ELSEVIER

Statistics & Probability Letters 54 (2001) 437–447

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Bayesian quantile regression

Keming Yu, Rana A. Moyeed*

Department of Mathematics and Statistics, University of Plymouth, Drake circus, Plymouth PL4 8AA, UK

Received January 2001; received in revised form April 2001

Abstract

The paper introduces the idea of Bayesian quantile regression employing a likelihood function that is based on the asymmetric Laplace distribution. It is shown that irrespective of the original distribution of the data, the use of the asymmetric Laplace distribution is a very natural and effective way for modelling Bayesian quantile regression. The paper also demonstrates that improper uniform priors for the unknown model parameters yield a proper joint posterior. The approach is illustrated via a simulated and two real data sets. © 2001 Elsevier Science B.V. All rights reserved

Keywords: Asymmetric Laplace distribution; Bayesian inference; Markov chain Monte Carlo methods; Quantile regression

1. Introduction

The classical theory of linear models is essentially a theory for models of conditional expectations. In many applications, however, it is fruitful to go beyond these models. Quantile regression is gradually emerging as a comprehensive approach to the statistical analysis of linear and nonlinear response models. A number of papers have recently appeared on the application of quantile regression (see Cole and Green, 1992; Royston and Altman, 1994; Buchinsky, 1998; Yu and Jones, 1998; He et al., 1998; Koenker and Machado, 1999). Quantile regression supplements the exclusive focus of least squares based methods on the estimation of conditional mean functions with a general technique for estimating families of conditional quantile functions. This greatly expands the flexibility of both parametric and nonparametric regression methods.

In strictly linear models, a simple approach to estimating the conditional quantiles is suggested in Koenker and Bassett (1978). Consider the following standard linear model:

$$y_t = \mu(\boldsymbol{x}_t) + \varepsilon_t,$$

where $\mu(\boldsymbol{x}_t)$ may be thought of as the conditional mean of y_t given the vector of regressors $\boldsymbol{x}_t$, and ε_t is the error term with mean zero and constant variance. It is not necessary to specify the distribution of the error

* Corresponding author.
E-mail address: r.moyeed@plymouth.ac.uk (R.A. Moyeed).

PII: S0167-7152(01)00124-9

term as it is allowed to take any form. Typically,

$$\mu(\boldsymbol{x}_t) = \boldsymbol{x}_t' \boldsymbol{\beta}$$

for a vector of coefficients $\boldsymbol{\beta}$. The pth $(0 < p < 1)$ quantile of ε_t is the value, q_p, for which $P(\varepsilon_t < q_p) = p$. The pth conditional quantile of y_t given $\boldsymbol{x}_t$ is then simply

$$q_p(y_t|\boldsymbol{x}_t) = \boldsymbol{x}_t'\boldsymbol{\beta}(p), \tag{1}$$

where $\boldsymbol{\beta}(p)$ is a vector of coefficients dependent on p.

The pth *regression quantile* $(0 < p < 1)$ is defined as any solution, $\hat{\boldsymbol{\beta}}(p)$, to the *quantile regression* minimisation problem

$$\min_{\boldsymbol{\beta}} \sum_t \rho_p(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}), \tag{2}$$

where the *loss function*

$$\rho_p(u) = u(p - I(u < 0)). \tag{3}$$

Equivalently, we may write (3) as

$$\rho_p(u) = u(pI(u > 0) - (1 - p)I(u < 0)),$$

or

$$\rho_p(u) = \frac{|u| + (2p - 1)u}{2}.$$

Contrary to the commonly used quadratic loss function for mean regression estimation, the quantile regression links to a special class of loss functions which has robust properties (see Huber, 1981). For example, the conditional median regression is more robust than the conditional mean regression in terms of outliers in the observations.

The use of Bayesian inference in generalized linear and additive models is quite standard these days. The relative ease with which Markov chain Monte Carlo (MCMC) methods may be used for obtaining the posterior distributions, even in complex situations, has made Bayesian inference very useful and attractive. Unlike conventional methods, Bayesian inference provides one with the entire posterior distribution of the parameter of interest. In addition, it allows for parameter uncertainty to be taken into account when making predictions. However, in the area of quantile regression, there is very little in the literature along the Bayesian lines, the only exception being Fatti and Senaoana (1998).

In this paper, we adopt a Bayesian approach to quantile regression which is quite different from what has been done previously in this context. Irrespective of the actual distribution of the data, Bayesian inference for quantile regression proceeds by forming the likelihood function based on the asymmetric Laplace distribution (Koenker and Bassett, 1978). In general one may choose any prior, but we show that the use of improper uniform priors produces a proper joint posterior.

