Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2011, Article ID 874907, 16 pages
doi:10.1155/2011/874907

Research Article

Power Prior Elicitation in Bayesian Quantile Regression

Rahim Alhamzawi and Keming Yu

Department of Mathematical Sciences, Brunel University, Uxbridge UBB 3PH, UK

Correspondence should be addressed to Keming Yu, keming.yu@brunel.ac.uk

Received 12 February 2010; Accepted 1 November 2010

Academic Editor: Tomasz J. Kozubowski

We address a quantile dependent prior for Bayesian quantile regression. We extend the idea of the power prior distribution in Bayesian quantile regression by employing the likelihood function that is based on a location-scale mixture representation of the asymmetric Laplace distribution. The propriety of the power prior is one of the critical issues in Bayesian analysis. Thus, we discuss the propriety of the power prior in Bayesian quantile regression. The methods are illustrated with both simulation and real data.

1. Introduction

Quantile regression models have been widely used for a variety of applications (Koenker [1]; Yu et al. [2]). Like standard or mean regression models, dealing with parameter and model uncertainty as well as updating information is of great importance for quantile regression and application. Since Yu and Moyeed [3] Bayesian inference quantile regression has attracted a lot of attention in the literature (Hanson and Johnson [4]; Tsionas [5]; Scaccia and Green [6]; Schennach [7]; Dunson and Taylor [8]; Geraci and Bottai [9]; Taddy and Kottas [10]; Yu and Stander [11]; Kottas and Krnjajić [12]; Lancaster and Jun [13]). These Bayesian inference models include Bayesian parametric, Bayesian semiparametric as well as Bayesian nonparametric models. However, almost all these models set priors independent of the values of quantiles, or the prior is the same for modelling different quantiles. This approach may result in inflexibility in quantile modelling. For example, a 95% quantile regression model should have different parameter values from the median quantile, and thus the priors used for modelling the quantiles should be different. It is therefore more reasonable to set different priors for different quantiles. In this paper, we address a quantile dependent prior for Bayesian quantile regression. Our idea is to set priors based on historical data. Although one can use improper prior in Bayesian quantile regression, the inference on current data could be more reliable and sensitive if there exist historical data gathered from similar previous studies. There are several methods to incorporate the historical data in the analysis of a current study. One of these methods is the power prior proposed by Ibrahim and Chen [14] which is constructed by raising the likelihood function of the historical data to a power parameter between 0 and 1. The power parameter represents the proportion of the historical data needed in the current study. The a priori idea for the power prior distribution belongs to Diaconis and Ylvisaker [15] and Morris [16] who studied conjugate priors for the exponential families, where they considered the power parameter as fixed constant which can be determined in advance. Ibrahim and Chen [14] developed this idea and considered the uncertainty case of the power parameter. They applied it in generalized linear mixed models, semiparametric proportional hazards models, and cure rate models for survival data. Chen et al. [17] examined the theoretical properties of power prior distribution for generalized linear models, while Ibrahim et al. [18] studied the optimality properties of the power prior, and Chen and Ibrahim [19] studied the relation between the power prior and hierarchical models and provided a formal justification of the power prior by examining formal analytical relationships between the power prior and hierarchical modelling in linear models.

Following the standard setup and notation for the power prior by Ibrahim and Chen [14], suppose that there exist historical data gathered from previous studies similar to the current study denoted by $D_0 = (n_0, y_0, x_0)$ along with a precision parameter a_0, $0 \le a_0 \le 1$, where n_0 denotes the sample size of the historical data, y_0 is an $n_0 \times 1$ historical data response vector, and $x'_{0i} = (1, x_{0i1}, x_{0i2}, \ldots, x_{0in})$ represent the $k+1$ known covariates from the historical data. The power parameter a_0; represents how much data from the previous study is to be used in the current study. There are two special cases for a_0; the first case $a_0 = 0$ corresponds to no incorporation of the data from previous study relative to the current study. The second case $a_0 = 1$ corresponds to full incorporation of the data from previous study relative to the current study. Therefore, a_0 controls the influence of the data gathered from previous studies that is similar to the current study; such control is important when the sample size of the current data is quite different from the sample size of the historical data or where there is heterogeneity between two studies (Ibrahim and Chen [14]). In generalized linear models, Ibrahim and Chen [14] defined the power prior of unknown parameters β based on the historical data as

$$\pi(\beta \mid D_0, a_0) \propto \left[L(\beta \mid D_0)\right]^{a_0} \pi_0(\beta \mid c_0), \tag{1.1}$$

where c_0 is a specified hyperparameter for the initial prior. Formulation (1.1) was initially elicited for a_0 as known parameter which can be determined previously, for example, by using expert beliefs or via a meta-analytic approach. Ibrahim and Chen [14] extend this idea by treating a_0 as random that is why the formulation becomes quite complicated. However, a random a_0 gives the researcher more freedom and flexibility in weighting the data gathered from previous studies. Thus Ibrahim and Chen [14] proposed a joint power prior distribution for (β, a_0) in generalized linear model of the form

$$\pi(\beta, a_0 \mid D_0) \propto \left[L(\beta \mid D_0)\right]^{a_0} \pi_0(\beta \mid c_0) \pi(a_0 \mid \gamma_0), \tag{1.2}$$

where c_0 and γ_0 are specified hyperparameter vectors. Power priors (1.1) and (1.2) will not have a closed form in general; however Ibrahim and Chen [14] suggested using a uniform prior for $\pi_0(\beta \mid c_0)$ and a beta prior for $\pi(a_0 \mid \gamma_0)$, or other choices, such as truncated normal or gamma priors. The advantage of employing these three priors for $\pi(a_0 \mid \gamma_0)$ is due to their similar theoretical and computational properties. Furthermore, the authors extend the original power prior to a situation where the set of covariates measured in the previous study is a subset from a set of covariates in the current data or when the historical data are not available. In addition they generalized power prior (1.2) to multiple data from previous studies, and power prior (1.2) becomes

$$\pi(\beta, a_0 \mid D_0) \propto \left\{ \prod_{j=1}^{M} \left[L(\beta \mid D_{0j}) \right]^{a_{0j}} \pi(a_{0j} \mid \gamma_0) \right\} \pi_0(\beta \mid c_0), \tag{1.3}$$

where M represent the size of previous studies, $a_0 = (a_{01}, \ldots, a_{0M})$, D_{0j} is the historical data for jth study, $j = 1, 2, \ldots, M$, and $D_0 = (D_{01}, \ldots, D_{0M})$.

