Sankhyā : The Indian Journal of Statistics
2016, Volume 78-A, Part 1, pp. 87-104

On Bayesian Quantile Regression Using a Pseudo-joint Asymmetric Laplace Likelihood

Karthik Sriram
Indian Institute of Management Ahmedabad, Ahmedabad, India
R. V. Ramamoorthi
Michigan State University, East Lansing, USA
Pulak Ghosh
Indian Institute of Management Bangalore, Bangalore, India

Abstract

We consider a pseudo-likelihood for Bayesian estimation of multiple quantiles as a function of covariates. This arises as a simple product of multiple asymmetric Laplace densities (ALD), each corresponding to a particular quantile. The ALD has already been used in the Bayesian estimation of a single quantile. However, the pseudo-joint ALD likelihood is a way to incorporate constraints across quantiles, which cannot be done if each of the quantiles is modeled separately. Interestingly, we find that the normalized version of the likelihood turns out to be misleading. Hence, the pseudo-likelihood emerges as an alternative. In this note, we show that posterior consistency holds for the multiple quantile estimation based on such a likelihood for a nonlinear quantile regression framework and in particular for a linear quantile regression model. We demonstrate the benefits and explore potential challenges with the method through simulations.

AMS (2000) *subject classification.* Primary 62J02; Secondary 62C10.
Keywords and phrases. Asymmetric Laplace density, Bayesian quantile regression, Pseudo-likelihood.

1 Introduction

The classical linear quantile regression as studied by Koenker and Bassett (1978), involved the following problem for a given $\tau \in (0, 1)$:

$$\min_{\beta_\tau \in \mathcal{R}^d} \sum_{i=1}^{N} \rho_\tau(Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_\tau), \tag{1}$$

where $\rho_\tau(u) = u(\tau - I_{(u \leq 0)})$ with $I_{(\cdot)}$ being the indicator function and $\mathcal{R}^d$, a finite dimensional Euclidean space. Further, Koenker (1984) and Zou and

Yuan (2008) studied its extension to jointly estimating multiple quantiles (i.e. for $\tau \in \{\tau_1, \ldots, \tau_k\}$) by considering the problem:

$$\min_{(\alpha_{\tau_1},\ldots,\alpha_{\tau_k},\boldsymbol{\beta})\in\mathcal{R}^{k+d}} \sum_{j=1}^{k}\sum_{i=1}^{N} \rho_{\tau_j}(Y_i - \alpha_{\tau_j} - \mathbf{X}_i^T\boldsymbol{\beta}). \tag{2}$$

They derived asymptotic properties for the estimators under the assumption that

$$Y_i = \alpha_0 + \mathbf{X_i^T}\boldsymbol{\beta_0} + \epsilon_{\mathbf{i}}, \text{ where } \epsilon_{\mathbf{i}} \overset{\mathbf{i.i.d.}}{\sim} \mathbf{F_0}, \tag{3}$$

which automatically implies varying intercepts $\alpha_\tau = \alpha_0 + F_0^{-1}(\tau)$ across quantiles but with common slopes $\boldsymbol{\beta}_\tau = \boldsymbol{\beta}_0$.

In an attempt to provide a Bayesian framework for modeling a single quantile, Yu and Moyeed (2001) proposed the idea of assuming the asymmetric Laplace density (ALD) for the response, i.e. $Y_i \sim f_\tau(y - \mu_{\tau,i})$, where

$$f_\tau(s) = \tau(1-\tau)\cdot e^{-\rho_\tau(s)}, \ -\infty < y < \infty. \tag{4}$$

They carried out linear Bayesian quantile regression by taking $\mu_{\tau,i} = \alpha_\tau + \mathbf{X_i^T}\boldsymbol{\beta}_\tau$ and endowing priors on the intercept and slope parameters. One motivation for such an approach is the fact that obtaining a maximum likelihood estimate (MLE) based on the ALD likelihood in (4) is equivalent to solving problem (1). While this can be considered a truly Bayesian procedure only if the response truly follows ALD, Yu and Moyeed (2001) argued using empirical studies that the results are satisfactory even if ALD happens to be a misspecification. Towards a more formal justification, Sriram et al. (2013) derived posterior consistency for the linear Bayesian quantile regression parameters based on the "misspecified" ALD model, under fairly general conditions.

Our aim in this paper is to investigate the extension of the ALD based Bayesian approach to jointly modeling multiple quantiles, thus enabling the incorporation of constraints across quantiles. The ALD based approach to Bayesian quantile regression has been used in many applications (e.g. Yu et al. 2005; Yue and Rue 2011). When estimation of more than one quantile is needed, some applications (e.g. Benoit and Van den Poel 2012) have used the ALD approach by applying it separately for each quantile. However, estimating one quantile at a time does not allow for applying model constraints (such as that of common slopes or monotonicity) to be imposed across different quantiles. Therefore, a natural question arises as to whether a working-likelihood motivated by ALD can be used for Bayesian joint estimation of multiple quantiles.

A natural extension to handle multiple quantiles would be to consider a working-likelihood obtained by taking the product of ALD densities (as in (4)) across different values of τ, i.e. $0 < \tau_1 < \tau_2 < \cdots < \tau_K < 1$. While it is natural to think of normalizing such a likelihood so that it integrates to 1, we will see later (Section 3) that such an approach could be misleading. Accordingly, denoting $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_K)$ and $\boldsymbol{\mu}_i = (\mu_{\tau_1,i}, \mu_{\tau_2,i}, \ldots, \mu_{\tau_K,i})$, we consider the following pseudo-likelihood for Y_i without the normalizing constant:

$$p_{\boldsymbol{\tau},\boldsymbol{\mu_i}}(y) = \prod_{j=1}^{K} f_{\tau_j}(y - \mu_{\tau_j,i}). \tag{5}$$

The conditional τ_j^{th} quantile of Y_i can then be modeled as $\mu_{\tau_j,i} = Q_{\tau_j}(\mathbf{X}_i)$, which could be a linear or even a non-linear model. Bayesian analysis can then be carried out by endowing a prior on the space of $(Q_{\tau_1}, Q_{\tau_2}, \ldots, Q_{\tau_K})$. For a true Bayesian inference, the "true" data generating mechanism needs to closely resemble the working-likelihood. However, since the likelihood (5) is essentially a misspecification of the true likelihood, the mathematical justification needs to be asymptotic in nature.