Section 2 of the paper, gives a brief overview of the asymmetric Laplace distribution. Section 3 describes the theoretical and computational framework of quantile regression from a Bayesian perspective. Section 4 shows that the use of improper priors for the quantile regression parameters leads to proper posterior distributions. In Section 5, we illustrate the implementation of Bayesian quantile regression through a simulated case-study and two real examples. We end with a short discussion in Section 6.

2. Asymmetric Laplace distribution

It can be easily shown that the minimization of the loss function (3) is exactly equivalent to the maximization of a likelihood function formed by combining independently distributed asymmetric Laplace densities. Let us recall the properties of the asymmetric Laplace distribution. A random variable U is said to follow the asymmetric Laplace distribution if its probability density is given by

$$f_p(u) = p(1-p)\exp\{-\rho_p(u)\}, \tag{4}$$

where $0 < p < 1$ and $\rho_p(u)$ is as defined in (3). When $p = 1/2$, (4) reduces to $1/4\exp(-|u|/2)$, which is the density function of a standard symmetric Laplace distribution. For all other values of p, the density in (4) is asymmetric. The mean of U is $(1-2p)/p(1-p)$ and it is positive only for $p > 1/2$. The variance given by $(1-2p+2p^2)/p^2(1-p)^2$ increases quite rapidly as p approaches 0 or 1.

One could also incorporate location and scale parameters μ and σ, respectively, in the density in (4) to obtain

$$f_p(u;\mu,\sigma) = \frac{p(1-p)}{\sigma}\exp\left\{-\rho_p\left(\frac{u-\mu}{\sigma}\right)\right\}.$$

3. Bayesian quantile regression

In the conventional generalized linear model, the estimates of the unknown regression parameters $\boldsymbol{\beta}$, as introduced in Section 1, are obtained by assuming that (i) conditional on $\boldsymbol{x}$, the random variables $Y_i, i = 1,\ldots,n$, are mutually independent with distributions $f(y;\mu_i)$ specified by the values of $\mu_i = \mathrm{E}[Y_i|\boldsymbol{x}_i]$; (ii) for some known link function g, $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$. In this well-known framework, a Gaussian distribution for the Y_i's with an identity link function yields the quadratic loss function estimator of $\boldsymbol{\beta}$.

When we are interested in the conditional quantile, $q_p(y_i|\boldsymbol{x}_i)$, rather than the conditional mean, $\mathrm{E}[Y_i|\boldsymbol{x}_i]$, we could still cast the problem in the framework of the generalized linear model, no matter what the original distribution of the data is, by assuming that (i) $f(y;\mu_i)$ is asymmetric Laplace and (ii) $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}(p) = q_p(y_i|\boldsymbol{x}_i)$, for any $0 < p < 1$.

In this paper, we adopt a Bayesian approach to inference. Although a standard conjugate prior distribution is not available for the quantile regression formulation, MCMC methods may be used for extracting the posterior distributions of unknown parameters. This, in fact, allows the use of virtually any prior distribution. Besides giving us the marginal and joint posterior distributions of all the unknown parameters, the Bayesian framework, implemented via the MCMC method, also provides a convenient way of incorporating parameter uncertainty into predictive inferences.

Given the observations, $\boldsymbol{y} = (y_1,\ldots,y_n)$, the posterior distribution of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\boldsymbol{y})$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{y}) \propto L(\boldsymbol{y}|\boldsymbol{\beta})\, p(\boldsymbol{\beta}), \tag{5}$$

where $p(\boldsymbol{\beta})$ is the prior distribution of $\boldsymbol{\beta}$ and $L(\boldsymbol{y}|\boldsymbol{\beta})$ is the likelihood function written as

$$L(\boldsymbol{y}|\boldsymbol{\beta}) = p^n(1-p)^n\exp\left\{-\sum_i \rho_p(y_i - \boldsymbol{x}_i'\boldsymbol{\beta})\right\} \tag{6}$$

using (4) with a location parameter $\mu_i = \boldsymbol{x}_i'\boldsymbol{\beta}$.

In theory, we could use any prior in (5), but in the absence of any realistic information one could use improper uniform prior distributions for all the components of $\boldsymbol{\beta}$. This choice is appealing as the resulting joint posterior distribution is proportional to the likelihood surface. We discuss the validity of using such a prior in the next section.