Section 2 of the paper gives a brief overview of likelihood function based on asymmetric type of Laplace distribution, and we define the power prior for Bayesian quantile regression. In Section 3, we discuss the propriety of the power prior. In Section 4 we describe in detail the location-scale mixture of normal representation, and we propose power priors by using this representation for Bayesian quantile regression. Section 5 contains two simulation studies with one real data, and we end with a short discussion in Section 6.

2. The Power Prior

Consider the quantile linear regression model

$$y_i = x_i' \beta_p + \varepsilon_i, \tag{2.1}$$

where $\{(x_i, y_i),\ i = 1, 2, \ldots, n\}$ are independent observations, y_i is the response variable, $x_i' = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$ represent the $(k+1)$ known covariates, $\beta_p' = (\beta_{0(p)}, \beta_{1(p)}, \ldots, \beta_{k(p)})$ is the $(k+1)$ unknown parameters, and ε_i, $i = 1, \ldots, n$, represent error terms which are independent and identically distributed errors. The distribution of the error is assumed unknown and is restricted to have the pth quantile equal to zero and $0 < p < 1$. Let $q_p(y \mid x)$ represent the conditional quantile of y_i given x_i. Then the relation between $q_p(y \mid x)$ and x can be modelled as $q_p(y \mid x) = x_i' \beta_p$.

Following Yu and Moyeed [3], we suppose that ε_i has an asymmetric Laplace distribution with the density

$$f(\varepsilon \mid p) = p(1-p) \exp\{-\rho_p(\varepsilon)\}, \tag{2.2}$$

where

$$\rho_p(u) = \begin{cases} p|u| & \text{if } u \geq 0, \\ (1-p)|u| & \text{if } u < 0. \end{cases} \tag{2.3}$$

We refer to Kotz et al. [20] for a nice comprehensive review about the asymmetric Laplace distribution. The mean and variance of the asymmetric Laplace distribution are, respectively, given by

$$E(\varepsilon_i) = \frac{1-2p}{p(1-p)}, \qquad \mathrm{Var}(\varepsilon_i) = \frac{1-2p+2p^2}{p^2(1-p)^2}. \tag{2.4}$$

It is known that the probability density function of the asymmetric Laplace distribution of y_i given a location parameter $\mu_i = x_i'\beta_p$ is given by

$$f(y_i \mid \beta_p) = p(1-p)\exp\left\{-(y_i - x_i'\beta_p)\left\{p - I_{y_i \le x_i'\beta_p}\right\}\right\}. \tag{2.5}$$

Let $D = (n, y_i, x_i)$ denote the data from the current study. Then, the likelihood function for the current study is given by

$$\begin{aligned} f(\beta_p \mid D) &= p^n(1-p)^n \prod_{i=1}^{n} \exp\left\{-(y_i - x_i'\beta_p)\left\{p - I_{y_i \le x_i'\beta_p}\right\}\right\} \\ &= p^n(1-p)^n \exp\left\{-\sum_{i=1}^{n}(y_i - x_i'\beta_p)\left\{p - I_{y_i \le x_i'\beta_p}\right\}\right\}. \end{aligned} \tag{2.6}$$

Suppose that there exists historical data from a previous study denoted by $D_0 = (n_0, y_0, x_0)$ measuring the same response variable and covariates as the current study, where n_0 denotes the sample size of the previous study, y_0 is an $n_0 \times 1$ response vector of the previous study, and $x_i' = (1, x_{0i1}, x_{0i2}, \ldots, x_{0ik})$ represent the $k+1$ known covariates from the previous study. Then the likelihood function based on the data from the previous study is defined by

$$L(\beta_p \mid D_0) = p^{n_0}(1-p)^{n_0} \exp\left\{-\sum_{i=1}^{n_0}(y_i - x_{0i}'\beta_p)\left\{p - I_{y_{0i} \le x_{0i}'\beta_p}\right\}\right\}. \tag{2.7}$$

From Ibrahim and Chen [14] we define the joint prior distribution of β_p and a_0 for Bayesian quantile regression as

$$\pi(\beta_p, a_0 \mid D_0) \propto \left[L(\beta_p \mid D_0)\right]^{a_0} \pi_0(\beta_p \mid c_0)\pi(a_0 \mid \gamma_0), \tag{2.8}$$

where $L(\beta_p \mid D_0)$ is the likelihood function for the historical data for quantile regression which is given by (2.7). We assume that the initial prior for β_p is uniform. However, other choices, including multivariate normal or a double exponential can be used. Yu and Stander [11] prove that all posterior moments for β_p exist under these priors.

3. The Propriety of Power Prior Distribution in Quantile Regression

The power prior proposed by Ibrahim and Chen [14] has been constructed to be a useful class of informative prior in Bayesian analysis. This prior depends on the availability of the historical data, and in the context of Bayesian analysis when such data are available the prior distribution should be proper because it is well known that any informative Bayesian analysis requires a proper prior distribution; thus the propriety of the power prior is of critical importance. In this section we discuss the propriety of the power prior distribution in Bayesian quantile regression.

Theorem 3.1. *Suppose that the initial prior distribution for β_p is a uniform prior and a_0 has a beta prior with hyperparameters $(\delta_0 > 0,\ \lambda_0 > 0)$. Then, the joint prior distribution* (2.8) *in quantile regression for (β_p, a_0) is proper. In other words*

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \left[L(\beta_p \mid D_0)\right]^{a_0} a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} da_0 d\beta_p < \infty. \tag{3.1}$$

Proof. See the appendix. □

Corollary 3.2. *Suppose that the initial prior distribution for β_p is a uniform prior and the random variable a_0 has a uniform prior. Then, the joint power prior distribution* (2.8) *in quantile regression for (β_p, a_0) is proper. In other words*

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \left[L(\beta_p \mid D_0)\right]^{a_0} da_0 d\beta_p < \infty. \tag{3.2}$$

This corollary is derived directly from Theorem 3.1 *because the uniform distribution is the special case of the beta distribution when $(\delta_0 = 1,\ \lambda_0 = 1)$ and the proof is omitted.*

Corollary 3.3. *Suppose that the initial prior distribution for β_p is uniform prior and a_0 is constant. Then, power prior* (1.1) *in quantile regression for β_p is proper. In other words*