In this note, we show that posterior consistency holds for the joint quantile estimation based on (5), for a nonlinear quantile modeling framework and in particular for a linear model. While this result is motivated from Sriram et al. (2013), it does not follow directly from there and is more general because it handles (a) a non-linear model formulation and (b) the case of jointly estimating multiple quantiles. In particular, for $K = 1$, the result would support the extension of the approach in Yu and Moyeed (2001) to a non-linear quantile model.

We further demonstrate the benefits and explore potential challenges with the joint modeling approach through simulations. The first simulation in Section 4.1 is an example where the quantile function is non-linear in the parameters and demonstrates how the joint modeling could be more beneficial than modeling each quantile separately. The pseudo likelihood (5) is essentially a misspecification of the true underlying distribution of the data. It is known that under misspecification, the asymptotic coverage of a 95 % credible interval cannot be expected to be equal to 95 % (see Kleijn and van der Vaart 2012). The second simulation in Section 4.2 investigates the coverage property of the credible intervals obtained from this approach. In particular, it investigates the credible intervals of the common slope parameter as we increase the number of quantiles jointly estimated.

The problem of simultaneously modeling multiple quantiles has recently led to the development of classical as well as Bayesian approaches. Key classical approaches include He (1997), Wu and Liu (2009), Chernozhukov et al. (2010), & Dette and Volgushev (2008). Key Bayesian approaches include Tokdar and Kadane (2012), Reich et al. (2011), Taddy and Kottas (2010), Dunson and Taylor (2005), & Yang and He (2012).

The ALD based working-likelihood presents itself as an interesting alternative worthy of investigation since it can accomodate linear as well as non-linear model formulations along with any type of prior specification. In Section 2, we present the main result followed by some examples and observations. Section 3 demonstrates the issue with using a normalized likelihood versus a pseudo likelihood. Section 4 presents the simulations and the last section concludes.

2 Main Result

In this section, we present the posterior consistency property for a non-linear Bayesian quantile regression model based on the misspecified working likelihood (5). Our results extend the work of Sriram et al. (2013) in two ways: (i) to the case of multiple quantiles and (ii) to apply to a class of functions that is more general than linear quantile functions. We hasten to note that although the working likelihood (5) is essentially a product of individual ALDs, the posterior consistency property (even for the linear quantile case) does not directly follow from Sriram et al. (2013). This is because the prior on the quantile parameters need not be multiplicative across quantiles. This is the case for example when a slope parameter may be fixed across quantiles, or when a specific monotonicity-ensuring prior is used. In such cases, the posterior likelihood may not necessarily split into a product of posteriors from individual ALDs.

Let $Y_{1:n} := (Y_1, Y_2, \cdots, Y_n)$ be a vector of n independent but non-identically distributed responses ($i.n.i.d$). We consider a univariate non-random covariate (X_i). We will assume that distribution of Y_i (denoted by P_{0i}) will depend on X_i. We will denote the finite product measure $\prod_{i=1}^{n} P_{0i}$ as well as the infinite product measure $\prod_{i=1}^{\infty} P_{0i}$ by P, and the corresponding expectation by $E[\cdot]$.

To keep the exposition simple, we derive results for jointly modeling two quantiles of $Y_i|X_i$, corresponding to $\boldsymbol{\tau} = (\tau_1, \tau_2)$. The arguments are easily extendable to more than two quantiles and to the case of multiple random or non-random covariates. Let X_i take values in $\mathcal{X}$, a subset of the real line $\Re$. The specified model for the response conditional on X_i is given by $Y_i|X_i \sim f_{\tau_1}(Y_i - Q_{\tau_1}(X_i)) \cdot f_{\tau_1}(Y_i - Q_{\tau_2}(X_i))$, where f_τ is as in (4), Q_{τ_1} and Q_{τ_2} belong to a class of functions $\mathcal{G}$ defined on $\mathcal{X}$, with a proper prior Π defined on $\mathcal{G} \times \mathcal{G}$. However, the true (but unknown) probability distribution of $Y_i|X_i$ is P_{0i} with the "true" τ^{th} conditional quantile given by $Q_{0\tau}(X_i)$, where $Q_{0\tau} \in \mathcal{G}$. To simplify notations, we write $\mathbf{Q} = (Q_{\tau_1}, Q_{\tau_2})$, and correspondingly the true quantile functions as $\mathbf{Q_0} = (Q_{0\tau_1}, Q_{0\tau_2})$. We define

$$g_{\mathbf{Q}}(y, x) := f_{\tau_1}(y - Q_{\tau_1}(x)) \cdot f_{\tau_2}(y - Q_{\tau_2}(x)). \tag{6}$$

Based on a prior Π for $\mathbf{Q}$, the posterior distribution is given by

$$\Pi\left(\mathbf{Q}|Y_{1:n}\right) \qquad \propto \prod_{i=1}^{N} g_{\mathbf{Q}}(Y_i, X_i) \cdot \Pi(\mathbf{Q}) \tag{7}$$

We discuss posterior consistency with respect to an empirical L_2 metric d_n defined as follows:

$$d_n^2(\mathbf{Q}, \mathbf{Q}_0) := \frac{1}{n}\sum_{i=1}^{n}[|Q_{\tau_1}(X_i) - Q_{0\tau_1}(X_i)|^2 + |Q_{\tau_2}(X_i) - Q_{0\tau_2}(X_i)|^2] \tag{8}$$

In the *i.n.i.d.* context, it is natural to consider such an empirical average metric (e.g. see Ghosal and van der Vaart 2007).

Our aim is to derive conditions under which the posterior probability

$$\Pi(d_n(\mathbf{Q}, \mathbf{Q_0}) > \epsilon | Y_{1:n}) \to \ 0 \ a.s. \ [P].$$

Towards deriving our results, we write $U_n^c := \{d_n(\mathbf{Q}, \mathbf{Q_0}) > \epsilon\}$ and the posterior probability as

$$\Pi(U_n^c|Y_{1:n}) := \frac{\int_{U_n^c} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})}{\int_{\mathcal{G}\times\mathcal{G}} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})}. \tag{9}$$

We make the following assumptions. The first assumption on positivity of Kullback-Leibler neighborhoods is standard and helps control the denominator of the posterior probability, as seen in Proposition 1 below.