4. Improper priors for parameters

In this section, we show that if we choose the prior of $\boldsymbol{\beta}$ to be improper uniform, then the resulting joint posterior distribution will be proper.

Theorem 1. *If the likelihood function is given by* (6) *and* $p(\boldsymbol{\beta}) \propto 1$, *then the posterior distribution of* $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\boldsymbol{y})$, *will have a proper distribution. In other words*

$$0 < \int \pi(\boldsymbol{\beta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta} < \infty,$$

or, *equivalently*,

$$0 < \int L(\boldsymbol{y}|\boldsymbol{\beta})\,p(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta} < \infty.$$

Proof. See Appendix A.

In practice one usually assumes that the components of $\boldsymbol{\beta}$ have independent improper uniform prior distributions which is a special case of the above theorem.

5. Applications

We describe our approach to Bayesian quantile regression through one simulated and two real examples. The real examples are based on the immunoglobulin-G and stack loss data sets.

We chose independent improper uniform priors for all the components of $\boldsymbol{\beta}$ in all the examples. We simulated realizations from the posterior distributions by means of a single-component Metropolis–Hastings algorithm (Gilks et al., 1996). Each of the parameters was updated using a random-walk Metropolis algorithm with a Gaussian proposal density centered at the current state of the chain. In all the examples, time series plots showed that the Markov chains converged almost within the first few iterations. However, we discarded the first 1000 runs in every case and then collected a sample of 5000 values from the posterior of each of the elements of $\boldsymbol{\beta}$.

5.1. Simulated data

We generated $n = 100$ observations from the model

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n,$$

assuming that $\mu = 5.0$ and $\varepsilon_i \sim \mathrm{N}(0, 1)$ for all $i = 1, \ldots, n$.

In this example, we have only one parameter μ, so that $q_p(y_i) = \beta(p)$. Fig. 1 shows the posterior histograms of $\beta(p)$ for $p = 0.05$, 0.25, 0.75 and 0.95. Table 1 compares the posterior means with the true values of $\beta(p)$.

Although, we show here the results from a normal sample, the procedure worked quite well with other error distributions as well, indicating that the use of the asymmetric Laplace distribution to model the quantile regression parameters was indeed quite satisfactory.

5.2. Immunoglobulin-G

This data set refers to the serum concentration (grams per litre) of immunoglobulin-G (IgG) in 298 children aged from 6 months to 6 years. A detailed discussion of this data set may be found in Isaacs et al. (1983)