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[L(\beta_p \mid D_0)\right]^{a_0} d\beta_p < \infty. \tag{3.3}$$

This corollary is derived directly from Corollary 3.2*, and the proof is omitted. It is straightforward to verify that the joint prior $\pi(\beta_p, a_0 \mid D_0)$ when β_p has a uniform prior is always proper in quantile regression, which also ensures the proper propriety of the joint posterior of (β_p, a_0).*

Theorem 3.4. *Suppose that the initial prior distribution for β_p is assumed to be independent, and each $\pi_0(\beta_{i(p)} \mid c_0) \propto \exp\{-(1/\lambda_i)|\beta_{i(p)} - \mu_i|\}$, a double-exponential with fixed μ_i, $\lambda_i > 0$, and a_0 has a beta prior with hyperparameters (δ_0, λ_0). Then, the joint prior distribution* (2.8) *in quantile regression for (β_p, a_0) is proper.*

4. Mixture Representation

Consider the linear model for quantile regression (2.1), where the error term ε has an asymmetric Laplace distribution with the pth quantile equal to zero. The probability density function of the asymmetric Laplace distribution with location parameter μ and skewness parameter p, $p \in (0, 1)$ is given by (2.2).

It is well known that the asymmetric Laplace distribution (2.2) can be viewed as a mixture of an exponential and a scaled normal distribution (Reed and Yu [21] and Kotz et al. [20]). This can be recognized in the following lemma.

Lemma 4.1. *Suppose that X is a random variable that follows the asymmetric Laplace distribution with density* (2.2)*, ξ is a standard normal random variable, and z is a standard exponential random variable. Then, one can represent X as a location-scale mixture of normals given by*

$$X =^d \frac{1-2p}{p(1-p)} z + \sqrt{\frac{2z}{p(1-p)}}\xi. \tag{4.1}$$

From this result we can equivalently represent the error term ε_i as a mixture of normal distributions, given by

$$\varepsilon_i = \theta z_i + \phi\sqrt{z_i}\xi_i, \tag{4.2}$$

where

$$\theta = \frac{1-2p}{p(1-p)}, \qquad \phi^2 = \frac{2}{p(1-p)}. \tag{4.3}$$

Following Reed and Yu [21], we assume that the conditional distribution of each y_i given z_i is normal with mean $x_i'\beta_p + \theta z_i$ and variance $\phi^2 z_i$ and the z_i given β_p are independent standard exponential variables. Letting $y' = (y_1, \ldots, y_n)$ and $z' = (z_1, \ldots, z_n)$, then, the joint density of (y, z) is given by

$$f(y, z \mid \beta_p) = \prod_{i=1}^{n} f(y_i \mid \beta_p, z_i)\pi(z_i \mid \beta_p), \tag{4.4}$$

$$\begin{aligned} f(y, z \mid \beta_p) &\propto \prod_{i=1}^{n}\left(z_i^{-1/2} \exp\left\{ -\frac{(y_i - x_i'\beta_p - \theta z_i)^2}{2\phi^2 z_i} \right\} \exp\{-z_i\} \right) \\ &= \left(\prod_{i=1}^{n} z_i^{-1/2} \right) \exp\left\{ -\sum_{i=1}^{n} \frac{(y_i - x_i'\beta_p - \theta z_i)^2}{2\phi^2 z_i} \right\} \exp\left\{ -\sum_{i=1}^{n} z_i \right\}. \end{aligned} \tag{4.5}$$

We then integrate out the exponential variable z, which leads to the likelihood $f(y \mid \beta_p)$, where

$$f(y \mid \beta_p) = \int f(y, z \mid \beta_p) dz. \tag{4.6}$$

4.1. The Power Prior for Mixture Representation

Suppose that we are interested in making inference about β_p on the normal distribution with unknown variance, by incorporating both the previous and current studies.

Following the standard setup and notation for the power prior distribution for mixture representation, we assume that only one historical data set exists, and it is given by $D_0 = (n_0, y_0, x_0)$, where n_0 is the sample size of the historical data, y_0 is the $n_0 \times 1$ response vector, and x_0 is the $n_0 \times (k+1)$ matrix of covariates.

Let $z_0' = (z_{01}, \ldots, z_{0n_0})$, where $z_{01}, \ldots, z_{0n_0}$ are standard exponential random variables. As a mixture representation, the joint density for the historical data of y_{0i} given z_{0i} is normal with mean $x_{0i}'\beta_p + \theta z_{0i}$ and variance $\phi^2 z_{0i}$, and each z_{0i} given β_p is independently and identically standard exponential distribution, which can be viewed as the prior distribution on z_{0i}. For $\pi_0(\beta_p \mid c_0)$ we choose a normal density as initial prior with mean 0 and variance $B = c_0 I$, that is, $\pi_0(\beta_p \mid c_0) \propto \exp(-(1/2c_0)\beta_p'\beta_p)$. The purpose of this choice is due to the fact that all posterior moments of β_p exist under the above prior as provided in the studies of Yu and stander [11]. It is also convenient if all covariates are measured on the same scale parameter. As a special case one may choose a uniform improper prior which is special case of beta distribution when $(\delta_0 = 1, \lambda_0 = 1)$ for $\pi_0(\beta_p \mid c_0)$, that is, $\pi_0(\beta_p|c_0) \propto 1$; this corresponds to $c_0 \to \infty$, and this choice is very convenient with the partially Gibbs sampler as provided by Reed and Yu [21]. We propose a prior distribution of β_p taking the form

$$\pi(\beta_p \mid D_0, a_0) \propto \left\{ \prod_{i=1}^{n_0} \int_{z_{0i}} \left[f(y_{0i} \mid \beta_p, z_{0i}) \right]^{a_0} \pi(z_{0i} \mid \beta_p) dz_{0i} \right\} \pi_0(\beta_p \mid c_0), \tag{4.7}$$

where $f(y_{0i} \mid \beta_p, z_{0i})$ and $f(z_{0i} \mid \beta_p)$ are the same $f(y_i \mid \beta_p, z_i)$ and $f(z_i \mid \beta_p)$ in (4.4) with (y_{0i}, z_{0i}) in place of (y_i, z_i) to represent the historical data. Since we view a_0 as a random quantity, the prior specification is completed by specifying a prior distribution for a_0. We take a beta prior for a_0 with parameter (δ_0, λ_0), or one may choose a uniform prior. Thus we propose a joint prior distribution for β_p and a_0 of the form

