomes computationally intensive for high-dimensional covariate spaces. However, if interest focuses on the posterior of conditional response densities $f(y \mid \boldsymbol{x}_0; G)$ (e.g., Figure 3), or corresponding conditional quantiles, for a small number of specified $\boldsymbol{x}_0$ values, the approach is feasible in higher dimensions. Moreover, as discussed above, for inference on conditional quantile regression functions for a small subset of the covariates (e.g., Figures 1 and 2), the input grid is over a lower dimensional space and the computational expense is reduced. Regardless, the proposed approach to inference for quantile regression is well-suited for problems with small to moderate number of covariates, and there is indeed a wide variety of such regression problems that are of interest in economics. For such settings, the methodology is very flexible as it allows both non-linear quantile curves as well as non-standard shapes for the conditional distribution of the response given the covariates. Moreover, the model does not rely on the additive nonparametric regression formulation and therefore can uncover interactions between covariates that might influence certain quantile regression curves. Finally, a key feature of the approach is that it enables simultaneous inference for any set of quantile curves of interest in a particular application.

4. DATA ILLUSTRATION

To illustrate the quantile regression methodology developed in Sections 2 and 3, we consider data used by Yafeh and Yoshua (2003) to investigate the relationship between shareholder concentration and several indices 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 was done there, we consider a single model proposed by Yafeh and Yoshua (2003) in which index *MH5*, consisting of general sales and administrative expenses deflated by sales, is the response y related to a four-dimensional covariate vector $\boldsymbol{x}$, which includes *Leverage* (ratio of debt to total assets), $\log(Assets)$, the *Age* of the firm, and *TOPTEN*, the percent of ownership held by the ten largest shareholders. The response and all four 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 multivariate normal distribution a suitable choice for the kernel of the DP mixture model in (1).

The model is implemented using the prior specification approach outlined in Section 2.2. In the absence of genuine prior 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. Results were insensitive to reasonable changes in the prior specification, e.g., doubling the observed data range for the response and covariates did not affect the posterior estimates in Figures 1 – 3. A gamma$(1, 0.2)$ prior is placed on the DP precision parameter α, implying $\mathrm{E}(n^*) \approx 18$. 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 150000 parameter draws recorded on every tenth iteration following a (conservative) burn-in of 50000 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. In Figure 1, we plot posterior point and 90% interval estimates for the response median and 90-th percentile as a function of each individual covariate. In addition, Figure 2 provides inference for the response median and 90-th percentile surfaces over the two-dimensional covariate space defined by *Leverage* 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 non-linear shapes in the estimated quantile curves as well as to quantify the associated uncertainty. The inference results displayed in Figure 2 suggest 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).

Finally, Figure 3 illustrates inference for the conditional response density $f(y \mid \boldsymbol{x}_0; G)$. Included are results for four values, $\boldsymbol{x}_0$, of the covariate vector $\boldsymbol{x} = (TOPTEN, Leverage, Age, \log(Assets))$, specifically, clockwise from top left, $\boldsymbol{x}_0 = (40, 0.3, 55, 11)$, $(35, 0.6, 55, 11)$, $(40, 0.3, 70, 13)$, and $(70, 0.8, 55, 11)$. This type of inference highlights the ability of the model to capture non-standard distributional features such as heavy tails, skewness, and multimodality. The posterior estimates in Figure 3 clearly indicate that the response distribution changes significantly throughout the covariate space in ways that can not be modeled with standard parametric forms. Inspection of the data scatterplots in Figure 1 makes it clear that the non-standard 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.