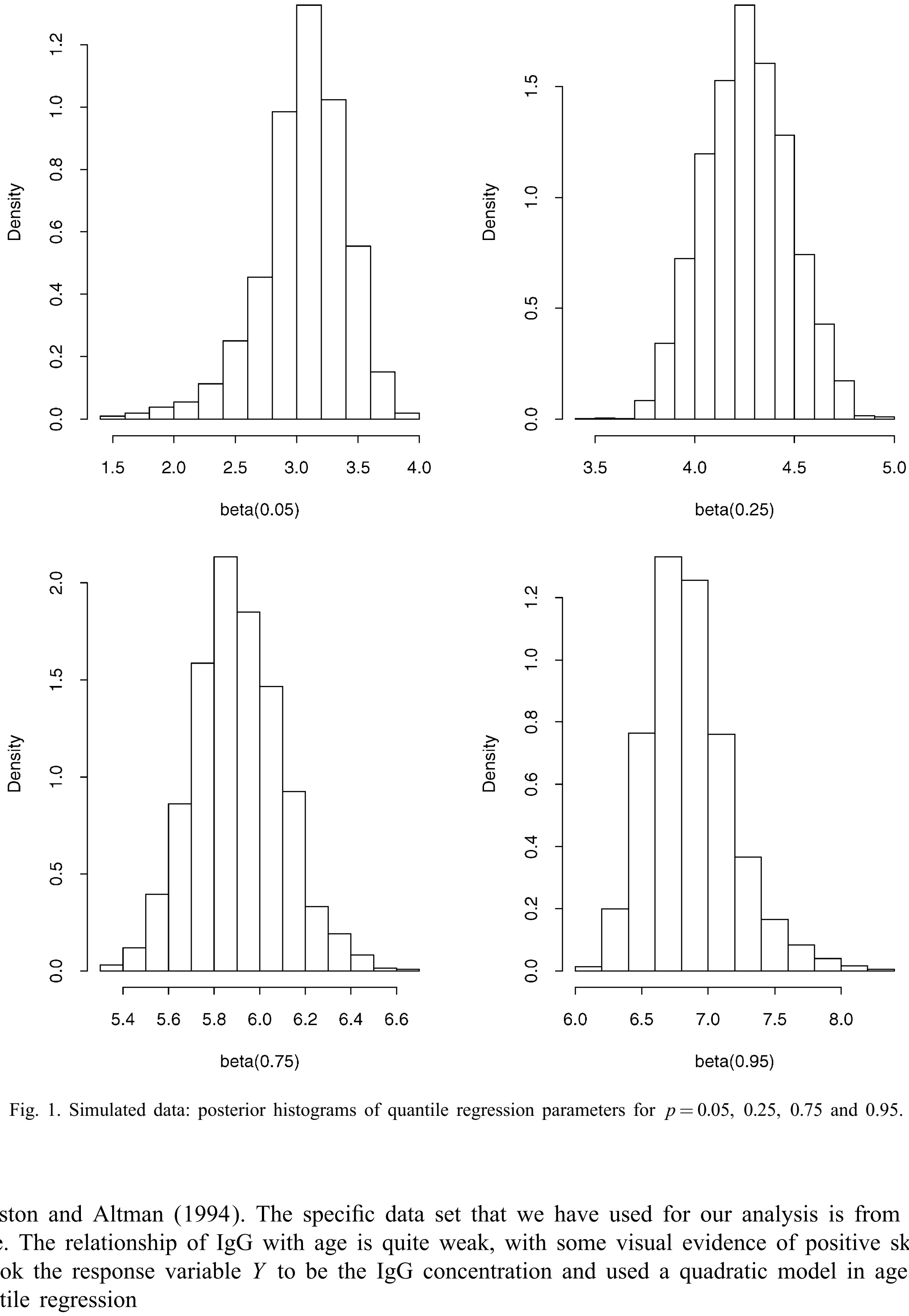

Fig. 1. Simulated data: posterior histograms of quantile regression parameters for $p = 0.05$, 0.25, 0.75 and 0.95.

and Royston and Altman (1994). The specific data set that we have used for our analysis is from the latter reference. The relationship of IgG with age is quite weak, with some visual evidence of positive skewness.

We took the response variable Y to be the IgG concentration and used a quadratic model in age, x, to fit the quantile regression

$$q_p(y|x) = \beta_0(p) + \beta_1(p)x + \beta_2(p)x^2,$$

Table 1
Posterior means, posterior standard deviations (S.D.) and true values of $\beta(p)$ for the simulated data

p	Mean $\beta(p)$	S.D. $\beta(p)$	True $\beta(p)$
0.05	3.055	0.365	3.355
0.25	4.262	0.264	4.326
0.75	5.904	0.289	5.674
0.95	6.864	0.503	6.645

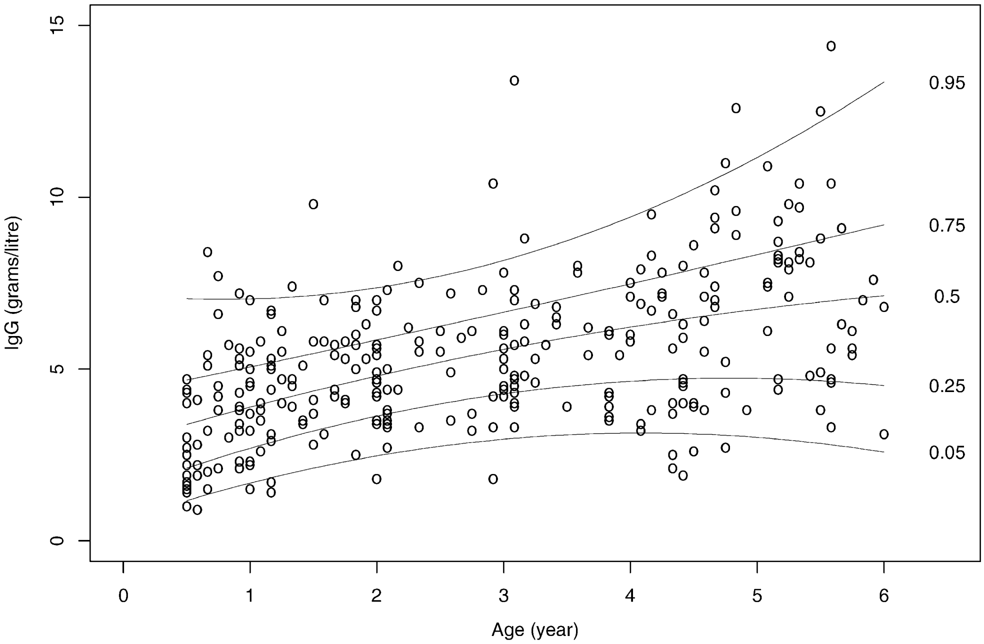

Fig. 2. IgG data: plot showing the quantile regression curves fitted to the data.