$$\pi(\beta_p, a_0 \mid D_0) \propto \left\{ \prod_{i=1}^{n_0} \int_{z_{0i}} \left[f(y_{0i} \mid \beta_p, z_{0i}) \right]^{a_0} \pi_0(z_{0i} \mid \beta_p) dz_{0i} \right\} \pi_0(\beta_p \mid c_0) \pi(a_0 \mid \gamma_0), \tag{4.8}$$

$$\begin{aligned} &\propto \prod_{i=1}^{n_0} \int_{z_{0i}} \left(z_{0i}^{-1/2} \exp\left\{ -a_0 \frac{(y_{0i} - x_{0i}'\beta_p - \theta z_{0i})^2}{2\phi^2 z_{0i}} \right\} \exp\{-z_{0i}\} dz_{0i} \right) \\ &\quad \times \exp\left\{ -\frac{1}{2c_0} \beta_p'\beta_p \right\} \times a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1}. \end{aligned} \tag{4.9}$$

We see that (4.8) will not have a closed form in general because it depends on the initial priors that we choose. Thus the joint posterior distribution of β_p and a_0 is given by

$$p(\beta_p, a_0 \mid D, D_0) \propto \left[\prod_{i=1}^{n} f(y_i \mid \beta_p, z_i) \right] \pi(\beta_p, a_0 \mid D_0). \tag{4.10}$$

Power prior (4.8) is constructed for one historical data, and this power prior can be easily generalized to multiple historical data. To generalized power prior (4.8) to multiple historical data, we assume that there are M historical studies denoted by $D_0 = (D_{01}, \ldots, D_{0M})$, where $D_{0j} = (n_{0j}, y_{0j}, x_{0j})$ represent the historical data based on the j study, $j = 1, \ldots, M$. Let $z_{0j}' = (z_{01j}, \ldots, z_{0n_{0j}})$, where $z_{01j}, \ldots, z_{0n_{0j}}$ are standard exponential random variables. We define a_{0j}

to be the power parameter for the jth study with beta prior distribution. Hence, the prior can be generalized as

$$\pi(\beta_p, a_0 \mid D_0) \propto \prod_{j=1}^{M}\left\{\prod_{i=1}^{n_{0j}}\int_{z_{0ij}}\left[f(y_{0ij} \mid \beta_p, z_{0ij})\right]^{a_{0j}}\pi_0(z_{0ij} \mid \beta_p)dz_{0ij}\right\} \times \pi_0(\beta_p \mid c_0)\pi(a_{0j} \mid \gamma_0), \tag{4.11}$$

where $a_0 = (a_{01}, \ldots, a_{0M})$, and each a_{0j} has a beta prior with the same hyperparameters (δ_0, λ_0).

4.2. Inference with Scale Parameter

In the previous section, we have considered the power prior distribution in quantile regression model without taking into account a scale parameter. One may be interested to introduce a scale parameter into the model for the proposed Bayesian inference. Suppose that $\tau > 0$ is the scale parameter. From now on, it is more convenient to work with $v_i = \tau z_i$ for the current data and with $v_{0i} = \tau z_{0i}$ for the historical data. We assume that only one historical data set exists, and it is given by $D_0 = (n_0, y_0, x_0)$. Let $v_0' = (v_{01}, \ldots, v_{0n_0})$. Then, the conditional distribution for each y_{0i} given v_{0i}, β_p, and τ is normal with mean $x_{0i}'\beta_p + \theta v_{0i}$ and variance $\tau\phi^2 v_{0i}$, that is, $y_{0i} \mid v_{0i}, \beta_p, \tau \sim N(x_{0i}'\beta_p + \theta v_{0i}, \tau\phi^2 v_{0i})$, and the v_{0i} given β_p and τ are independent and identically distributed exponential variables with rate parameter τ. The conditional distribution of v_{0i} given β_p and τ can be viewed as prior distribution on v_{0i}. It will be more convenient to work with the following priors:

$$\tau \sim \Gamma(l_0, s_0), \qquad \beta_p \mid \tau \sim N_k(0, B_0), \quad B_0 = c_0 I, \quad c_0 \longrightarrow \infty, \tag{4.12}$$

where l_0, s_0, and B_0 are known parameters. For a_0 we take a beta prior with parameter (δ_0, λ_0). Now, the specification of the power prior distribution is completed, and thus we propose a joint prior distribution for β_p, τ, and a_0 of the form

$$\pi(\beta_p, \tau, a_0 \mid D_0) \propto \left\{\prod_{i=1}^{n_0}\int_{v_{0i}}\left[f(y_{0i} \mid \beta_p, \tau, v_{0i})\right]^{a_0}\pi_0(v_{0i} \mid \beta_p, \tau)dv_{0i}\right\} \times \pi_0(\beta_p \mid c_0)\pi(\tau)\pi(a_0 \mid \gamma_0), \tag{4.13}$$

$$\propto \left\{\prod_{i=1}^{n_0}\int_{v_{0i}}(\tau v_{0i})^{-1/2}\exp\left\{-a_0\frac{(y_{0i} - x_{0i}'\beta_p - \theta v_{0i})^2}{2\phi^2\tau v_{0i}}\right\}\tau\exp\{-\tau v_{0i}\}dv_{0i}\right\} \times \exp\left\{-\frac{1}{2c_0}\beta_p'\beta_p\right\} \times (\tau)^{l_0-1}\exp\{-s_0\tau\}a_0^{\delta_0-1}(1-a_0)^{\lambda_0-1}. \tag{4.14}$$

Then, the joint posterior distribution of β_p, τ, and a_0 is given by

$$p(\beta_p, \tau, a_0 \mid D, D_0) \propto \left[\prod_{i=1}^{n} f(y_i \mid \beta_p, \tau, v_i)\right] \pi(\beta_p, \tau, a_0 \mid D_0). \tag{4.15}$$

Power prior (4.13) can be easily generalized to M historical data, and the generalized distribution can be given as

$$\pi(\beta_p, \tau, a_0 \mid D_0) \propto \prod_{j=1}^{M} \left\{ \prod_{i=1}^{n_{0j}} \int_{v_{0ij}} [f(y_{0ij} \mid \beta_p, \tau, v_{0ij})]^{a_{0j}} \pi_0(v_{0ij} \mid \beta_p, \tau) dv_{0ij} \right\} \times \pi_0(\beta_p \mid c_0) \pi(\tau) \pi(a_{0j} \mid \gamma_0). \tag{4.16}$$

5. Numerical Examples

In this section, our aim is to compare the posterior means of parameters of interest after incorporating the current and historical data with the mean of true values for both studies. In addition, we will demonstrate the behaviour of the prior under several choices of prior parameters.