ASSUMPTION 1. Π *is a proper prior and for any* $\delta > 0$

$$\Pi\left(\left\{\boldsymbol{Q}: \ \frac{1}{n}\sum_{i=1}^{n} E\log\frac{g_{\mathbf{Q_0}}(Y_i, X_i)}{g_{\boldsymbol{Q}}(Y_i, X_i)} < \delta \ \forall \ n, \ \sum_{i=1}^{\infty}\frac{1}{i^2}E\log^2\frac{g_{\mathbf{Q_0}}(Y_i, X_i)}{g_{\boldsymbol{Q}}(Y_i, X_i)} < \infty\right\}\right) > 0.$$

Using Lemma 1 of Sriram et al. (2013), it can be easily seen that

$$\frac{1}{n}\sum_{i=1}^{n} E\log\frac{g_{\mathbf{Q_0}}(Y_i,X_i)}{g_{\mathbf{Q}}(Y_i,X_i)} \leq \mathbf{e}(\mathbf{Q},\mathbf{Q_0}), \tag{10}$$

where $\mathbf{e}(\cdot,\cdot)$ is the supnorm metric given by

$$\mathbf{e}(\mathbf{Q_1},\mathbf{Q_2}) := \sup_{x\in\mathcal{X}}(|Q_{1\tau_1}(x)-Q_{2\tau_1}(x)|+|Q_{1\tau_2}(x)-Q_{2\tau_2}(x)|). \tag{11}$$

It follows that Assumption 1 would be satisfied if Π puts a positive probability on all sup-norm neighborhoods of $\mathbf{Q_0}$.

PROPOSITION 1. *Suppose Assumption 1 holds. Then for any* $\beta > 0$

$$e^{n\beta}\cdot\int_{\mathcal{G}\times\mathcal{G}}\prod_{i=1}^{n}\frac{g_{\boldsymbol{Q}}(Y_i,X_i)}{g_{\mathbf{Q_0}}(Y_i,X_i)}d\Pi(\boldsymbol{Q}) \to \ \infty \ a.s.\ [P].$$

We skip the proof of the proposition, which is similar to Lemma 4.4.1 of Ghosh and Ramamoorthi (2003).

The next assumption is on the covariate space and the class of quantile functions.

ASSUMPTION 2.

(a) The covariate space $\mathcal{X}$ *is compact.*

(b) $\exists\ M > 0$ *such that* $\sup_{x\in\mathcal{X}}(|Q_{\tau_1}(x)|+|Q_{\tau_2}(x)|) \leq M\ \forall\ Q_{\tau_1},Q_{\tau_2}\in\mathcal{G}$.

This leads to the following important lemma that can be used to control the numerator of the posterior probability in (9).

LEMMA 1. *If Assumption 2 holds, then* $\exists\ \alpha\in\ (0,1)$ *and a constant* $C > 0$ *such that the following hold for any* $\boldsymbol{Q}\in\mathcal{G}\times\mathcal{G}$*:*

a) $$E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^\alpha \leq e^{-C|Q_\tau(x)-Q_{0\tau}(x)|^2} \quad \text{for } \tau\in\{\tau_1,\ \tau_2\}.$$

b) $$E\left(\frac{g_{\boldsymbol{Q}}(y,x)}{g_{\mathbf{Q_0}}(y,x)}\right)^\alpha \leq e^{-C(|Q_{\tau_1}(x)-Q_{0\tau_1}(x)|^2+|Q_{\tau_2}(x)-Q_{0\tau_2}(x)|^2)}.$$

c) $$E\left(\prod_{i=1}^{n}\frac{g_{\boldsymbol{Q}}(Y_i,X_i)}{g_{\mathbf{Q_0}}(Y_i,X_i)}\right)^\alpha \leq e^{-Cnd_n^2(\boldsymbol{Q},\mathbf{Q_0})}.$$

In the interest of flow, we defer the proof of the lemma to Appendix A.

The next assumption is essentially a sieve and entropy condition that helps bound the numerator of the posterior probability.

ASSUMPTION 3. *For any $b > 0$, $c > 0$, suppose $\mathcal{G} \times \mathcal{G} \subseteq W_n \cup V_n$ such that,*

(i) $\Pi(V_n) \leq K_1 \cdot e^{-cn}$ *for some constant* K_1,

(ii) For any $\delta > 0$, $\mathbb{N}(\delta, W_n, \boldsymbol{e}) \leq K_2 \cdot e^{nb}$ *for some constant* K_2, *where* $\mathbb{N}(\delta, W_n, \boldsymbol{e}) :=$ *the number of supnorm balls of radius* δ *that cover* W_n, *where the supnorm* $\boldsymbol{e}(\cdot, \cdot)$ *is as defined in (11).*

We now derive our main result using the above assumptions. We then discuss some examples and make some observations.

THEOREM 1. *Under Assumptions 1, 2 and 3,*

$$\Pi\left(d_n(\mathbf{Q}, \mathbf{Q_0}) > \epsilon \mid Y_{1:n}\right) \to 0 \ a.s.\ [P].$$

PROOF. Recall $U_n^c = \{\mathbf{Q} : d_n(\mathbf{Q}, \mathbf{Q_0}) > \epsilon\}$. For b and c to be chosen later, let $\mathcal{G} \times \mathcal{G} = W_n \cup V_n$ as in Assumption 3. Also, for $\delta > 0$ to be chosen later, let $\{B_j, 1 \leq j \leq \mathbb{N}(\delta/2, W_n, \mathbf{e})\}$ be an open cover of W_n. Using the inequality in (10) (with $\mathbf{Q}_j$ in place of $\mathbf{Q}_0$), we note that for $\mathbf{Q}, \mathbf{Q}_j \in B_j \cap U_n^c$, $\prod_{i=1}^n \frac{g_{\mathbf{Q}}(Y_i,X_i)}{g_{\mathbf{Q_j}}(Y_i,X_i)} \leq e^{n \cdot \mathbf{e}(\mathbf{Q},\mathbf{Q_j})} \leq e^{n\delta}$. Now, for α as in Lemma 1 and using Cauchy-Schwartz inequality, we get