for $0 < p < 1$. Fig. 2 shows the plot of the data along with the quantile regression lines for values of $p = 0.05$, 0.25, 0.5, 0.75 and 0.95. Each point on the curves in Fig. 2 is the mean of the predictive posterior distribution given by

$$\pi(y|\boldsymbol{y}) = \int f(y; \boldsymbol{\beta}(p))\pi(\boldsymbol{\beta}(p)|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\beta}(p).$$

We could also obtain desired credible intervals around these curves using the MCMC samples of $\boldsymbol{\beta}(p)$. Fig. 3 displays the quantile regression curves with their 95% pointwise credible intervals corresponding to the extreme quantiles $p = 0.05$ and 0.95.

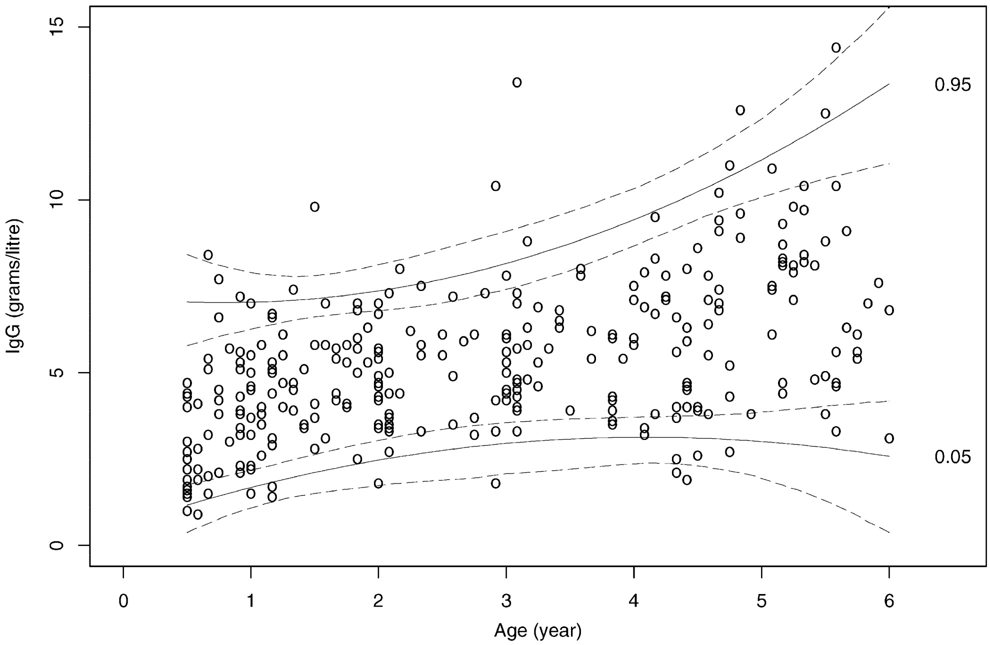

Fig. 3. IgG data: quantile regression curves with the corresponding pointwise 95% credible intervals for $p = 0.05$ and 0.95 (solid line, mean and dashed lines, 2.5% lower and upper credible intervals).

Table 2
Estimates and 95% intervals for the 0.95 quantile regression parameters of the stack loss data

Parameters	2.5% quantile	97.5% quantile	Mean	Median
$\beta_0(0.95)$	−92.546	34.332	−44.259	−50.269
$\beta_1(0.95)$	0.180	1.453	0.751	0.737
$\beta_2(0.95)$	−0.165	2.731	1.478	1.549
$\beta_3(0.95)$	−1.016	0.601	−0.098	−0.045

5.3. Stack loss

As our next example, we consider Brownlee's (1965) much-studied stack loss data, also discussed in Venables and Ripley (2000). The data are from the operation of a plant for the oxidation of ammonia to nitric acid, measured on 21 consecutive days. There are three explanatory variables: air flow to the plant, x_1, cooling water inlet temperature, x_2, and acid concentration, x_3. The response variable, Y, is the percentage of ammonia lost (times 10).