Example 5.1. We simulate two data sets, the first one for the current study and the second for the previous study. For the current study we generate 100 observations from the model $y_i = \mu + \varepsilon_i$ assuming that $\mu = 5.0$ and $\varepsilon_i \sim N(0, 1)$.

For the historical data we use the same model with 50 observations and $\mu = 6.0$. In this example we have used only one parameter μ. Table 1 compares the posterior means with the means of true values for $q_p(y_i) = \beta_p$ at 5 different quantiles, namely, 90%, 75%, 50%, 25%, and 10%. We conduct sensitive analysis with respect to five different choices for (δ_0, λ_0) for five different quantiles. For computation we construct a Markov chain via the Metropolis-Hastings (MH) algorithm. We ran the algorithm for 15000 iterations and discarded the first 5000 as burn in. Figures 1, 2, and 3 compare the posterior densities of β_p for $p = 0.90, 0.50$, and 0.10, respectively, for improper prior with the posterior densities of β_p for the power prior with parameters $(\mu_{a_0}, \sigma_{a_0}) = (0.50, 0.078)$ and $(\mu_{a_0}, \sigma_{a_0}) = (0.99, 0.010)$. Clearly, the power prior is more informative than improper prior, due to the small range of posterior densities.

Note that as shown in Chen et al. [17] it is easier to specify the prior mean and standard deviation of a_0 from the following equations:

$$\mu_{a_0} = \frac{\delta_0}{(\delta_0 + \lambda_0)}, \qquad \sigma_{a_0} = \left(\mu_{a_0}(1 - \mu_{a_0})\right)^{1/2} (\delta_0 + \lambda_0 + 1)^{-1/2}. \tag{5.1}$$

Furthermore they have shown that the investigator must choose μ_{a_0} small if he/she wishes low weight to the historical data and must choose $\mu_{a_0} \geq 0.5$ if he/she wishes more weight to the historical data.

In this example we use power prior (2.8), taking uniform prior for β_p and beta prior for a_0. Under specific quantile level, we see that as the weight for the historical data increases the

Table 1: Posterior means, posterior standard deviations (SD), and mean of the true values of $\beta_{(p)}$.

p	(δ_0, λ_0)	$(\mu_{a_0}, \sigma_{a_0})$	Mean $\beta_{(p)}$	SD $\beta_{(p)}$	Mean of the true values of $\beta_{(p)}$
0.90	(5,5)	(0.50, 0.151)	6.410	0.2348	**6.7816**
	(20,20)	(0.50, 0.078)	6.735	0.2514	
	(30,30)	(0.50, 0.064)	6.776	0.2326	
	(50,1)	(0.98, 0.019)	6.837	0.2311	
	(100,1)	(0.99, 0.010)	**6.843**	0.2260	
0.75	(5,5)	(0.50, 0.151)	5.771	0.1563	**6.1745**
	(20,20)	(0.50, 0.078)	5.991	0.1692	
	(30,30)	(0.50, 0.064)	6.025	0.1668	
	(50, 1)	(0.98, 0.019)	6.094	0.1635	
	(100,1)	(0.99, 0.010)	**6.109**	0.1609	
0.50	(5,5)	(0.50, 0.151)	5.097	0.1559	**5.5000**
	(20,20)	(0.50, 0.078)	5.273	0.1477	
	(30,30)	(0.50, 0.064)	5.316	0.1451	
	(50,1)	(0.98, 0.019)	5.382	0.1424	
	(100,1)	(0.99, 0.010)	**5.383**	0.1411	
0.25	(5,5)	(0.50, 0.151)	4.466	0.1622	**4.8255**
	(20,20)	(0.50, 0.078)	4.600	0.1464	
	(30,30)	(0.50, 0.064)	4.614	0.1607	
	(50,1)	(0.98, 0.019)	4.645	0.1523	
	(100,1)	(0.99, 0.010)	**4.645**	0.1437	
0.10	(5,5)	(0.50, 0.151)	3.911	0.2250	**4.2185**
	(20,20)	(0.50, 0.078)	3.993	0.2066	
	(30,30)	(0.50, 0.064)	4.019	0.2014	
	(50, 1)	(0.98, 0.019)	4.038	0.1990	
	(100,1)	(0.99, 0.010)	**4.053**	0.1965	

posterior mean of β_p increases. This is a comforting feature because it is consistent with what we expect from the data. This implies that the posterior mean for the parameters of interest is quite robust for the different weights for power parameter.

More noticeably, when $(\delta_0 = 100, \lambda_0 = 1)$, that is, we give more weight to the historical data, we see that the posterior mean is very close to the mean of the true values. In addition, under specific quantile level, we found that as the weight for the historical data increases the standard deviation tends to decrease.

Example 5.2. For a mixture representation with scale parameter, we simulate two data sets, the first one for the current study and the second for the previous study. For the current study we generate a data set of $n = 50$ observations from the model $y_i = \beta_{0(p)} + \beta_{1(p)} x_i + 1/11(11 + x_i)\varepsilon_i$, where x_i are random uniform numbers on the interval $(0, 10)$ and $\varepsilon_i \sim N(0,1)$. We restricted

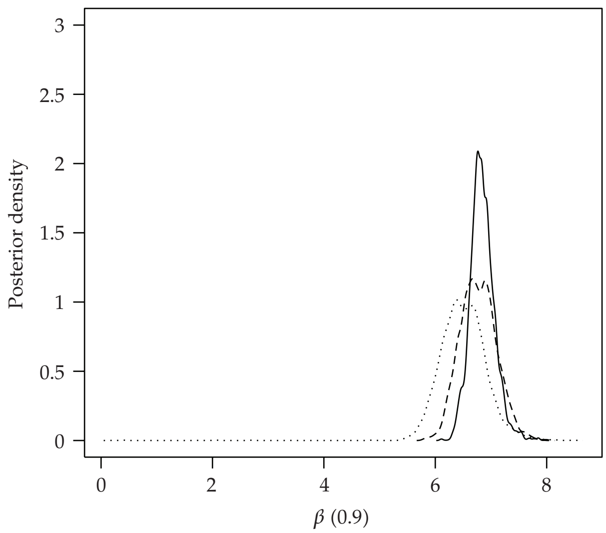

Figure 1: Plots of posterior densities for $\beta_{0.90}$, where the dotted curve is for improper uniform prior, the dashed and solid curves are for power priors with parameters $(\mu_{a_0}, \sigma_{a_0}) = (0.50, 0.078)$ and $(\mu_{a_0}, \sigma_{a_0}) = (0.99, 0.010)$, respectively.