$$\begin{aligned}
&E\left(\int_{B_j \cap U_n^c} \prod_{i=1}^n \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right)^{\frac{\alpha}{2}} \\
&\leq \sqrt{E\left(\int_{B_j \cap U_n^c} \prod_{i=1}^n \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_j}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right)^{\alpha}} \cdot \sqrt{E\left(\prod_{i=1}^n \frac{g_{\mathbf{Q_j}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i}\right)^{\alpha}} \\
&\leq e^{\frac{n\delta}{2}} \cdot e^{-\frac{n d_n^2(\mathbf{Q_1}, \mathbf{Q_0})}{2}} \cdot \Pi(B_j)^{\frac{\alpha}{2}} \leq e^{-\frac{n(\epsilon^2 - \delta)}{2}}.
\end{aligned}$$

Choosing $\delta = \frac{\epsilon^2}{2}$, $b = \epsilon^2/8$, using the fact that $W_n \subseteq \cup_j B_j$, and applying part (ii) of Assumption 3, we get

$$E\left(\int_{U_n^n \cap W_n} \prod_{i=1}^n \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right)^{\frac{\alpha}{2}} \leq \sum_j E\left(\int_{B_j \cap U_n^c} \prod_{i=1}^n \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right)^{\frac{\alpha}{2}}$$

$$
\begin{aligned}
&\leq \mathbb{N}(\delta/2, W_n, \mathbf{e}) \cdot e^{-\frac{n\epsilon^2}{4}} \leq K_2 \cdot e^{nb} \cdot e^{-\frac{n\epsilon^2}{4}} \\
&\leq K_2 \cdot e^{-\frac{n\epsilon^2}{8}}
\end{aligned}
$$

By Markov's inequality,

$$
\begin{aligned}
&P\left(\int_{U_n^n \cap W_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q}) > e^{-2n\beta_0}\right) \\
&\leq e^{n\alpha\beta_0} \cdot E\left(\int_{U_n^c \cap W_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right)^{\frac{\alpha}{2}} \\
&\leq K_2 \cdot e^{-n(\frac{\epsilon^2}{8} - \beta_0)}.
\end{aligned}
$$

Choosing a small enough β_0 and applying Borel-Cantelli lemma gives

$$
e^{n\beta_0} \cdot \int_{U_n^c \cap W_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q}) \to \ 0 \ a.s. \ [P].
$$

This along with Proposition 1 gives

$$
\Pi(U_n^c \cap W_n | Y_{1:n}) \to \ 0 \ a.s. \ [P].
$$

To complete the proof, it is now enough to argue that $\Pi(U_n^c \cap V_n | Y_{1:n}) \to 0 \ a.s. \ [P]$. To see this, note by Assumption 2 that $\log \frac{g_{\mathbf{Q}}(Y_i,X_i)}{g_{\mathbf{Q_0}}(Y_i,X_i)} \leq 2M$. Hence,

$$
\begin{aligned}
&P\left(\int_{U_n^n \cap V_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q}) > e^{-2n\beta_0}\right) \\
&\leq e^{n\beta_0} \cdot E\left(\int_{U_n^c \cap V_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q})\right) \\
&\leq e^{n\beta_0 + 2nM} \cdot \Pi(V_n) \leq K_1 \cdot e^{n(2M + \beta_0 - c)}.
\end{aligned}
$$

The last step uses part (i) of Assumption 3. For a large enough choice of c, Borel-Cantelli lemma can be applied to obtain

$$
e^{n\beta_0} \cdot \int_{U_n^c \cap V_n} \prod_{i=1}^{n} \frac{g_{\mathbf{Q}}(Y_i, X_i)}{g_{\mathbf{Q_0}}(Y_i, X_i)} d\Pi(\mathbf{Q}) \to \ 0 \ a.s. \ [P].
$$

This along with Proposition 1 completes the proof.

REMARK 1 (Parametric example). Theorem would hold when the quantile function is a smooth linear or nonlinear function in finitely many parameters, which take values in a compact set. In this case, Assumptions 2 and 3 would follow due to compactness and Assumption 1 would hold as long as Euclidean neighborhoods around the true parameter value get positive prior probabilities.

REMARK 2 (Non-parametric example). As a nonparametric example, let $(Q_{\tau_1}(x), Q_{\tau_2}(x))$ be modeled as $(H_1(\eta(x)), H_2(\eta(x)))$, which is uniquely determined by a real valued function $\eta(\cdot)$ and known bounded uniformly continuous functions $H_1(\cdot)$, $H_2(\cdot)$, on a compact $\mathcal{X} \subset \mathbb{R}$ such that $H_2(x) \geq H_1(x)\ \forall\ x \in \mathcal{X}$. Let $\eta(\cdot)$ be endowed with a Gaussian process prior with mean $\mu(x)$ and covariance kernel $\sigma(x', x) := \frac{1}{\tau}\sigma_0(\lambda x', \lambda x)$, along with priors Π_λ and Π_τ on the parameters λ and τ respectively. Ghosal and Roy (2006) consider this set up with a univariate function of the Gaussian process for binary regression. Our extension to model two quantiles via the bivariate function $(H_1(\eta), H_2(\eta))$ proceeds along the same lines. For simplicity of our exposition, we assume σ_0 is once continuously differentiable. Suppose that the true quantile functions are $(H_1(\eta_0), H_2(\eta_0))$. Under the assumption that the mean function $\mu(\cdot)$ and $\eta_0(\cdot)$ belong to the Reproducing Kernel Hilbert Space (RKHS) of the Gaussian process prior, Theorem 4 of their paper ensures that supnorm neigborhoods of η_0 get positive probability and hence our Assumption 1 holds. Compactness of $\mathcal{X}$ and boundedness of H_1 and H_2 imply Assumption 2. Towards obtaining Assumption 3, for a $c > 0$, suppose sequences λ_n and τ_n are such that $\Pi_\lambda(\lambda < \lambda_n) < e^{-cn}$ and $\Pi_\tau(\tau > \tau_n) < e^{-cn}$. Further, for any $b_1 > 0$, $b_2 > 0$ suppose $\exists$ sequence M_n such that $M_n^2 \geq b_1 n\lambda_n^2/\tau_n$ and $M_n \leq b_2 n$, then Lemmas 1 and 2 of their paper imply part (i) and (ii) of Assumption 3 respectively. While their assumption on the sequences require that the sequences exist for some value of $c > 0$, our result essentially requires that such sequences can be obtained for some value of c exceeding a certain minimum threshold. Such a requirement is standard for obtaining posterior consistency in misspecified models.

