ORIGINAL ARTICLE

A Discrete Density Approach to Bayesian Quantile and Expectile Regression with Discrete Responses

Xi Liu[1,2] · Xueping Hu[1,2] · Keming Yu[1,2]

Abstract
For decades, regression models beyond the mean for continuous responses have attracted great attention in the literature. These models typically include quantile regression and expectile regression. But there is little research on these regression models for discrete responses, particularly from a Bayesian perspective. By forming the likelihood function based on suitable discrete probability mass functions, this paper introduces a discrete density approach for Bayesian inference of these regression models with discrete responses. Bayesian quantile regression for discrete responses is first developed, and then this method is extended to Bayesian expectile regression for discrete responses. The posterior distribution under this approach is shown not only coherent irrespective of the true distribution of the response, but also proper with regarding to improper priors for the unknown model parameters. The performance of the method is evaluated via extensive Monte Carlo simulation studies and one real data analysis.

Keywords Bayesian inference · Discrete asymmetric Laplace distribution · Discrete asymmetric normal distribution · Discrete responses · Expectile regression · Posterior consistency · Quantile regression

1 Introduction

Regression models for dealing with responses following a non-normal distribution have been drawing significant attention in the literature. For example, quantile regression and expectile regression have been widely developed in the literature and increasingly applied to a greater variety of scientific questions. See, [1–5] and among others.

✉ Keming Yu
   keming.yu@brunel.ac.uk

1   Anqing Normal University, Anqing, China

2   Brunel University London, London, England

⧏ Springer

Typically, quantile regression estimates various conditional quantiles of a response or dependent random variable, including the median (0.5th quantile). Putting different quantile regressions together provide a more complete description of the underlying conditional distribution of the response than a simple mean regression. This is particularly useful when the conditional distribution is asymmetric or heterogeneous or fat-tailed or truncated. Quantile regression has been widely used in statistics and numerous application areas. Bayesian quantile regression for continuous responses has received increasing attention from both theoretical and empirical viewpoints. See a recent review by [6] for the first type of Bayesian quantile regression methods ([7–9] and among others) based on asymmetric Laplace distribution (ALD) likelihood function. But among numerous application areas of regression models, discrete observations such as integer values (e.g., -2, -1, 0, 1, 2, 3, etc.) on a response are easily collected. In particular, many big data nowadays contain discrete observations such as number of online transaction, number of days in hospital, number of votes and so on. Classic regression models for discrete responses include logistic, Poisson and negative Binomial regression. Discrete responses are generally skewed, the mean-based regression analysis would not be sufficient for a complete analysis ([10]).

However, quantile regression for discrete responses receives far less attention than for continuous responses in the literature. Binary quantile regression was first introduced by [11]. Then, several authors ([12–14] and among others) developed different smoothed estimation techniques (nonparametric or semiparametric methods) for the binary quantile regression model under frequentist approaches. Based on the idea of linking ALD to latent variables in Bayesian Tobit quantile regression ([15]), the papers by [16, 17] and among others did proposed Bayesian inference binary quantile regression. Based on a normal-exponential mixture representation of ALD, [18] and [19] extended Bayesian binary quantile regression to Bayesain ordinal quantile regression. Then, [20] further extended it to Bayesain inference of single-index quantile regression for ordinal data. [21, 22] and [23] applied these discrete quantile regression methods in economics, energy and education, respectively.

But all these research methods are not a direct Bayesian quantile regression approach for discrete responses, or deal with quantile regression for general discrete responses. Similarly, there is little research on expectile regression for discrete responses, let alone from a Bayesian perspective ([24]). A semi-parametric jittering approach for general count quantile regression has been introduced ([25]), but some degree of smoothness has to be artificially imposed on the approach. Quantile regression for count data may be achieved via density regression as shown in [26], but this approach may result in a global estimation of regression coefficients. In this paper, we propose Bayesian inference quantile regression for discrete responses via introducing a discrete version of ALD-based likelihood function. This approach not only keeps the 'local property' of quantile regression, but also enjoys the coherency and finite posterior moments of the posterior distribution. Along this line, we then introduce Bayesian expectile regression for discrete responses, which proceeds by forming the likelihood function based on a discrete asymmetric normal distribution (DAND). Section 2 introduces a discrete asymmetric Laplace distribution (DALD) and discusses its natural link with quantile regression for discrete responses. Sections 3 and 4 detail this Bayesian approach for

quantile regression and expectile regression with discrete responses, respectively. Section 5 illustrates the numerical performance of the proposed method. Section 6 concludes with a brief discussion.