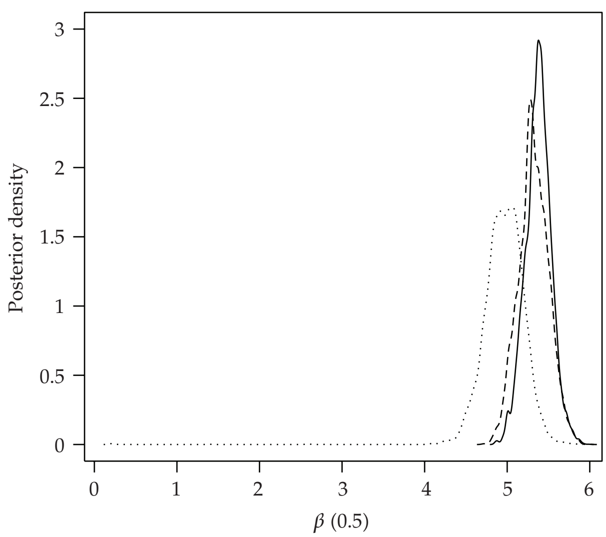

Figure 2: Plots of posterior densities for $\beta_{0.50}$, where the dotted curve is for improper uniform prior, the dashed and solid curves are for power priors with parameters $(\mu_{a_0}, \sigma_{a_0}) = (0.50, 0.078)$ and $(\mu_{a_0}, \sigma_{a_0}) = (0.99, 0.010)$, respectively.

$\beta_{0(p)} = 10$ and $\beta_{1(p)} = -1$. For the previous study we generate $n_0 = 150$ observations from the above model with $\beta_{0(p)} = 9$ and $\beta_{1(p)} = -1.2$.

We use initial prior $N(0, 10^6)$ on all regression parameters and $\Gamma(10^{-3}, 10^{-3})$ on all scale parameters. Then we ran MCMC algorithm for 11000 iterations and discarded the first 1000 as burn in. We then compute the posterior means of the parameters at 5 different quantiles, namely, 90%, 75%, 50%, 25%, and 10%. We conduct sensitive analysis with respect to five

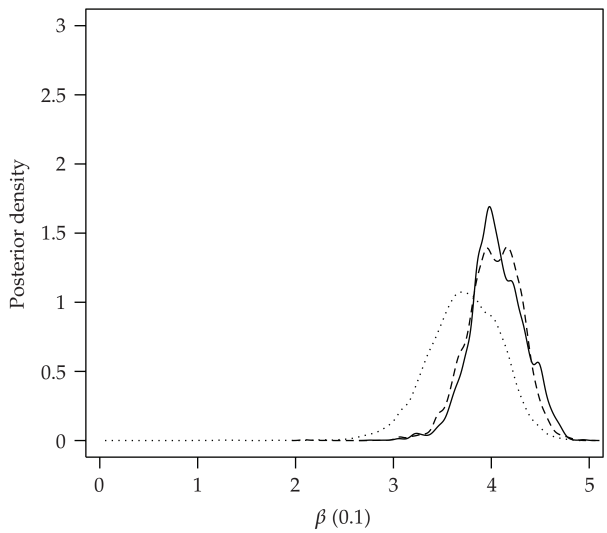

Figure 3: Plots of posterior densities for $\beta_{0.10}$, where the dotted curve is for improper uniform prior, the dashed and solid curves are for power priors with parameters $(\mu_{a_0}, \sigma_{a_0}) = (0.50, 0.078)$ and $(\mu_{a_0}, \sigma_{a_0}) = (0.99, 0.010)$, respectively.

different weights for the power parameter, namely, 10%, 25%, 50%, 75%, and 90%. The results are summarized in Table 2. Based on the results in Table 2 for each quantile, it is consistent in the sense that the posterior mean of β_p either increases or decreases steadily as the weight of the historical data increases. Under specific quantile level, we also found that as the weight for the historical data increases the posterior standard deviations for all parameters tend to decrease.

Example 5.3. We consider data from the British Household Panel Survey. The data were originally collected by the ESRC Research Centre on Microsocial Change at the University of Essex and analyzed by Yu et al. [22]. The data represent the wage distribution among British workers between 1991 and 2001. We use the data for the year 2000 as current data and for 1994 as historical data. Four covariates and intercept are included in the analysis. The relation between response variable and covariates are given by the following model:

$$\ln(Y_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 D_i + \varepsilon_i, \tag{5.2}$$

where S_i is the number of years of schooling, E_i is the potential experience (approximated by the age minus years of schooling minus 6), and D_i is equal to 1 for public sector workers and 0 otherwise. In this example we fixed the power parameter at five weights, namely, 0.10, 0.25, 0.50, 0.75, and 0.90. The results are summarized in Table 3. From Table 3, we see that as the weight for the historical data increases, the posterior mean for each regression coefficient either decreases or increases. We also found that as the weight for the historical data increases, the posterior standard deviations for all parameters tend to decrease.