REMARK 3. In applications involving ALD, it is common to include a scale parameter. Given that the model is misspecified, the scale parameter does not have a clear interpretation. However, including such a parameter can be computationally useful. We believe that the above approach can be modified in the lines of Sriram et al. (2013), to include a prior on the scale parameters $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_K)$ by defining the pseudo likelihood as:

$$p_{\boldsymbol{\tau},\boldsymbol{\mu_i},\boldsymbol{\sigma}}(y) = \prod_{j=1}^{K} f_{\tau_j}\left(\frac{y - \mu_{\tau_j,i}}{\sigma_j}\right). \tag{12}$$

REMARK 4. When the parameter space is compact, it is possible to obtain posterior conistency for joint modeling of quantiles using the pseudo-joint ALD likelihood, with respect to the supnorm metric $\mathbf{e}(\cdot,\cdot)$ as in (11) instead of the empirical L_2 metric $d_n(\cdot,\cdot)$. For a single ALD, such a result is briefly discussed in Section 6 of a recent forthcoming article by Ramamoorthi et al. (2015). We believe that these arguments can be extended to the pseudo joint likelihood for modeling multiple quantiles as well. As an additional assumption, such an extension would require a denseness condition or an in-fill condition on the covariate space in the lines of Assumption B in Section 6 of their paper.

3 Pseudo-likelihood Versus a Normalized Proper Likelihood

The likelihood (5) does not integrate to 1 w.r.t. y. It is not hard to evaluate the (albeit cumbersome) normalizing constant. However, if we include this constant as part of the working likelihood, it turns out that the Kullback-Liebler divergence of the working likelihood w.r.t. the true underlying density may not be minimized at the desired quantiles of interest. This fact is captured in the following simple proposition.

PROPOSITION 2. *Let $f(y)$ denote the density obtained after normalizing the density (5) for the particular case when $K = 2$, $\tau_1 = 1 - \tau_2 = .25$ and $\mathcal{Y} = (-\infty, \infty)$. Let p_0 be the true underlying density which is continuous. Then the Kullback-Leibler divergence $E_{p_0} \log \frac{p_0}{f}$ is minimized when $\mu^{\tau_1} = \mu^{\tau_2} =$ median of p_0.*

We skip the proof of this proposition as it follows from elementary calculus. It is well known in the case of misspecified models that posterior consistency is attained at the density that minimizes the Kullback-Leibler divergence w.r.t. the true density. Therefore, a consequence of the proposition is that the resulting quantile estimates from normalized density cannot be posterior consistent for the true quantile values and hence could be misleading

4 Simulations

4.1. Simulation 1: a Constrained Non-linear Model. Here, we consider a simple example involving the estimation of a non-linear quantile model with constraints. Suppose that the actual distribution of Y_i given X_i is a gamma distribution with shape parameter = 0.5 and scale parameter is $(1+X_i)$. We simulate 500 data points from this model. Suppose that we are interested in estimating quantiles at $\tau_1 = 0.05$, $\tau_2 = 0.5$ and $\tau_3 = 0.95$, i.e. median and two extreme quantiles. A property of this *gamma* distribution is that the τ^{th} quantile of $Y_i|X_i$ is of the form $(\alpha_0 + \beta_0 X_i) \times \gamma_0(\tau)$, where $\alpha_0 = \beta_0 = 1$ and $\gamma_0(\tau)$ is the τ^{th} quantile of a gamma distribution with shape =.5 and scale =1. We demonstrate in this example that the joint quantile formulation based on pseudo joint ALD likelihood to estimate $(\alpha_0, \beta_0, \gamma_{0\tau_1}, \gamma_{0\tau_2}, \gamma_{0\tau_3})$ performs better than running a quantile regression for each of the quantiles separately.

In practice, we will not know the true underlying distribution, but suppose we know that the extreme quantiles can be approximately modeled as a scalar multiple of the intermediate quantile. Then, the model formulation for the quantiles of $Y_i|X_i$ is as follows:

$$Q_{\tau_j}(X_i) = (\alpha + \beta X_i) \times \gamma_{\tau_j}, \text{ where } \gamma_{\tau_3} > \gamma_{\tau_2} > \gamma_{\tau_1}. \tag{13}$$

Note that while the model is linear in X_i, it is non-linear in the parameters $(\alpha, \beta, \gamma_{\tau_1}, \gamma_{\tau_2}, \gamma_{\tau_3})$. A simple approach to estimate such a model is to run a linear quantile regression in X_i separately for each of the quantiles at τ_1, τ_2 and τ_3. Based on this simple approach, the first row of Table 1 shows the root mean squared error (RMSE), the coverage ratio (COV) and average length (LEN) of the 95 % credible interval across 500 replications of the simulated data.

The approach based on separately estimating the model at each quantile ignores the constraints in the model, i.e. that the parameters (α, β) are common across quantiles and that $\gamma_{\tau_3} > \gamma_{\tau_2} > \gamma_{\tau_1}$. A better approach is to account for these constraints. This can be easily formulated by using pseudo-joint ALD likelihood for Y_i along with (13). As priors for (α, β), we take a relatively flat prior using a product of Gaussian distributions centered at 0.1 and variance 10. For identifiability reasons and without loss of generality, we take $\gamma_{\tau_3} = 1$ and endow a uniform prior on the set $\{(\gamma_{\tau_1}, \gamma_{\tau_2}) : 0 < \gamma_{\tau_1} < \gamma_{\tau_2} < 1\}$. Including the scale parameter of ALD usually improves the performance of the markov chain monte carlo method in simulating the posterior distribution. Accordingly, we consider the pseudo-joint likelihood with scale parameters as in (12), and use independent gamma priors with shape 10 and scale 0.5 for $(\sigma_{\tau_1}, \sigma_{\tau_2}, \sigma_{\tau_3})$. The second row of Table 1 shows the root mean squared error (RMSE), the coverage ratio (COV) and average length (LEN) of the 95 % credible interval based on this approach across 500 replications of the simulated data.