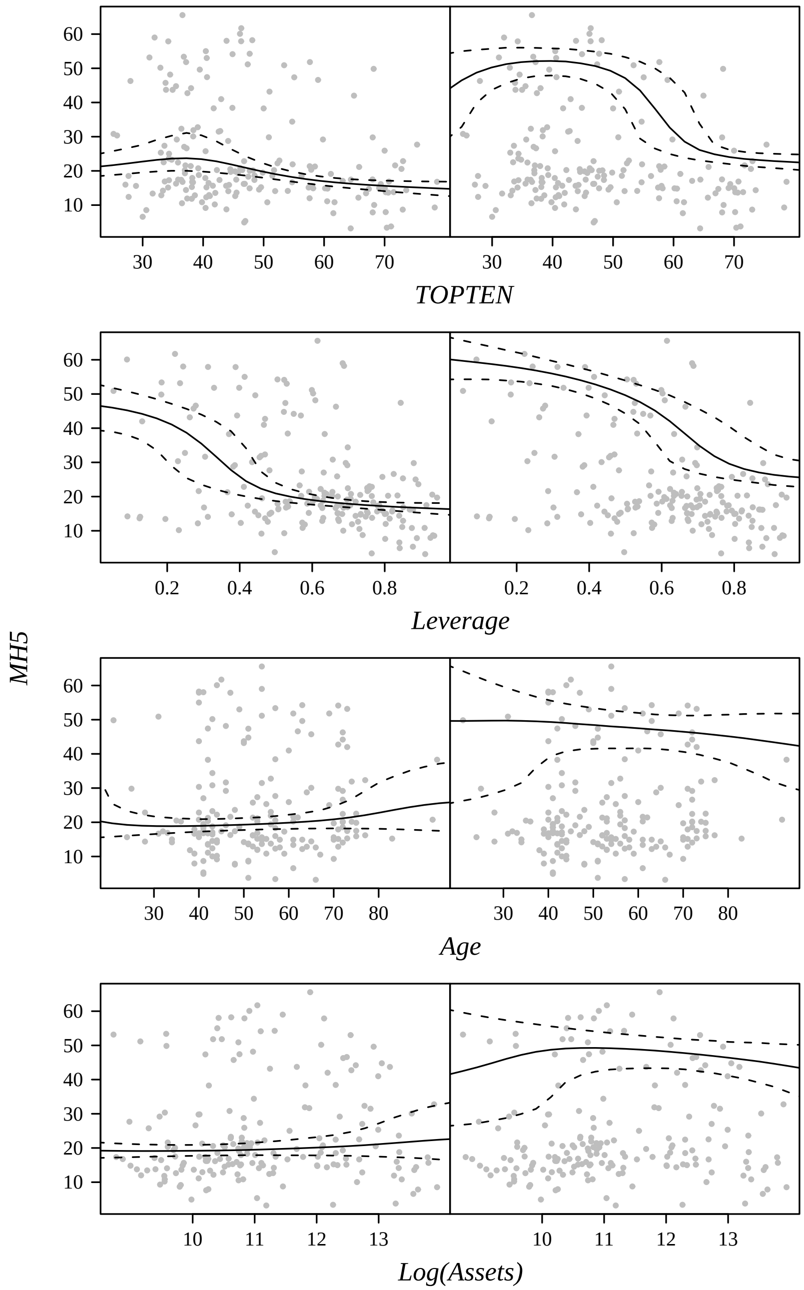

Figure 1: Moral hazard data. Posterior estimates for median regression (left column) and 90-th percentile regression (right column) for $MH5$ conditional on each individual covariate. The solid lines are posterior mean estimates and dashed lines contain a 90% posterior interval. Data scatterplots are shown in grey.

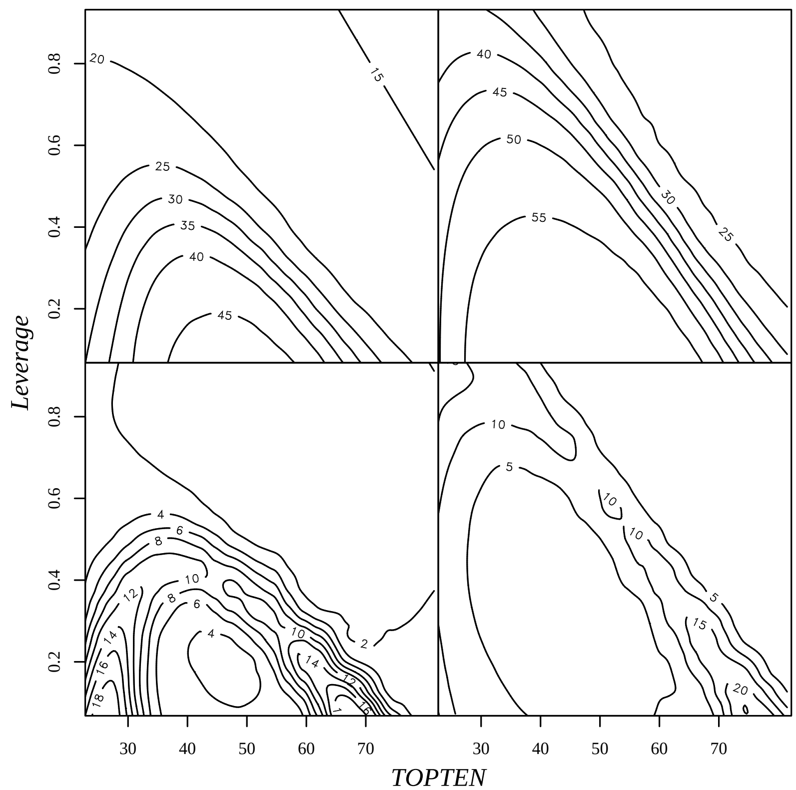

Figure 2: Moral hazard data. Posterior estimates of median surfaces (left column) and 90-th percentile surfaces (right column) for $MH5$ conditional on *Leverage* and $TOPTEN$. The posterior mean is shown on the top row and the posterior interquartile range on the bottom.