Table 2: Posterior means, posterior standard deviations (SD), and mean of the true values of $\beta_{(p)}$.

p	a_0	Mean $\beta_{0(p)}$	SD $\beta_{0(p)}$	Mean of the true values of $\beta_{0(p)}$	Mean $\beta_{1(p)}$	SD $\beta_{1(p)}$	Mean of the true values of $\beta_{1(p)}$
0.90	0.10	10.2190	0.4731	10.7816	−1.1840	0.1042	−0.9835
	0.25	10.2550	0.2960		−1.1738	0.0591	
	0.50	10.5200	0.1573		−1.1569	0.0315	
	0.75	10.7500	0.2127		−1.1060	0.0332	
	0.90	10.9400	0.1311		−1.0743	0.0194	
0.75	0.10	9.7010	0.3316	10.1745	−1.1911	0.0611	−1.0387
	0.25	9.7030	0.2934		−1.1869	0.0639	
	0.50	9.7930	0.2214		−1.1710	0.0455	
	0.75	10.0100	0.1852		−1.1680	0.0333	
	0.90	10.1620	0.1636		−1.1652	0.0301	
0.50	0.10	9.2095	0.2414	9.5000	−1.1938	0.0275	−1.1000
	0.25	9.2560	0.1952		−1.1957	0.0233	
	0.50	9.2600	0.1046		−1.1958	0.0176	
	0.75	9.2885	0.0871		−1.1968	0.0143	
	0.90	9.3080	0.0735		−1.1971	0.0112	
0.25	0.10	9.2820	0.3552	8.8255	−1.2590	0.0718	−1.1613
	0.25	9.1890	0.2489		−1.2650	0.0462	
	0.50	8.9910	0.1841		−1.2690	0.0340	
	0.75	8.8230	0.1660		−1.2760	0.0313	
	0.90	8.7270	0.1492		−1.2810	0.0279	
0.10	0.10	8.8240	0.3272	8.2184	−1.1940	0.0640	−1.2165
	0.25	8.6460	0.2171		−1.1920	0.0433	
	0.50	8.3880	0.1556		−1.2030	0.0292	
	0.75	8.1900	0.1723		−1.2430	0.0315	
	0.90	8.0980	0.1171		−1.2600	0.0256	

6. Discussion

In this paper, we have demonstrated the use of power prior in Bayesian quantile regression that incorporates both historical and current data. The advantage of the method is that the prior distribution is changing automatically when we change the quantile. Thus, we have prior distribution for each quantile, and the prior is proper. In addition, we proposed joint prior distributions using a mixture of normal representation of the asymmetric Laplace distribution. The behavior of the power prior is clearly quite robust with different weights for power parameter. We use random power parameter in the first example that can be determined via the hyperparameters of beta distribution, and we compare the posterior

Table 3: Posterior means of $\beta_{(p)}$ for the real data. In the parentheses are standard deviations of β_p.

p	a_0	Mean $\beta_{0(p)}$	Mean $\beta_{1(p)}$	Mean $\beta_{2(p)}$	Mean of $\beta_{3(p)}$	Mean of $\beta_{4(p)}$
0.90	0.10	7.2114 (0.432)	0.0237 (0.035)	0.0201 (0.019)	−0.0005 (0.017)	−0.1036 (0.021)
	0.25	7.3455 (0.441)	0.0240 (0.039)	0.0193 (0.013)	−0.0002 (0.021)	−0.0900 (0.019)
	0.50	7.3701 (0.357)	0.0212 (0.031)	0.0109 (0.011)	−0.0002 (0.018)	−0.0864 (0.013)
	0.75	7.3704 (0.332)	0.0210 (0.027)	0.0109 (0.009)	−0.0001 (0.014)	−0.0819 (0.012)
	0.90	7.3732 (0.263)	0.0201 (0.022)	0.0106 (0.009)	−0.0001 (0.012)	−0.0827 (0.007)
0.75	0.10	6.8264 (0.337)	0.0231 (0.026)	0.0252 (0.013)	−0.0005 (0.015)	−0.0455 (0.027)
	0.25	7.0158 (0.227)	0.0228 (0.011)	0.0252 (0.019)	−0.0001 (0.014)	−0.0328 (0.022)
	0.50	7.0173 (0.316)	0.0216 (0.011)	0.0159 (0.012)	−0.0004 (0.010)	−0.0145 (0.017)
	0.75	7.0408 (0.216)	0.0203 (0.010)	0.0117 (0.010)	−0.0004 (0.011)	−0.0097 (0.016)
	0.90	7.0391 (0.117)	0.0191 (0.010)	0.0112 (0.008)	−0.0004 (0.011)	−0.0085 (0.013)
0.5	0.10	6.3933 (0.221)	0.0269 (0.013)	0.0354 (0.018)	−0.0008 (0.022)	0.0137 (0.024)
	0.25	6.7117 (0.117)	0.0250 (0.009)	0.0306 (0.013)	−0.0006 (0.020)	0.0471 (0.019)
	0.50	6.7130 (0.113)	0.0149 (0.010)	0.0265 (0.012)	−0.0006 (0.017)	0.0487 (0.018)
	0.75	6.7163 (0.113)	0.0193 (0.008)	0.0110 (0.009)	−0.0002 (0.018)	0.0631 (0.016)
	0.90	6.7928 (0.105)	0.0136 (0.008)	0.0110 (0.009)	−0.0002 (0.012)	0.0633 (0.013)
0.25	0.10	6.2386 (0.328)	0.0216 (0.024)	0.0165 (0.019)	−0.0003 (0.019)	0.0794 (0.018)
	0.25	6.3479 (0.317)	0.0201 (0.029)	0.0162 (0.017)	−0.0002 (0.024)	0.0897 (0.016)
	0.50	6.3624 (0.306)	0.0177 (0.018)	0.0139 (0.023)	−0.0002 (0.018)	0.0921 (0.011)
	0.75	6.3703 (0.219)	0.0167 (0.015)	0.0146 (0.013)	−0.0002 (0.014)	0.0937 (0.009)
	0.90	6.3986 (0.201)	0.0142 (0.014)	0.0120 (0.012)	−0.0004 (0.013)	0.0937 (0.007)
0.1	0.10	5.8857 (0.357)	0.0200 (0.019)	0.0238 (0.025)	−0.0006 (0.017)	0.0766 (0.017)
	0.25	5.9255 (0.311)	0.0142 (0.018)	0.0301 (0.013)	−0.0007 (0.016)	0.1022 (0.023)
	0.50	5.9308 (0.299)	0.0114 (0.023)	0.0329 (0.011)	−0.0007 (0.015)	0.1239 (0.018)
	0.75	5.9550 (0.271)	0.0110 (0.014)	0.0302 (0.015)	−0.0006 (0.012)	0.1403 (0.018)
	0.90	5.9592 (0.248)	0.0095 (0.013)	0.0366 (0.012)	−0.0008 (0.012)	0.1496 (0.014)

mean of the intercept with the mean of true values. In the second example we show the behavior of the power prior distribution when the power parameter is a fixed parameter and can be determined using expert beliefs or via a meta-analytic approach, and we compare the posterior mean of parameter of interest with the mean of true values for both studies. In the third example, we also use fixed power parameter, and we compare the posterior mean for different weights for the historical data. The power prior is a very useful class of informative prior distribution for Bayesian quantile regression. It also seems to be useful in many applications such as model selection and carcinogenicity studies.