To this set of data we fitted a quantile regression model of the form

$$q_p(y|\boldsymbol{x}) = \beta_0(p) + \beta_1(p)x_1 + \beta_2(p)x_2 + \beta_3(p)x_3.$$

Table 2 shows the posterior mean, median and 95% credible intervals for each of the 0.95 quantile regression parameters. Fig. 4 exhibits the empirical samples from the joint posterior distributions of the quantile regression

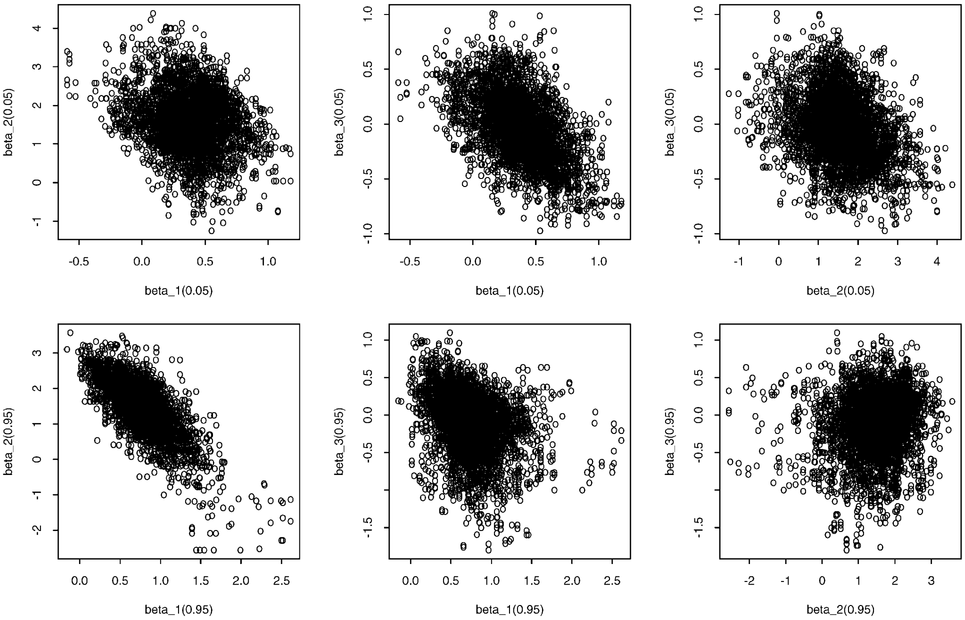

Fig. 4. Stack loss data: plots showing the empirical samples from the joint distributions of quantile regression parameters for $p = 0.05$ and 0.95.

coefficients for $p = 0.05$ and 0.95. The relationship between the regression coefficients can be readily seen from this figure.

6. Discussion

In this paper, we have shown how Bayesian inference may be undertaken in the context of quantile regression. This is achieved by putting the Bayesian quantile regression problem in the framework of the generalized linear model and using the asymmetric Laplace distribution to form the likelihood function. We have shown through one simulated and two real examples the usefulness of this approach. The posterior distributions of the unknown model parameters were obtained by using MCMC methods implemented in S-Plus. The use of the asymmetric Laplace distribution makes the method robust and the Bayesian approach which is fairly easy to implement provides complete univariate and joint posterior distributions of parameters of interest.

Although, we have chosen improper flat priors in our examples, there is scope of using other priors in a relatively straightforward fashion. Our approach of estimating the quantile functions may be extended to spatial and random effects models. This is currently being investigated.

Appendix A. Proof of Theorem 1

Lemma 1. *When the vector* $\boldsymbol{\beta}$ *has only one component, the pth conditional quantile is given by* $q_p(y|x) = \beta(p)$. *For the sake of simplicity, we write* $\beta = \beta(p)$. *If we assume that* $p(\beta) \propto 1$, *then*

$$0 < \int L(\mathbf{y}|\beta)\,\mathrm{d}\beta < +\infty.$$

Proof. Without loss of generality, suppose that the observations $\{y_j\}_{j=1}^n$ are sorted as $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$, then given p and n,

$$\begin{aligned}\int_{-\infty}^{+\infty} L(\boldsymbol{y}|\beta)\,\mathrm{d}\beta &= p^n(1-p)^n \int_{-\infty}^{+\infty} \exp\left[-\sum_{j=1}^n (y_j-\beta)\{p - I(y_j \leqslant \beta)\}\right]\mathrm{d}\beta \\ &\geqslant p^n(1-p)^n \int_{y_n}^{+\infty} \exp\left[-\sum_{j=1}^n (y_j-\beta)\{p - I(y_j \leqslant \beta)\}\right]\mathrm{d}\beta \\ &= p^n(1-p)^n \int_{y_n}^{+\infty} \exp\left\{(1-p)\sum_{j=1}^n (y_j-\beta)\right\}\mathrm{d}\beta \\ &= p^n(1-p)^n \exp\left\{(1-p)\left(y_n - \sum_{j=1}^n y_j\right)\Big/ (n(1-p))\right\} \\ &\geqslant p^n(1-p)^{(n-1)}/n \\ &> 0.\end{aligned}$$