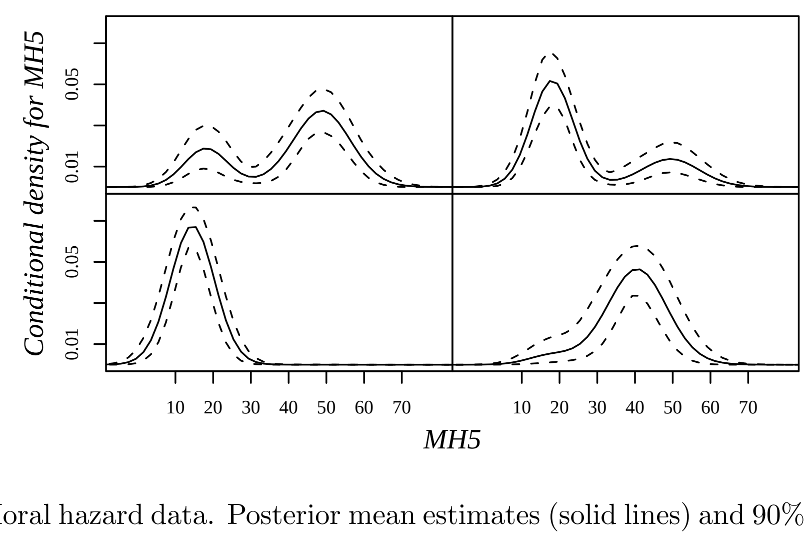

Figure 3: Moral hazard data. Posterior mean estimates (solid lines) and 90% interval estimates (dashed lines) for four conditional densities $f(y \mid \boldsymbol{x}_0; G)$ (see Section 4 for the values of $\boldsymbol{x}_0$).

5. MODEL ELABORATIONS

Section 5.1 discusses possible modifications of the model developed in Section 2 to handle problems with both categorical and continuous covariates. Section 5.2 develops the extension to nonparametric Tobit quantile regression. A data example that illustrates both of these extensions is presented in Section 5.3.

5.1 Incorporating Categorical Covariates

Here, we discuss possible extensions of the modeling framework of Section 2.1 to incorporate both continuous covariates, $\boldsymbol{x}_c$, and categorical covariates, $\boldsymbol{x}_d$, where $\boldsymbol{x} = (\boldsymbol{x}_c, \boldsymbol{x}_d)$.

A natural extension of the DP mixture model in (1) and (2) involves replacing the multivariate normal distribution with a mixed continuous/discrete specification for the mixture kernel $k(y, \boldsymbol{x}_c, \boldsymbol{x}_d; \boldsymbol{\theta})$. One possible specification emerges from independent components for $(y, \boldsymbol{x}_c)$ and $\boldsymbol{x}_d$. The former can be a multivariate normal distribution, as in Section 2.1, and the latter would be assigned an appropriate multivariate discrete distribution. In its simplest form, this discrete distribution would comprise independent components for the individual elements of $\boldsymbol{x}_d$. More generally, $k(y, \boldsymbol{x}_c, \boldsymbol{x}_d; \boldsymbol{\theta})$ 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, $k(y, \boldsymbol{x}_c, \boldsymbol{x}_d) = \Pr(\boldsymbol{x}_d \mid y, \boldsymbol{x}_c) k(y, \boldsymbol{x}_c)$, where, for example, with a single binary covariate x_d, a (linear) logistic form could be used for $\Pr(x_d = 1 \mid y, \boldsymbol{x}_c)$. The latter setting will perhaps be more appropriate given the direction of conditioning involving the response variable. In this case, we could have $k(y, \boldsymbol{x}_c, \boldsymbol{x}_d) = k(y, \boldsymbol{x}_c \mid \boldsymbol{x}_d) \Pr(\boldsymbol{x}_d)$, and use a multivariate normal density for $k(y, \boldsymbol{x}_c \mid \boldsymbol{x}_d)$ with parameters that are functions of $\boldsymbol{x}_d$. A simpler formulation would be $k(y, \boldsymbol{x}_c, \boldsymbol{x}_d) = k(y \mid \boldsymbol{x}_c, \boldsymbol{x}_d) k(\boldsymbol{x}_c) \Pr(\boldsymbol{x}_d)$, using, say, a normal density for $k(y \mid \boldsymbol{x}_c, \boldsymbol{x}_d)$ with mean that is a function of $\boldsymbol{x}_c$ and $\boldsymbol{x}_d$.