2 Discrete Asymmetric Laplace Distribution

Let Y be a real-valued random variable with its τ-th $(0 < \tau < 1)$ quantile μ $(-\infty < \mu < \infty)$, and then it is well-known that μ could be found by minimizing the expected loss of Y with respect to the loss function (or check function) $\rho_\tau(y) = y(\tau - I(y < 0))$, or $\min_\mu E_{F_0(Y)}\rho_\tau(Y - \mu)$, where $F_0(Y)$ denotes the distribution function of Y, which is usually unknown in practice.

When Y is a continuous random variable, the inference based on the loss function $\rho_\tau(y - \mu)$ was linked to a maximum likelihood inference based on an $\text{ALD}(\mu, \tau)$ with local parameter μ and shape parameter τ:

$$f(y; \mu, \tau) = \tau(1 - \tau) \exp\left\{-\rho_\tau(y - \mu)\right\}. \tag{1}$$

Now, if Y is a discrete random variable, let Y take integer values in $\mathbb{Z}$. We first derive a discrete version of ALD or a DALD and then show that the τth quantile μ can also be estimated via this DALD.

To this end, note that the corresponding cumulative distribution function (c.d.f.) of an ALD in Eq.(1) can be written as:

$$F(y; \mu, \tau) = \begin{cases} 1 - (1 - \tau)\exp\left\{-\tau(y - \mu)\right\}, & y \geq \mu, \\ \tau \exp\left\{(1 - \tau)(y - \mu)\right\}, & y < \mu. \end{cases} \tag{2}$$

Let $S(y; \mu, \tau)$ be the survival function of this ALD, which is given by:

$$\begin{aligned} S(y; \mu, \tau) &= 1 - F(y; \mu, \tau) \\ &= \begin{cases} (1 - \tau)\exp\left\{-\tau(y - \mu)\right\}, & y \geq \mu, \\ 1 - \tau \exp\left\{(1 - \tau)(y - \mu)\right\}, & y < \mu, \end{cases} \end{aligned} \tag{3}$$

then, according to [27], the probability mass function (p.m.f.) of a DALD can be defined as:

$$\phi(y; \mu, \tau) = \begin{cases} S(y; \mu, \tau) - S(y + 1; \mu, \tau), & y \in \mathbb{Z}, \\ 0, & otherwise, \end{cases} \tag{4}$$

with $S(y; \mu, \tau)$ in Eq.(3). Note that when $y \geq \mu$,

$$S(y;\mu,\tau) - S(y+1;\mu,\tau)$$

$$= (1-\tau)\left(\exp\{-\tau(y-\mu)\} - \exp\{-\tau(y+1-\mu)\} \right)$$

$$= (1-\tau)\left(\exp\{-\tau(y-\mu)\} - \exp\{-\tau(y-\mu)\}\exp\{-\tau\} \right)$$

$$= (1-\tau)\exp\{-\tau(y-\mu)\}\left(1 - \exp\{-\tau\} \right)$$

$$= -(1-\tau)\left(\exp\{-\tau\} - 1 \right)\exp\{-\tau(y-\mu)\}$$

$$= \rho_\tau(-\mathrm{sgn}(y-\mu))\left[\exp\{-\rho_\tau(\mathrm{sgn}(y-\mu))\} - 1\right]\exp\{-\rho_\tau(y-\mu)\},$$

and similar deviation of it in the case of $y < \mu$. It follows:

$$\phi(y;\mu,\tau) = \rho_\tau(-\mathrm{sgn}(y-\mu))\left[\exp\{-\rho_\tau(\mathrm{sgn}(y-\mu))\} - 1\right]$$
$$\exp\{-\rho_\tau(y-\mu)\} \qquad (5)$$
$$y = \cdots, -1, 0, 1, \cdots,$$