Using the Cauchy–Schwarz inequality $(\int f(x)g(x)\,\mathrm{d}x)^2 \leqslant \int f(x)^2\,\mathrm{d}x \int g(x)^2\,\mathrm{d}x$ repeatedly, we obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} L(\boldsymbol{y}|\beta)\,\mathrm{d}\beta &= p^n(1-p)^n \int_{-\infty}^{+\infty} \prod_{j=1}^n \exp[-(y_j-\beta)\{p - I(y_j \leqslant \beta)\}]\,\mathrm{d}\beta \\ &\leqslant p^n(1-p)^n \prod_{j=1}^{n-1}\left(\int_{-\infty}^{+\infty} \exp[-2^j(y_j-\beta)\{p - I(y_j \leqslant \beta)\}]\,\mathrm{d}\beta\right)^{1/2^j} \\ &\quad \times \left(\int_{-\infty}^{+\infty} \exp[-2^n(y_n-\beta)\{p - I(y_n \leqslant \beta)\}]\,\mathrm{d}\beta\right)^{1/2^n} \\ &= p^n(1-p)^n \times \prod_{j=1}^{n-1} \frac{1}{2^{j/2^j}} \times \frac{1}{2^{n-1}2^{n-1}} \\ &\quad \times \frac{1}{(p(1-p))^{\sum_{j=1}^{n-1} 1/2^j}} \times \frac{1}{(p(1-p))^{2^{n-1}}}.\end{aligned}$$

Note that $p(1-p) \leqslant 1/4$, $\prod_{j=1}^{n-1} 1/2^{j/2^j} < \infty$, and $\sum_{j=1} 1/2^j < 2$. Thus Lemma 1 is proved.

Proof of Theorem 1. Let $\boldsymbol{\beta}_{-q}$ and $\mathbf{x}'_{\mathbf{j}_{-q}}$ denote $\boldsymbol{\beta}$ and $\mathbf{x}'_{\mathbf{j}}$ without the qth component. Without loss of generality, suppose that $x_{qj} > 0$ $(j = 1, \ldots, n)$ and let $v_j(\boldsymbol{\beta}) = (y_j - \mathbf{x}'_{\mathbf{j}_{-q}}\boldsymbol{\beta}_{-q})/x_{qj}$, $(j = 1, 2, \ldots, n)$, and assume that $v_1(\boldsymbol{\beta}) \leqslant v_2(\boldsymbol{\beta}) \leqslant \cdots \leqslant v_n(\boldsymbol{\beta})$. Then $y_j \leqslant \mathbf{x}'_{\mathbf{j}}\boldsymbol{\beta}$ if and only if $v_j(\boldsymbol{\beta}) \leqslant \beta_q$ and $y_j - \mathbf{x}'_{\mathbf{j}}\boldsymbol{\beta} = (v_j(\boldsymbol{\beta}) - \beta_q)x_{qj}$. Thus, when $p(\boldsymbol{\beta}) \propto 1$, we get

$$\begin{aligned}
\int L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta} &= p^n(1-p)^n \int \exp\left[-\sum_{j=1}^{n}(y_j - \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\{p - I(y_j \leqslant \mathbf{x}_\mathbf{j}^\mathrm{T}\boldsymbol{\beta})\}\right]\,\mathrm{d}\boldsymbol{\beta} \\
&= p^n(1-p)^n \exp\left[-\sum_{j=1}^{n} x_{qj}(v_j(\boldsymbol{\beta}) - \beta_q)\{p - I(\beta_q \geqslant v_j(\boldsymbol{\beta}))\}\right]\,\mathrm{d}\boldsymbol{\beta} \\
&\geqslant p^n(1-p)^n \int \mathrm{d}\boldsymbol{\beta}_{-q} \int_{v_n(\boldsymbol{\beta})}^{+\infty} \exp\left((1-p)\sum_{j=1}^{n} x_{qj}(v_j(\boldsymbol{\beta}) - \beta_q)\right)\,\mathrm{d}\beta_q \\
&= p^n(1-p)^n \frac{1}{(1-p)\sum_{j=1}^{n} x_{qj}} \\
&\quad \times \int \exp\left[-(1-p)\sum_{j=1}^{n} x_{qj}\{v_n(\boldsymbol{\beta}) - v_j(\boldsymbol{\beta})\}\right]\,\mathrm{d}\boldsymbol{\beta}_{-q}.
\end{aligned}$$