A different modeling strategy arises by retaining the multivariate normal mixture structure for the response and continuous covariate variables and placing a version of a dependent DP prior (MacEachern, 2000) on the collection of mixing distributions corresponding to the different levels of the categorical covariates. For instance, with a single binary covariate, the data vector can be decomposed into two groups, $\{\boldsymbol{z}_{ij} = (y_{ij}, \boldsymbol{x}_{ij}) : i = 1, ..., n_j\}$, $j = 1, 2$, associated with the two levels of the categorical covariate. (Here, $\boldsymbol{x}$ consists of continuous covariates only.) Then, for $j = 1, 2$, the $\boldsymbol{z}_{ij}$ are assumed to arise from the DP mixture in (1) given group-specific mixing distributions G_j. The model is completed with a dependent DP prior for (G_1, G_2), say, in the spirit of Tomlinson and Escobar (1999) and De Iorio et al. (2004), or Gelfand and Kottas (2001) if stochastic order restrictions for the categorical covariate levels are plausible, as in, e.g., treatment-control settings. In some situations this approach will result in more flexible inference, however, specifying and implementing the dependent DP mixture model becomes challenging with an increasing number of categorical covariates. Results from this line of research in the context of modeling for survival analysis problems will be reported in a future article.

5.2 Tobit Quantile Regression

There are several regression applications that involve constrained observations for the response variable, and possibly 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 independent and normally distributed with mean $\boldsymbol{x}_i^T\boldsymbol{\beta}$ 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 $\boldsymbol{x}_i$ 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 we can thus set without loss of generality $y_i^0 = 0$ (as in our data example of Section 5.3). Formally, this data structure corresponds to (fixed) left censoring. However, there is a subtle difference with more traditional survival analysis applications, since in economics settings, the latent variable y^* may exist only conceptually, e.g., as a particular *utility* functional formulated based on empirical and/or theoretical studies.

The classical semiparametric literature includes several estimation techniques for both the mean regression and quantile regression Tobit models (see, e.g., Buchinsky and Hahn, 1998, and further references therein). Again, these approaches do not include probabilistic modeling for the latent response distribution and are thus 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, more recently, Yu and Stander (2007) studied linear Tobit quantile regression with asymmetric Laplace errors.

The modeling framework developed in Sections 2 and 3 can be utilized to provide a flexible nonparametric approach to inference for Tobit quantile regression. Again, we start with a DP mixture model, $f(y^*, \boldsymbol{x}; G) = \int k(y^*, \boldsymbol{x}; \boldsymbol{\theta}) \mathrm{d}G(\boldsymbol{\theta})$, $G \mid \alpha, \boldsymbol{\psi} \sim \mathrm{DP}(\alpha, G_0(\boldsymbol{\psi}))$, for the joint distribution of the latent response variable y^* and the vector of covariates $\boldsymbol{x}$. The mixture kernel can be defined by a multivariate normal with continuous covariates (as in Section 2) or involve discrete components when categorical covariates are available (as discussed in Section 5.1). The first stage of the hierarchical model for the data, $(y_i, \boldsymbol{x}_i)$, $i = 1, ..., n$, is built again from conditional independence given the mixing parameters $\boldsymbol{\theta}_i$, $i = 1, ..., n$, but is modified with respect to (2) to replace the (conditional) response kernel density with its corresponding distribution function for all i with $y_i = 0$. The analogous modifications to the MCMC posterior simulation method of Section 3 yield the full posterior for G, α, $\boldsymbol{\psi}$ and the $\boldsymbol{\theta}_i$, $i = 1, ..., n$. We provide more details in Section 5.3 with the concrete DP mixture model used for our data illustration.

As in Section 3, full and exact inference for any set of quantile regression curves emerges from the posterior realizations for the conditional response density $f(\cdot \mid \boldsymbol{x}_0; G)$ over grid values $\boldsymbol{x}_0$ in the covariate space. Note that here, for any specified point $y_0 > 0$ associated with fully observed responses, $f(y_0 \mid \boldsymbol{x}_0; G)$ in the notation of Section 3 is given through $f(y_0 \mid y^* = y_0 > 0, \boldsymbol{x}_0; G)$. Hence, inference for Tobit quantile regression is based on the conditional response density, given $\boldsymbol{x}$, arising from the underlying DP mixture $f(y^*, \boldsymbol{x}; G)$, conditionally also on $y^* > 0$. Moreover, using the posterior realizations for $f(y^* \mid \boldsymbol{x}; G)$, we can obtain the posterior for $\Pr(y^* \leq 0 \mid \boldsymbol{x}_0; G)$. A collection of these posteriors for a set of specified $\boldsymbol{x}_0$ 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 $\boldsymbol{x}$, the proposed modeling approach enables potentially different structure 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 the relationships above.