and the loss function (or check function) is

$$\rho_\tau(u) = \frac{|u| + (2\tau-1)u}{2}.$$

Remark 1 One could also incorporate scale parameter σ in Eq.(4) to obtain

$$\phi(y;\mu,\tau) = \rho_\tau(-\mathrm{sgn}(y-\mu))\left[\exp\left\{-\rho_\tau\left(\mathrm{sgn}\left(\frac{y-\mu}{\sigma}\right)\right)\right\} - 1\right]$$
$$\exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\},$$
$$y = \cdots, -1, 0, 1, \cdots.$$

According to [6], any fixed σ can be utilised to obtain asymptotically valid posterior inference and make the results asymptotically invariant. Here, we simply fix σ as 1.

Given a sample $Y = (Y_1, Y_2, \cdots, Y_n)$ of the discrete response Y whose distribution $F_0(y)$ may be unknown, consider the DALD-based likelihood function for μ:

$$L(Y|\mu) = \prod_{i=1}^{n} \qquad (6)$$
$$\left[\rho_\tau(-\mathrm{sgn}(Y_i-\mu))\left[\exp^{-\rho_\tau(\mathrm{sgn}(Y_i-\mu))} - 1\right]\exp\{-\rho_\tau(Y_i-\mu)\}\right].$$

Then, we have

$$\operatorname*{argmax}_{\mu} L(\boldsymbol{Y}|\mu)$$

$$= \operatorname*{argmax}_{\mu} \log L(\boldsymbol{Y}|\mu)$$

$$= \operatorname*{argmax}_{\mu} \left\{ - \sum_{i=1}^{n} \rho_\tau (Y_i - \mu) \right\}$$

$$= \operatorname*{argmin}_{\mu} \sum_{i=1}^{n} \rho_\tau (Y_i - \mu).$$

This means that the estimation of the τ-th quantile μ of a discrete random variable Y with respect to the loss function $\rho_\tau(\cdot)$ is equivalent to maximization of the likelihood function Eq.(6) based on the DALD. According to [28], a Bayesian inference of μ can be developed. That is, if $\pi(\mu)$ represents prior beliefs about the τ-th quantile μ, and $\boldsymbol{Y}$ is observed data from the unknown distribution $F_0(Y)$ of the discrete random variable Y, then a posterior $\pi(\mu|\boldsymbol{Y})$ which is valid and coherent update of $\pi(\mu)$ can be obtained via the DALD-based likelihood function Eq.(6) and is given by:

$$\pi(\mu|\boldsymbol{Y}) \propto \pi(\mu)\, L(\boldsymbol{Y}|\mu). \tag{7}$$

Coherency here means if ν denotes a probability measure on the space of μ, then ν is named coherent if

$$\iint \rho_\tau (Y - \mu) dF_0(Y)\nu(d\mu) \leq \iint \rho_\tau (Y - \mu) dF_0(Y)\nu_1(d\mu),$$

for all other probability measure ν_1 on the space of μ in terms of expected loss of Y given by $E_{F_0(Y)}\rho_\tau(Y - \mu)$. This coherency property aims to ensure the consistency of posterior from the proposed inference even if the 'working likelihood' in Eqs. (3)–(A1) is misspecified.

3 Bayesian Quantile Regression with Discrete Responses

Generalized linear models (GLMs) extend the linear modelling capability to scenarios that involve non-normal distributions $f(y;\mu)$ or heteroscedasticity, with $f(y;\mu)$ specified by the values of $\mu = E[Y|X = x]$ conditional on x, including to involve a known link function g, $g(\mu) = x^T \boldsymbol{\beta}$. Specifically, GLMs also apply to the so-called 'exponential' family of models, which typically include Poisson regression with log-link function.

When we are interested in the conditional quantile $Q_Y(\tau|x)$ of a discrete response, according to [7], 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) $f(y;\mu)$ follows a DALD in the form of Eqs.(5) or (A1) and (ii) $g(\mu) = x^T \boldsymbol{\beta}(\tau) = Q_Y(\tau|x)$ for any $0 < \tau < 1$.