Appendix

Proof of Theorem 3.1

To prove the joint prior distribution is proper prior, that is,

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \left[L(\beta_p \mid D_0)\right]^{a_0} a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} da_0 d\beta_p < \infty, \tag{A.1}$$

note that

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \ln \left[L(\beta_p \mid D_0)\right]^{a_0} a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} da_0 d\beta_p \\
&\quad = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} -\sum_{i=1}^{n_0} (y_{0i} - x'_{0i}\beta_p) \left[p - I_{\{y_{0i} \le x'_{0i}\beta_p\}}\right] d\beta_p \int_{0}^{1} a_0 da_0 \\
&\qquad + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \ln \left[a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1}\right] da_0 d\beta_p \\
&\quad = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \ln \left\{ \exp \left\{ -\sum_{i=1}^{n_0} (y_{0i} - x'_{0i}\beta_p) \left[p - I_{\{y_{0i} \le x'_{0i}\beta_p\}}\right] \right\} \right\} \left(\frac{1}{2}\right) d\beta_p + K,
\end{aligned} \tag{A.2}$$

where

$$K = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \ln \left[a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1}\right] da_0 d\beta_p. \tag{A.3}$$

Then

$$\begin{aligned}
&\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \left[L(\beta_p \mid D_0)\right]^{a_0} a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} da_0 d\beta_p \\
&\quad = K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n_0} (y_{0i} - x'_{0i}\beta_p) \left[p - I_{\{y_{0i} \le x'_{0i}\beta_p\}}\right] \right\} \right\} d\beta_p.
\end{aligned} \tag{A.4}$$

Following Yu and Moyeed [3], this integral is finite:

$$0 < \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{1} \left[L(\beta_p \mid D_0)\right]^{a_0} a_0^{\delta_0 - 1} (1 - a_0)^{\lambda_0 - 1} da_0 d\beta_p < \infty. \tag{A.5}$$

Acknowledgments

The authors wish to thank Professor Tomasz J. Kozubowski and two anonymous referees for helpful comments and suggestions, which have led to an improvement of this paper.

References

[1] R. Koenker, *Quantile Regression*, vol. 38 of *Econometric Society Monographs*, Cambridge University Press, Cambridge, UK, 2005.

[2] K. Yu, Z. Lu, and J. Stander, "Quantile regression: applications and current research areas," *Journal of the Royal Statistical Society D. The Statistician*, vol. 52, no. 3, pp. 331–350, 2003.

[3] K. Yu and R. A. Moyeed, "Bayesian quantile regression," *Statistics & Probability Letters*, vol. 54, no. 4, pp. 437–447, 2001.

[4] T. Hanson and W. O. Johnson, "Modeling regression error with a mixture of Polya trees," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1020–1033, 2002.

[5] E. G. Tsionas, "Bayesian quantile inference," *Journal of Statistical Computation and Simulation*, vol. 73, no. 9, pp. 659–674, 2003.

[6] L. Scaccia and P. J. Green, "Bayesian growth curves using normal mixtures with nonparametric weights," *Journal of Computational and Graphical Statistics*, vol. 12, no. 2, pp. 308–331, 2003.

[7] S. M. Schennach, "Bayesian exponentially tilted empirical likelihood," *Biometrika*, vol. 92, no. 1, pp. 31–46, 2005.

[8] D. B. Dunson and J. A. Taylor, "Approximate Bayesian inference for quantiles," *Journal of Nonparametric Statistics*, vol. 17, no. 3, pp. 385–400, 2005.

[9] M. Geraci and M. Bottai, "Quantile regression for longitudinal data using the asymmetric Laplace distribution," *Biostatistics*, vol. 8, no. 1, pp. 140–154, 2007.

[10] M. Taddy and A. Kottas, "A nonparametric model-based approach to inference for quantile regression," Tech. Rep., UCSC Department of Applied Math and Statistics, 2007.

[11] K. Yu and J. Stander, "Bayesian analysis of a Tobit quantile regression model," *Journal of Econometrics*, vol. 137, no. 1, pp. 260–276, 2007.

[12] A. Kottas and M. Krnjajić, "Bayesian semiparametric modelling in quantile regression," *Scandinavian Journal of Statistics*, vol. 36, no. 2, pp. 297–319, 2009.

[13] T. Lancaster and S. J. Jun, "Bayesian quantile regression methods," *Journal of Applied Econometrics*, vol. 25, no. 2, pp. 287–307, 2010.

[14] J. G. Ibrahim and M.-H. Chen, "Power prior distributions for regression models," *Statistical Science*, vol. 15, no. 1, pp. 46–60, 2000.

[15] P. Diaconis and D. Ylvisaker, "Conjugate priors for exponential families," *The Annals of Statistics*, vol. 7, no. 2, pp. 269–281, 1979.

[16] C. N. Morris, "Natural exponential families with quadratic variance functions: statistical theory," *The Annals of Statistics*, vol. 11, no. 2, pp. 515–529, 1983.

[17] M.-H. Chen, J. G. Ibrahim, and Q.-M. Shao, "Power prior distributions for generalized linear models," *Journal of Statistical Planning and Inference*, vol. 84, no. 1-2, pp. 121–137, 2000.

[18] J. G. Ibrahim, M.-H. Chen, and D. Sinha, "On optimality properties of the power prior," *Journal of the American Statistical Association*, vol. 98, no. 461, pp. 204–213, 2003.

[19] M.-H. Chen and J. G. Ibrahim, "The relationship between the power prior and hierarchical models," *Bayesian Analysis*, vol. 1, no. 3, pp. 551–574, 2006.

[20] S. Kotz, T. J. Kozubowski, and K. Podgórski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Financ*, Birkhäuser, Boston, Mass, USA, 2001.

[21] C. Reed and K. Yu, "A Partially collapsed Gibbs sampler for Bayesian quantile regression," Tech. Rep., Department of Mathematical Sciences, Brunel University, 2009.

[22] K. Yu, P. Van Kerm, and J. Zhang, "Bayesian quantile regression: an application to the wage distribution in 1990s Britain," *Sankhyā*, vol. 67, no. 2, pp. 359–377, 2005.