5.3 Data Example

To illustrate the extensions developed in Sections 5.1 and 5.2, 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) presents a systematic analysis of theoretical and statistical assumptions used in empirical models of female labor supply. The sample considered by Mroz (1987) consists of 753 married white women between the ages of 30 and 60, with 428 of them working at some time during year 1975. The 428 fully observed responses, y_i, are given by the wife's *work* (in 100 hours) during year 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 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 purely illustrative analysis, we consider *number of children* as the single covariate, 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".

Although the response variable can be treated as continuous (non-zero observed responses range from 12 to 4950 hours), the covariate is a categorical variable (with values that range from 0 to 8 children). As discussed in Section 5.1, there are several possible choices for the DP mixture kernel. Here, we consider the simple setting with $k(y^*, x; \boldsymbol{\theta})$ comprising independent normal and Poisson components, a version that is sufficient for our illustrative purposes. (In other applications, a similar model based on negative binomial, rather than Poisson, components for the mixture kernel could be considered as a robust alternative.) Specifically, we work with the following DP mixture model,

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

for the latent labor supply response and number of children covariate. Here, $\mathrm{N}(\cdot; \mu, \sigma^2)$ denotes the density of the normal distribution with mean μ and variance σ^2, and $\mathrm{Po}(\cdot; \lambda)$ the probability mass function of the Poisson distribution with mean λ. Moreover, 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 $\boldsymbol{\psi} = (\psi_1, \psi_2, \psi_3, \psi_4)$.

The results reported below 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, respectively. The remaining parameters of G_0 are set to $c = 2$ and $d = 1$. We have experimented 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 100000 parameter draws recorded on every fifth iteration following a (conservative) burn-in period of 50000 iterations.

Let $p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n \mid \alpha, \boldsymbol{\psi})$ be the, analogous to (3), Pólya urn prior for the mixing parameters $\boldsymbol{\theta}_i = (\mu_i, \sigma_i^2, \lambda_i)$, that results after integrating G over its DP prior, and set $I_0 = \{i : y_i = 0\}$

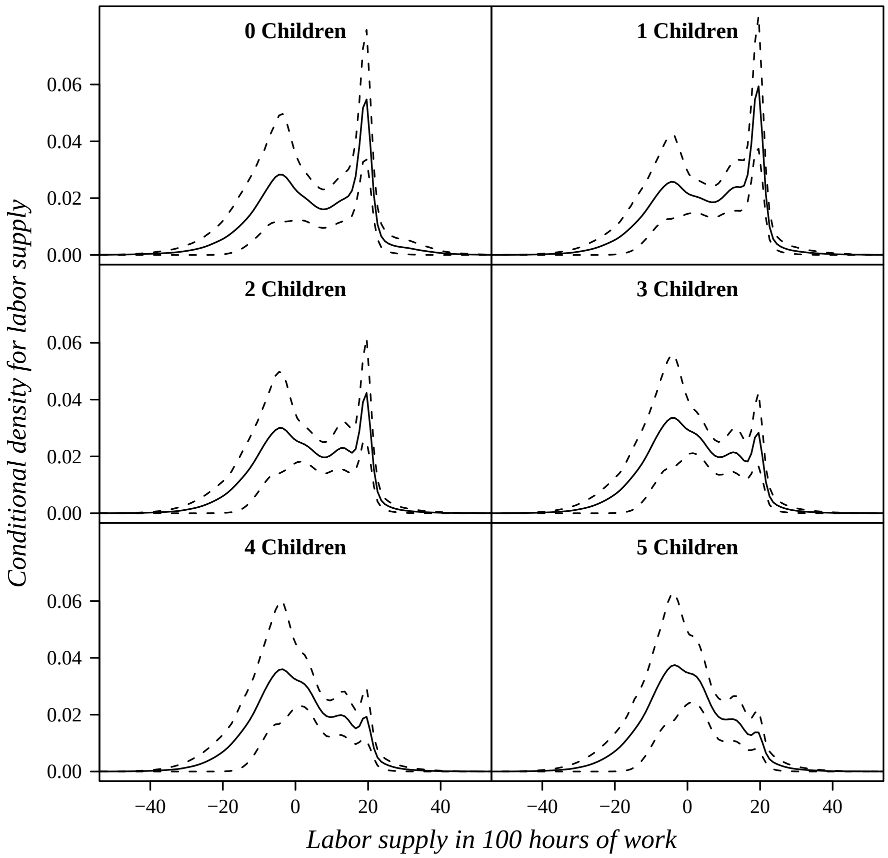

Figure 4: Female labor supply data. Posterior estimates for $f(y^* \mid x; G)$ given $x = 0, ..., 5$ children. Solid and dashed lines correspond to posterior mean and 90% posterior interval estimates, respectively.