When covariate information such as a covariate vector X is available, quantile regression denoted by $Q_Y(\tau|X)$ for μ is introduced. Without loss of generality,

consider a linear regression model for $Q_Y(\tau|X)$: $Q_Y(\tau|X) = X^T\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the regression parameter vector.

Given observations $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$ of the discrete response Y, one of the aims in regression analysis is the inference of $\boldsymbol{\beta}$. Let $\pi(\boldsymbol{\beta})$ be the prior distribution of $\boldsymbol{\beta}$, then the posterior distribution of $\boldsymbol{\beta}$, $\pi(\boldsymbol{\beta}|\boldsymbol{Y})$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{Y}) \propto \pi(\boldsymbol{\beta}) L(\boldsymbol{Y}|\boldsymbol{\beta}), \tag{8}$$

where the likelihood function $L(\boldsymbol{Y}|\boldsymbol{\beta})$ is given by:

$$L(\boldsymbol{Y}|\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\rho_\tau(-\mathrm{sgn}(Y_i - X_i^T\boldsymbol{\beta})) \left[\exp^{-\rho_\tau(\mathrm{sgn}(Y_i - X_i^T\boldsymbol{\beta}))} - 1 \right] \right.$$
$$\left. \exp\left\{ -\rho_\tau(Y_i - X_i^T\boldsymbol{\beta}) \right\} \right].$$

The numerical computation of the posterior distribution can be carried out by the Metropolis-Hastings algorithm. That is, we first generate a candidate $\boldsymbol{\beta}^*$ according to a random walk, which results in a symmetric proposal distribution in the Metropolis algorithm. Then, accept or reject $\boldsymbol{\beta}^*$ for $\boldsymbol{\beta}$ according to the acceptance probability $p(\boldsymbol{\beta}^*|\boldsymbol{\beta}) = \min\left(1, \frac{L(y|\boldsymbol{\beta}^*)}{L(y|\boldsymbol{\beta})}\right)$ around 30%. We always suggest to throw away at last 30% of the iterations at the beginning of an MCMC run or 'burn-in' period to make sure the chain to reach stationarity.

Besides the coherent property discussed in Sect. 2 for posterior distribution $\pi(\boldsymbol{\beta}|\boldsymbol{Y})$, it is important to verify the existence of the posterior distribution when the prior of $\boldsymbol{\beta}$ is improper, i.e,

$$0 < E\{\pi(\boldsymbol{\beta}|\boldsymbol{Y})\} < \infty,$$

or, equivalently,

$$0 < E\{\pi(\boldsymbol{\beta}) L(\boldsymbol{Y}|\boldsymbol{\beta})\} < \infty.$$

Moreover, it is preferable to check that the existence of posterior moments of the regression parameters is entirely unaffected by improper priors and quantile index τ ([29] and among others), i.e,

$$E\left[\left(\prod_{j=0}^{m} |\beta_j|^{r_j}\right)\Big|\boldsymbol{Y}\right] < \infty, \tag{9}$$

where r_j denotes the order of the moments of β_j.

To this end, we have the following conclusion:

Theorem 3.1 *Assume the posterior is given by Eq.(8) and $\pi(\boldsymbol{\beta}) \propto 1$, then all posterior moments of $\boldsymbol{\beta}$ in Eq.(9) exist.*

The proof of Theorem 3.1 is available in the Supplementary Materials.

4 Bayesian Expectile Regression for Discrete Responses

Instead of defining the τ-th quantile of a response Y by $\underset{\mu}{\arg\min} E\big(\rho_\tau(Y - \mu)\big)$, [30] defined the θ-th expectile of Y by

$$Expectile_\theta(Y) = \underset{\mu}{\arg\min} E\Big(\rho_\theta^{(E)}(Y - \mu)\Big), \tag{10}$$

in terms of an asymmetric quadratic loss function

$$\rho_\theta^{(E)}(u) = u^2|\theta - I(u < 0)|,$$

where $\theta \in (0, 1)$ determines the degree of asymmetry of the loss function. Note that θ is typically not equal to τ, although there is a one-to-one relationship between τ-th quantile and θ-th expectile ([31]).