Table 1: (**Simulation 1**) Comparison of results when quantiles are estimated separately versus jointly based on the pseudo joint ALD likelihood with constraints

Method		α_{τ_1}	β_{τ_1}	α_{τ_2}	β_{τ_2}	α_{τ_3}	β_{τ_3}
Quantiles	RMSE	0.0151	0.0021	0.2369	0.0421	1.3837	0.2775
estimated	COV	100 %	100 %	99.6 %	91.4 %	80.2 %	52 %
separately	LEN	0.192	0.0184	1.2366	0.1449	3.4309	0.4073
Quantiles	RMSE	0.0025	0.0011	0.1605	0.0336	1.3101	0.2692
estimated	COV	97 %	100 %	79.2 %	83 %	79.2 %	52.8 %
jointly	LEN	0.0095	0.0065	0.4038	0.0897	3.254	0.3937

The entries within each cell of the table are root mean squared error (RMSE), coverage ratio (COV) and length (LEN) of 95 % credible intervals across 500 replications of the simulated data

Comparing the results from the two approaches, we see that the latter approach based on the pseudo-joint likelihood with constraints, consistently has lower RMSE for the parameters across all quantiles. This difference is more pronounced for the quantile at $\tau = 0.05$, which would have been difficult to estimate without knowledge of the constraints across quantiles. This clearly demonstrates the utility of pseudo-joint likelihood in helping incorporate constraints across quantiles. The coverages of the two methods are comparable and the 95 % credible intervals are seen to be smaller when the quantiles are estimated jointly. However, as noted in the introduction, since the pseudo-joint ALD is essentially a misspecification of the true data generating mechanism, the coverages of the 95 % credible intervals cannot be expected to be close to 95 %. An appropriate correction to the credible intervals may be of further research interest.

4.2. Simulation 2: Linear Model with a Common Slope Parameter. The Bayesian posterior leads to the 95 % credible intervals for the parameters. However, given that the pseudo-joint ALD model is essentially a misspecification, the (frequentist) asymptotic coverage of the credible intervals may not match 95 %.(see Kleijn and van der Vaart 2012). Here, we empirically investigate this coverage property for a linear quantile model. Of special interest to us, is also the common slope parameter across quantiles.

We simulate data from a “true” model, whose quantiles are given by

$$Q_{0\tau}(X_i, Z_i) = \alpha_{0\tau} + \beta_{0\tau} X_i + \gamma_0 Z_i,\ i = 1, 2, \ldots, N.$$

Note that γ_0 is a common slope for all quantiles and $(\alpha_{0\tau}, \beta_{0\tau})$ vary across quantiles. More specifically, for the simulation, we take $\gamma_0 = 0.4$, $\alpha_\tau = 0.01 + 0.02\tau$ and $\beta_\tau = 0.03 + 0.04\tau + 0.09\tau^2$.

We specify the model as:

$$Q_\tau(X_i, Z_i) = \alpha_\tau + \beta_\tau X_i + \gamma Z_i,$$

where γ is a common slope parameter across quantiles. Estimation is done for three quantiles corresponding to $(\tau_1, \tau_2, \tau_3) = (.25, .5, .75)$. We consider scenarios arising from three different prior specifications on the parameters, all independent normals centered at 0 but with three different values for the prior standard deviation viz., 1, 10 and 100, and three choices for the data size (N) viz., 10,100 and 1000. For the posterior computation, we use the pseudo joint ALD likelihood as in (12) along with independent gamma priors with shape 10 and scale 0.5 for the scale parameters. We generate 1000 replications of the simulation for each scenario. For each replication in a given scenario (i.e a given prior standard deviation and a given data size), data is simulated from the true model and the 95 % credible intervals for all the parameters are computed based on the specified model and the given prior. Then the length of the interval is recorded and it is checked whether the interval contains the true parameter value. After the 1000 replications are completed, the proportion of replications whose credible interval contains the true parameter value (i.e. coverage) is computed along with the average length of the intervals. Table 2 shows the coverage results and average length of intervals (shown within parenthesis) across the 1000 replications for the different scenarios . For misspecified models, we expect that the coverage from such intervals would either be systematically under or over 95 %. Since this is anyway an asymptotic property, we can see that for small sample size (n = 10), the coverage percentages are far from what is desired. With increasing sample size ($N = 100, 1000$), the coverage percentage improves and also the length of intervals decreases. As expected for misspecified models, the coverage percentage is not always close to the desired 95 % even for large N. We can also see that the coverage percentages for the slope parameters are usually lower than 95 % and that for the intercept are higher than 95 %.

Interestingly, at large sample sizes, the coverage for the common slope parameter (γ) is slightly better than that of the other slope parameters β_{τ_2} and β_{τ_3}, which vary across quantiles. This may be due to the borrowing of information across quantiles about the common slope. This seems to suggest that if there is indeed a common slope parameter, then we should be able to estimate it with better precision by jointly modeling larger number of quantiles. To explore this point and also a point raised by a referee on potential over-narrowing of credible intervals for the common slope parameter, we conduct further investigation. We estimate the common slope parameter by considering the pseudo joint-ALD model with three, five, seven and