and $I_1 = \{i : y_i > 0\}$. Then, the posterior for α, $\boldsymbol{\psi}$ and the $\boldsymbol{\theta}_i$ is proportional to

$$p(\alpha)p(\boldsymbol{\psi})p(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n \mid \alpha, \boldsymbol{\psi}) \prod_{i \in I_0} \Phi(-\mu_i/\sigma_i) \prod_{i \in I_1} \mathrm{N}(y_i; \mu_i, \sigma_i^2) \prod_{i=1}^{n} \mathrm{Po}(x_i; \lambda_i)$$

where $\Phi(\cdot)$ is the standard normal distribution function. We sample from the full posterior, that includes also G, using an MCMC method similar to the one described in Section 3. The structure of the Metropolis-Hastings steps for the $\boldsymbol{\theta}_i$ remains the same. However, when

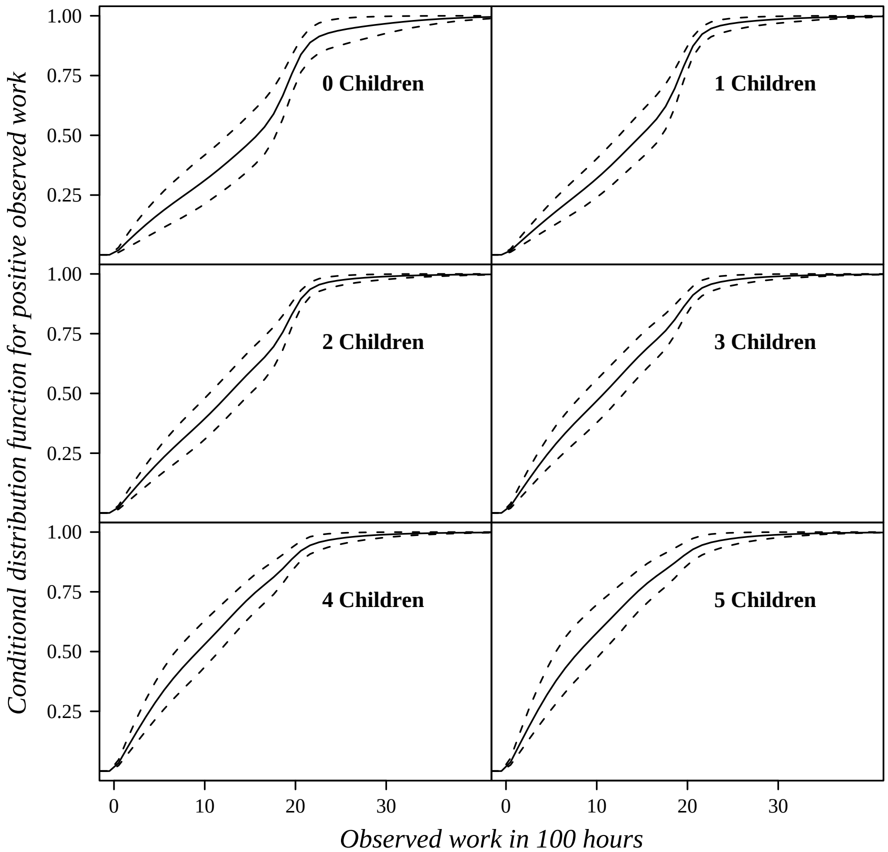

Figure 5: Female labor supply data. Posterior estimates for $\Pr(y^* < u \mid y^* > 0, x; G)$ for $x = 0, ..., 5$ children. The solid lines are posterior mean estimates and dashed lines indicate 90% posterior interval estimates.

resampling, for $j = 1, ..., n^*$, from

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

the posterior full conditionals for μ_j^* and σ_j^{2*} are no longer available in a form from which it is easy to draw. 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 DP precision parameter is, again, updated using the method from Escobar and West (1995). Finally, the posterior full conditionals for all four hyperparameters in $\boldsymbol{\psi}$ have standard forms.

As in Section 3, 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 number of children covariate. Posterior estimates for the conditional densities $f(y^* \mid x; G)$, corresponding to $x = 0, ..., 5$ children, are shown in Figure 4. The estimated latent response densities have non-standard shapes that change with the covariate value in a fashion that is difficult to describe with a parametric regression relationship. The peak around 2000 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 is exposing density structure that would have been missed under standard parametric assumptions for the latent response distribution, e.g., the models developed by Chib (1992) and Yu and Stander (2007) based on normal and asymmetric Laplace distributions, respectively.