Corresponding to $\rho_\theta^{(E)}(u)$, we can define an asymmetric normal distribution (AND) whose density function is given by

$$f^{(E)}(y;\mu,\theta) = k \begin{cases} \exp\big\{-\theta(y - \mu)^2\big\}, & y \geq \mu, \\ \exp\big\{(\theta - 1)(y - \mu)^2\big\}, & y < \mu, \end{cases} \tag{11}$$

where $k = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\theta(1-\theta)}}{\sqrt{\theta}+\sqrt{1-\theta}}$, μ and θ are the location parameter and shape parameter, respectively.

The corresponding c.d.f. of the AND can be written as:

$$F^{(E)}(y;\mu,\theta)$$
$$= \begin{cases} k\sqrt{\frac{\pi}{\theta}}\Phi\Big(\sqrt{2\theta}(y - \mu)\Big) + \frac{k}{2}\Big(\sqrt{\frac{\pi}{1-\theta}} - \sqrt{\frac{\pi}{\theta}}\Big), & y > \mu, \\ k\sqrt{\frac{\pi}{1-\theta}}\Phi\Big(\sqrt{2(1-\theta)}(y - \mu)\Big), & y \leq \mu, \end{cases} \tag{12}$$

where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution.

Therefore, based on the survival function $S^{(E)}(y;\mu,\theta) = 1 - F^{(E)}(y;\mu,\theta)$, we can derive the p.m.f. of the DAND by following the same procedure as in Eq.(4). In fact, note that

$$\rho_\theta(\text{sgn}(y - \mu)) = \begin{cases} \theta(y - \mu), & y > \mu, \\ -(1 - \theta)(y - \mu) & y \leq \mu, \end{cases}$$

then, for $y \in Z$,

$$\phi^{(E)}(y;\mu,\tau) = S^{(E)}(y;\mu,\theta) - S^{(E)}(y + 1;\mu,\theta)$$
$$= F^{(E)}(y + 1;\mu,\theta) - F^{(E)}(y;\mu,\theta),$$

we have

$$\phi^{(E)}(y;\mu,\tau) = k\sqrt{\frac{\pi}{\rho_\theta(\operatorname{sgn}(y-\mu))}}\left[\Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(y-\mu))}(y+1-\mu)\right)\right.$$
$$\left. - \Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(y-\mu))}(y-\mu)\right)\right], \quad y = \cdots, -1, 0, 1, \cdots. \tag{13}$$

Now, if Y is a discrete random variable with unknown distribution function $F_0(y)$, then given a sample $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)$ of Y, the θ-th expectile of Y is estimated by the minimization of the loss function $\rho_\theta^{(E)}$ or $\operatorname{argmin}_\mu \sum_{i=1}^n \rho_\theta^{(E)}(Y_i - \mu)$. Consider the DAND-based likelihood function:

$$L^{(E)}(\boldsymbol{Y}|\mu) = \prod_{i=1}^n \left[k\sqrt{\frac{\pi}{\rho_\theta(\operatorname{sgn}(Y_i-\mu))}}\left[\Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i+1-\mu)\right)\right.\right.$$
$$\left.\left. - \Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i-\mu)\right)\right]\right]. \tag{14}$$

We can see that the expectile μ can also be estimated equivalently by the maximization of the likelihood function $L^{(E)}(\boldsymbol{Y}|\mu)$ in Eq.(14). In fact,

$$\operatorname*{argmax}_\mu L^{(E)}(\boldsymbol{Y}|\mu)$$
$$= \operatorname*{argmax}_\mu \prod_{i=1}^n \left[\Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i+1-\mu)\right)\right.$$
$$\left. -\Phi\left(\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i-\mu)\right)\right]$$
$$= \operatorname*{argmax}_\mu \prod_{i=1}^n \int_{\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i-\mu)}^{\sqrt{2\rho_\theta(\operatorname{sgn}(Y_i-\mu))}(Y_i+1-\mu)} \phi(u)du$$

(According to Lagrange mean value theorem $\int_a^b \phi(u)du = \phi(\xi)(b-a)$,)

$$= \operatorname*{argmax}_\mu \left[\exp\left\{-\rho_\theta(\operatorname{sgn}(Y_i-\mu))\sum_i^n (Y_i-\mu)^2\right\}\right]$$
$$= \operatorname*{argmax}_\mu \left[-\rho_\theta(\operatorname{sgn}(Y_i-\mu))\sum_i^n (Y_i-\mu)^2\right]$$
$$= \operatorname*{argmin}_\mu \sum_{i=1}^n \rho_\theta^{(E)}(Y_i-\mu),$$

where $\phi(\cdot)$ denotes the p.d.f. of the standard normal distribution.