Table 2: (**Simulation 2**) Coverage of 95 % credible intervals

Prior sd	N	α_{τ_1}	β_{τ_1}	α_{τ_2}	β_{τ_2}	α_{τ_3}	β_{τ_3}	γ
1	10	78 %	61 %	77 %	57 %	72 %	55 %	51 %
		(0.54)	(0.05)	(0.61)	(0.06)	(0.65)	(0.07)	(0.72)
1	100	95 %	83 %	92 %	74 %	91 %	72 %	78 %
		(0.17)	(0.02)	(0.22)	(0.03)	(0.24)	(0.03)	(0.21)
1	1000	97 %	84 %	97 %	79 %	95 %	75 %	82 %
		(0.05)	(0.01)	(0.06)	(0.01)	(0.07)	(0.01)	(0.06)
10	10	73 %	60 %	74 %	56 %	71 %	50 %	50 %
		(0.56)	(0.05)	(0.61)	(0.06)	(0.65)	(0.06)	(0.76)
10	100	95 %	84 %	92 %	76 %	92 %	73 %	79 %
		(0.17)	(0.02)	(0.23)	(0.03)	(0.24)	(0.03)	(0.21)
10	1000	97 %	84 %	96 %	75 %	95 %	72 %	83 %
		(0.05)	(0.01)	(0.06)	(0.01)	(0.07)	(0.01)	(0.06)
100	10	73 %	59 %	76 %	55 %	74 %	53 %	51 %
		(0.58)	(0.06)	(0.62)	(0.06)	(0.66)	(0.06)	(0.77)
100	100	95 %	84 %	93 %	77 %	91 %	71 %	82 %
		(0.17)	(0.02)	(0.23)	(0.03)	(0.24)	(0.03)	(0.21)
100	1000	96 %	84 %	97 %	77 %	95 %	77 %	82 %
		(0.05)	(0.01)	(0.06)	(0.01)	(0.07)	(0.01)	(0.06)

Entries in the table show coverage (in %) and the average length (shown within parenthesis) based on 1000 replications

nine quantiles as separate cases. The coverage of the 95 % credible intervals for the common slope γ and the average length of credible interval, across

Table 3: (**Simulation 2**) Coverage of 95 % Credible Intervals for Common Slope Parameters (when different number of quantiles are jointly modeled)

N	Three	Five	Seven	Nine
10	51 %	42 %	37 %	32 %
	(0.72)	(0.55)	(0.42)	(0.34)
100	78 %	77 %	74 %	70 %
	(0.21)	(0.13)	(0.09)	(0.07)
1000	82 %	84 %	81 %	85 %
	(0.08)	(0.04)	(0.02)	(0.01)

Entries in the table show coverage (in %) and the average length (shown within parenthesis) based on 1000 replications, for the case when prior sd = 1. The column headings refer to the number of quantiles jointly modeled

1000 replications for each of these cases is shown in Table 3. The results are shown at different sample sizes for the scenario when the prior variance is 1. For the smaller sample sizes (i.e. $N = 10$ or 100), as we estimate more number of quantiles, the length of credible interval decreases. However, this is accompanied by a decrease in coverage ratio. For the larger sample size, i.e. $N = 1000$, as we estimate more number of quantiles, there is more stability and no more a clear decline in coverages, while the length of credible intervals still decreases. This suggests two things, (i) if indeed there is a common slope parameter, at least with large sample sizes, more accuracy can be achieved by jointly estimate multiple quantiles and (ii) for any fixed sample size, there may be a limit to how many quantiles one can jointly estimate, without overly narrowing down the credible intervals for the common slope parameter. We defer a detailed theoretical investigation of this issue to a future work.

5 Conclusion

We examined the use of a pseudo-likelihood based on the asymmetric Laplace density, which arises as a natural approach to consider for jointly estimating multiple quantiles. While the idea of considering a product of such densities to handle multiple quantiles arises naturally, our focus on the "pseudo" versus a "normalized" version of the likelihood is motivated by the interesting observation that the estimates from the latter approach could be misleading. We showed that posterior consistency for a non-linear quantile regression model (and in particular for the linear model) holds under some reasonable assumptions. The ALD has been a computationally attractive approach for modeling a single quantile. We expect similar computational benefits to carry over to estimating multiple quantiles based on the pseudo-likelihood. We discussed a simulation example where such a method is seen to be beneficial. The simulation study also highlighted a few issues that may be worthy of future research. First, since the likelihood is essentially a misspecification, the 95 % credible interval is not expected to have an asymptotic coverage of 95 %. Second, if the model has a common slope parameter across quantiles, for a fixed sample size, estimating too many quantiles jointly with this approach may lead to over narrowing of intervals for the common slope parameter. A theoretical investigation of these issues and potential remedial steps could be of future research interest.

Acknowledgments. We thank the Editor, Associate Editor and the referees for their insightful observations and helping us significantly improve the quality of the paper.

Appendix

A Details of the Proof of Lemma 1

PROOF OF LEMMA 1. The statements (b) and (c) are immediate from (a). Hence, we will focus on showing (a). Statement (a) can be obtained as a consequence of some results, particularly Corollary 1 of Theorem 8 in a recent article by Ramamoorthi et al. (2015). However, for completeness, we reproduce the arguments here.

It follows from Lemma 1 of Sriram et al. (2013) that $\log \frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))} \leq |Q_\tau(x) - Q_{0\tau}(x)|$ and by Lemma 2 of the same paper that $Q_{0\tau}(X_i) = arg\min_{Q_\tau \in \mathcal{G}} E \log \frac{p_0 i(Y_i)}{f_\tau(Y_i - Q_\tau(X_i))}$, for any $\tau \in (0,1)$. Therefore, by Assumption 2, $\exists\ M > 0$ such that $\log \frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))} \leq M$. Hence, $\exists\ 0 < \alpha < 1$ that does not depend on Q_τ such that $\alpha \cdot \left|\log \frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right| \leq \alpha \cdot |Q_\tau(x)) - Q_{0\tau}(x))| < 1/2$. We note that when $t < 1, e^t < \frac{1}{1-t}$, and further that when $t < 1/2$, $e^t - 1 < t/(1-t) < 2t$. Applying this inequality, we get

$$e^{\alpha \cdot \log \frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}} - 1 < 2\alpha \cdot \log \frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}.$$

Therefore, $E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^\alpha < 1 - 2\alpha \cdot E \log \frac{f_\tau(y-Q_{0\tau}(x))}{f_\tau(y-Q_\tau(x))}$. Since $0 \leq 2\alpha \cdot E \log \frac{f_\tau(y-Q_{0\tau}(x))}{f_\tau(y-Q_\tau(x))} < 1$, we have $E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^\alpha \leq 1$.