Non-standard features are also seen in response distributions for positive observed work. This is illustrated in Figure 5, which shows posterior estimates for $\Pr(y^* < u \mid y^* > 0, x; G) = \Pr(0 < y^* < u, x; G)/\Pr(y^* > 0, x; G)$, i.e., the conditional distribution function at $u > 0$, given positive observed work and given x; results are plotted for $x = 0, ..., 5$ children. For any value of x, working with a grid of u values, posterior realizations for $\Pr(y^* < u \mid y^* > 0, x; G)$ are given by

$$\Pr(y^* < u \mid y^* > 0, x; G_b) = \frac{\sum_{r=1}^{R_b} \omega_{rb} \text{Po}(x; \tilde{\lambda}_{rb}) \left[\Phi((u - \tilde{\mu}_{rb})/\tilde{\sigma}_{rb}) - \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb}) \right]}{\sum_{r=1}^{R_b} \omega_{rb} \text{Po}(x; \tilde{\lambda}_{rb}) \left[1 - \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb}) \right]},$$

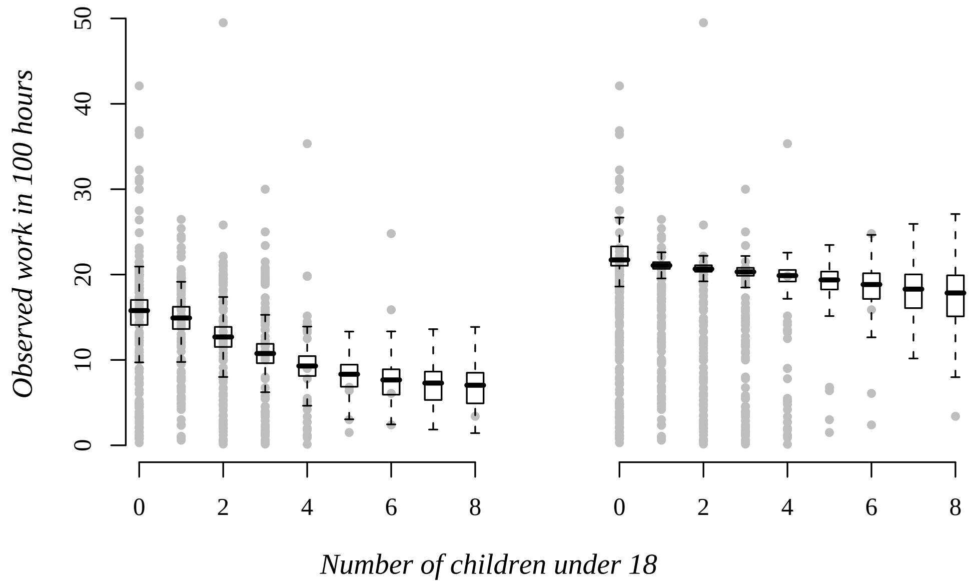

Figure 6: Female labor supply data. Posterior samples of positive observed work median (left panel) and 90-th percentile (right panel) given the realized values of the covariate. The positive data observations are shown in grey.

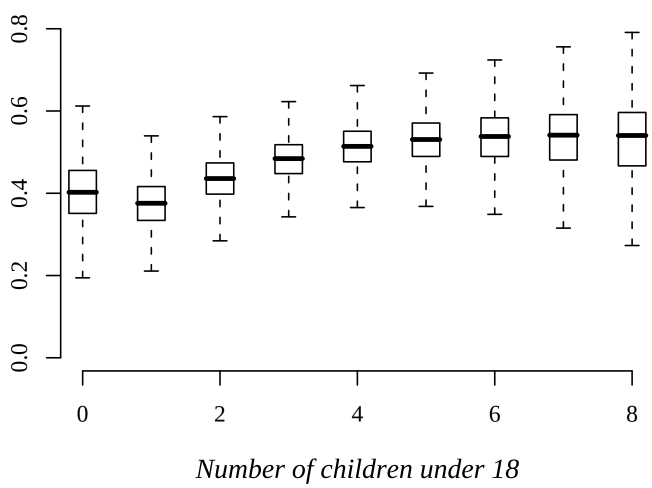

Figure 7: Female labor supply data. Posterior samples for $\Pr(y^* \leq 0 \mid x; G)$.

where, following the notation of Section 3, $G_b = \{\omega_{rb}, (\tilde{\mu}_{rb}, \tilde{\sigma}^2_{rb}, \tilde{\lambda}_{rb}) : r = 1, ..., R_b\}$ is the b-th posterior realization for G, with $b = 1, ..., B$ $(= 10000)$.