Expanding $\sum_j x_{qj}\{v_n(\boldsymbol{\beta}) - v_j(\boldsymbol{\beta})\}$ as a linear function of $\boldsymbol{\beta}_{-q}$, and using Lemma 1, we see that $\int L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta} > 0$.

Now using the Cauchy–Schwarz inequality repeatedly, we obtain

$$\begin{aligned}
\int L(\boldsymbol{y}|\boldsymbol{\beta})p(\boldsymbol{\beta})\,\mathrm{d}\boldsymbol{\beta} &= p^n(1-p)^n \int_{-\infty}^{+\infty} \prod_{j=1}^{n} \exp[-(y_j - \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\{p - I(y_j \leqslant \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\}]\,\mathrm{d}\boldsymbol{\beta} \\
&\leqslant p^n(1-p)^n \prod_{i=1}^{n-1} \left(\int_{-\infty}^{+\infty} \exp[-2^j(y_i - \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\{p - I(y_j \leqslant \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\}]\,\mathrm{d}\boldsymbol{\beta}\right)^{1/2^j} \\
&\quad \times \left(\int_{-\infty}^{+\infty} \exp[-2^n(y_n - \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\{p - I(y_n \leqslant \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\}]\,\mathrm{d}\boldsymbol{\beta}\right)^{1/2^n}.
\end{aligned}$$

For any $1 \leqslant j \leqslant n$, suppose that $x_{qj} \neq 0$. Let

$$I = \int_{-\infty}^{+\infty} \exp[-2^j(y_j - \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\{p - I(y_j \leqslant \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})\}]\,\mathrm{d}\boldsymbol{\beta}$$

be the jth term of the integral product above. Writing $\boldsymbol{\gamma}' = (\boldsymbol{\gamma}'_{-q}, \gamma_q) = (\boldsymbol{\beta}'_{-q}, \mathbf{x}_\mathbf{j}'\boldsymbol{\beta})$ gives the Jacobian of transformation $|\mathrm{d}\boldsymbol{\beta}/\mathrm{d}\boldsymbol{\gamma}| = 1/|x_{qj}|$. Thus,

$$I = \frac{1}{|x_{qj}|} \int \exp[-2^j(y_j - \gamma_q)\{p - I(y_j < \gamma_q)\}]\,\mathrm{d}\boldsymbol{\gamma}.$$

Using Lemma 1, it follows that

$$\int_{-\infty}^{+\infty} L(\boldsymbol{y}|\beta)\,\mathrm{d}\beta < +\infty.$$

References

Buchinsky, M., 1998. Recent advances in quantile regression models. J. Hum. Res. 33, 88–126.

Cole, T.J., Green, P.J., 1992. Smoothing reference centile curves: LMS method and penalized likelihood. Statist. Medic. 11, 1305–1319.

Fatti, L.P., Senaoana, E.M., 1998. Bayesian updating in reference centile charts. J. R. Statist. Soc. A 161, 103–115.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Introducing Markov chain Monte Carlo. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 1–19.

He, X., Ng, P., Portnoy, S., 1998. Bivaraite quantile smoothing splines. J. R. Statist. Soc. B 60, 537–550.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Isaacs, D., Altman, D.G., Tidmarsh, C.E., Valman, H.B., Webster, A.D.B., 1983. Serum Immunoglobin concentrations in preschool children measured by laser nephelometry: reference ranges for IgG, IgA, IgM. J. Clin. Pathol. 36, 1193–1196.

Koenker, R., Bassett, G.S., 1978. Regression quantiles. Econometrica 46, 33–50.

Koenker, R., Machado, J., 1999. Goodness of fit and related inference processes for quantile regression. J. Amer. Statist. Assoc. 94, 1296–1309.

Royston, P., Altman, D.G., 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). Appl. Statist. 43, 429–467.

Venables, W.N., Ripley, B.D., 2000. Modern Applied Statistics with S-PLUS, 3rd Edition. Springer, New York.

Yu, K., Jones, M.C., 1998. Local linear regression quantile estimation. J. Amer. Statist. Assoc. 93, 228–238.