Without loss of generality, assume $\alpha = \frac{1}{2^{K-1}}$ for some $K > 1$. Define $a := \left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^{\frac{1}{2^K}}$. Then

$$\begin{aligned} E\left|\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))} - 1\right| &= E\left|a^{2^k} - 1\right| = E\left[|a-1| \cdot \left|1 + a^2 + a^3 + \cdots + a^{2^k-1}\right|\right] \\ &\leq E\left(|a-1|^2\right)^{\frac{1}{2}} \cdot \left(E\left|1 + a^2 + a^3 + \cdots + a^{2^k-1}\right|^2\right)^{\frac{1}{2}}. \end{aligned}$$

For the first term on the right hand side of the above inequality, we have

$$\begin{aligned} E|a-1|^2 &= E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^{\frac{1}{2^{k-1}}} + 1 - 2E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^{\frac{1}{2^k}} \\ &\leq 2\left(1 - E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^{\frac{1}{2^k}}\right). \end{aligned}$$

Note that every term within the expansion $E\left|1+a^2+a^3+\cdots+a^{2^k-1}\right|^2$ is of the form $Ea^l = E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^{\frac{l}{2^k}}$, where $l \le 2^{k+1}-2$. In particular, the second term is bounded by some constant multiple of e^{2M}. Hence, we can conclude that there exists some constant C such that

$$E\left|\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}-1\right| \le C\cdot\sqrt{1-E\left(\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}\right)^\alpha} \tag{14}$$

The proof of statement (a) will be complete if we show that for some C_0

$$E\left|\frac{f_\tau(y-Q_\tau(x))}{f_\tau(y-Q_{0\tau}(x))}-1\right| \quad\ge\quad C_0|Q_\tau(x)-Q_{0\tau}(x))|. \tag{15}$$

First, it can be checked using the form of asymmetric Laplace density that

$$\left|\frac{f_\tau(Y-Q_\tau(x))}{f_\tau(Y-Q_{0\tau}(x))}-1\right| \ge \begin{cases} \left(1-e^{-(Q_\tau(x)-Q_{0\tau}(x))(1-\tau)}\right)\cdot I_{(Z\le 0)}, & if\ Q_\tau(x)-Q_{0\tau}(x)\ge 0 \\ \left(1-e^{(Q_\tau(x)-Q_{0\tau}(x))\tau}\right)\cdot I_{(Z>0)} & ,if\ Q_\tau(x)-Q_{0\tau}(x)<0 \end{cases}.$$

where $Z = Y - Q_{0\tau}(X)$. Since $|Q_\tau(x) - Q_{0\tau}(x)|$ is uniformly bounded (by Assumption 2), we can further say that there exists a constant $C_0 > 0$ such that

$$\left|\frac{f_\tau(Y-Q_\tau(x))}{f_\tau(Y-Q_{0\tau}(x))}-1\right| \ge C_0|Q_\tau(x)-Q_{0\tau}(x)|\cdot(I_{Z\le 0}\cdot I_{\theta-\theta_0\ge 0}+I_{Z>0}\cdot I_{\theta-\theta_0<0}).$$

Equation (15) follows by noting that $E\left[I_{Z\le 0}|X\right] = P(Z\le 0|X) = \tau$.

References

BENOIT, D.F. and VAN DEN POEL, D. (2012). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *J. Appl. Econometrics* **27**, 1174–1188.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78**, 1093–1125.

DETTE, H. and VOLGUSHEV, S. (2008). Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70**, 609–627.

DUNSON, D.B. and TAYLOR, J.A. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Stat.* **17**, 385–400.

GHOSAL, S. and ROY, A. (2006). Posterior consistency of gaussian process prior for non-parametric binary regression. *Ann. Statist.* **34**, 2413–2429.

GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35**, 192–223.

GHOSH, J.K. and RAMAMOORTHI, R.V. (2003). *Bayesian nonparametrics.* Springer.

HE, X. (1997). Quantile curves without crossing. *Amer. Statist.* **51**, 186–192.

KLEIJN, B.J.K. and VAN DER VAART, A.W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electron. J. Stat.* **6**, 354–381.

KOENKER, R. (1984). A Noteon L-estimates for linear models. *Statist. Probab. Lett.* **2**, 322–325.

KOENKER, R. and BASSETT, G.J. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

RAMAMOORTHI, R., SRIRAM, K. and MARTIN, R. (2015). *On posterior concentration in misspecified models.* (Forthcoming) Available online at http://bayesian.org/BA/19667.

REICH, B.J., FUENTES, M. and DUNSON, D.B. (2011). Bayesian spatial quantile regression. *J. Amer. Statist. Assoc.* **106**, 6–20.

SRIRAM, K., RAMAMOORTHI, R.V. and GHOSH, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Anal.* **8**, 479–504.

TADDY, M.A. and KOTTAS, A. (2010). A Bayesian nonparametric approach to inference for quantile regression. *J. Bus. Econom. Statist.* **28**, 357–369.

TOKDAR, S. T. and KADANE, J. B. (2012). Simultaneous linear quantile regression: a semiparametric Bayesian approach. *Bayesian Anal.* **7**, 51–72.

WU, Y. and LIU, Y. (2009). Stepwise multiple quantile regression estimation using non-crossing constraints. *Stat. Interface* **2**, 299–310.

YANG, Y. and HE, X. (2012). Bayesian empirical likelihood for quantile regression. *Annals Stat.* **40**, 1102–1131.

YU, K. and MOYEED, R.A. (2001). Bayesian quantile regression. *Statist. Probab. Lett.* **54**, 437–447.

YU, K., VAN KERM, P. and ZHANG, J. (2005). Bayesian quantile regression: an application to the wage distribution in 1990s Britain. *Sankhyā* **57**, 359–377.

YUE, Y.R. and RUE, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Statist. Data Anal.* **55**, 84–96.

ZOU, H. and YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals Stat.* **36**, 1108–1126.

Karthik Sriram
Production and Quantitative
Methods Area, Indian Institute
of Management Ahmedabad,
Ahmedabad, India
E-mail: karthiks@iimahd.ernet.in

R. V. Ramamoorthi
Statistics and Probability,
Michigan State University,
East Lansing, MI, USA
E-mail: ramamoor@stt.msu.edu

Pulak Ghosh
Department of Quantitative
Methods and Information Systems,
Indian Institute of Management
Bangalore, Bangalore, India
E-mail: pulak.ghosh@iimb.ernet.in

Paper received: 12 December 2013.