Next, inference about conditional quantiles $q_p(x)$ for positive observed work proceeds based on these posterior realizations. In particular, for any specified p and any value x for the number of children, the posterior samples $\{q_{pb}(x) : b = 1, ..., B\}$ for $q_p(x)$ are obtained (with interpolation) from $p = \Pr(y^* < q_{pb}(x) \mid y^* > 0, x; G_b)$. As an illustration, Figure 6 plots 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 rate of decrease in the median and 90-th percentile regression relationships between positive observed work and number of children. Note also 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 $q_{0.5}(x)$. This difference in the posterior uncertainty around the right tail of the conditional distribution functions at $x = 0$ and at $x = 1, 2, 3, 4$ children is also reflected in the corresponding posterior estimates in Figure 5.

Finally, as discussed in Section 5.2, of interest might be inference for $\Pr(y^* \le 0 \mid x; G)$, i.e., the probability of zero hours of observed work given the number of children. For any value of $x = 0, ..., 8$, posterior samples for this probability arise from $\Pr(y^* \le 0 \mid x; G_b) = \left[\sum_{r=1}^{R_b} \omega_{rb} \mathrm{Po}(x; \tilde{\lambda}_{rb}) \Phi(-\tilde{\mu}_{rb}/\tilde{\sigma}_{rb})\right] / \sum_{r=1}^{R_b} \omega_{rb} \mathrm{Po}(x; \tilde{\lambda}_{rb})$, for $b = 1, ..., B$. Boxplots of these posterior samples are shown in Figure 7, indicating fairly similar relationship between the covariate and the censoring mechanism for the response when $x = 0, 1$ children; a noticeable increase in the probability of zero 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 model-based, fully inferential approach for quantile regression. The modeling approach utilizes 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 presented a Markov chain Monte Carlo posterior simulation method for such inference. The modeling framework allows incorporation of both categorical and continuous covariates as well as partially observed responses. We have discussed related extensions, in the process, developing an approach to fully nonparametric Tobit quantile regression. Finally, we have provided illustrations of the methodology with two data examples.

ACKNOWLEDGMENTS

The work of the second author was supported in part by the National Science Foundation under Award DMS-0505085.

REFERENCES

Amemiya, T. (1984), "Tobit models: A survey," *Journal of Econometrics*, 24, 3-61.

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

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

Buchinsky, M., and Hahn, J. (1998), "An alternative estimator for the censored quantile regression model," *Econometrica*, 66, 653-671.

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

Chamberlain, G., and Imbens, G.W. (2003), "Nonparametric applications of Bayesian inference," *Journal of Business and Economic Statistics*, 21, 12-18.

Chib, S. (1992), "Bayes inference in the Tobit censored regression model," *Journal of Econometrics*, 51, 79-99.

De Iorio, M., Müller, P., Rosner, G.L., and MacEachern, S.N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205-215.

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

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

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

Gelfand, A.E., and Kottas, A. (2001), "Nonparametric Bayesian modeling for stochastic order," *Annals of the Institute of Statistical Mathematics*, 53, 865-876.

Gelfand, A.E., and Kottas, A. (2002), "A Computational Approach for Full Nonparametric

Bayesian Inference under Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 11, 289-305.

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

Hanson, T., Branscum, A., and Johnson, W.O. (2005), "Bayesian nonparametric modeling and data analysis: An introduction," in *Handbook of Statistics, volume 25: Bayesian Thinking, Modeling and Computation* (eds. D.K. Dey and C.R. Rao), Amsterdam: Elsevier, pp. 245-278.

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

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

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

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

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

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

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

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

Lijoi, A., Mena, R.H., and Prünster, I. (2005), "Hierarchical mixture modeling with normalized inverse-Gaussian priors," *Journal of the American Statistical Association*, 100, 1278-1291.

MacEachern, S.N. (2000), "Dependent Dirichlet Processes," Technical Report, Department of Statistics, The Ohio State University.

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.

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

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

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

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

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

Tobin, J. (1958), "Estimation of relationships for limited dependent variables," *Econometrica*, 26, 24-36.

Tomlinson, G., and Escobar, M. (1999), "Analysis of Densities," Research Report, Department of Public Health Sciences, University of Toronto.

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

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

Yafeh, Y., and Yoshua, O. (2003), "Large shareholders and banks: who monitors and how?" *The Economic Journal*, 113, 128-146.

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

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

Yu, K., and Stander, J. (2007), "Bayesian analysis of a Tobit quantile regression model," *Journal of Econometrics*, 137, 260-276.