Again, according to [28], a Bayesian inference of the expectile μ can be developed. That is, a coherent posterior $\pi(\mu|\boldsymbol{Y})$ for the update of $\pi(\mu)$ exists and is given by $\pi(\mu|\boldsymbol{Y}) \propto \pi(\mu)L^{(E)}(\boldsymbol{Y}|\mu)$ with the likelihood function $L^{(E)}(\boldsymbol{Y}|\mu)$ in Eq.(14). Along with the similar discussion in Sect. 3, we can prove that the posterior distribution under this Bayesian inference is proper with regarding to improper priors

for regression parameter β in the expectile regression model $\mu = X^T \beta$, if covariate information X is available. The corresponding proofs are available in the Supplementary Materials.

5 Numerical Analysis

In this section, we implement the proposed method to illustrate the Bayesian quantile regression for discrete responses via Monte Carlo simulation studies and one real data analysis. In all numerical analyses, we follow standard practice by setting the variance of random-walk MH sampling to tune the proposal distribution to get around a 0.25–0.3 acceptance rate. We discard the first 10000 of 20000 runs in every case of MCMC outputs and then collect a sample of 10000 values to calculate the posterior distribution of each of regression coefficients in β. All numerical experiments are carried out on one Intel Core i5-3470 CPU (3.20GMHz) processor and 8 GB RAM.

5.1 Simulated Example 1

Consider a simple regression model for which the sample $Y_i(i = 1, 2, \cdots, n)$ is counts and follows a Poisson distribution with parameter 3 and a Binomial distribution with parameters 20 and 1/5, respectively. 500 simulations for each case of $\tau \in \{0.05, 0.25, 0.50, 0.75, 0.95\}$ and $n \in \{200, 1000\}$ are performed. The quantile regression $Q_\tau(Y) = \beta(\tau)$ is a constant depending on τ only. Table 1 compares the posterior means with the true values of $\beta(\tau)$ for each case under 500 simulations. Moreover, the expectile regression $Expectile_\theta(Y) = \beta(\theta)$ is also a constant depending on the θ-th expectile. Table 2 compares the posterior means with the true values of $\beta(\theta)$

Table 1 Posterior mean and posterior standard deviations (S.D.) of $\beta(\tau)$ for quantiles from simulated example 5.1

τ	$n = 200$		$n = 1000$		
	Mean	S.D.	Mean	S.D.	True value
Case 1 : $Y \sim$ Pois(3)					
0.05	1.191	0.119	1.037	0.024	1
0.25	2.103	0.072	2.009	0.006	2
0.50	3.097	0.069	3.007	0.006	3
0.75	4.316	0.157	4.149	0.043	4
0.95	6.438	0.321	6.228	0.116	6
Case 2 : $Y \sim$ Binom(20, 1/5)					
0.05	1.255	0.110	1.028	0.007	1
0.25	3.139	0.078	3.011	0.009	3
0.50	4.175	0.109	4.030	0.011	4
0.75	5.453	0.182	5.441	0.066	5
0.95	7.430	0.310	7.166	0.115	7

Table 2 Posterior mean and posterior standard deviations (S.D.) of $\beta(\theta)$ for expectiles from simulated example 5.1

θ	$n = 200$		$n = 1000$		
	Mean	S.D.	Mean	S.D.	True value
Case 1 : $Y \sim \text{Pois}(3)$					
0.05	1.266	0.103	1.242	0.049	1.24
0.25	2.268	0.072	2.270	0.034	2.27
0.50	2.943	0.070	2.972	0.033	3
0.75	3.662	0.077	3.717	0.033	3.80
0.95	5.029	0.136	5.016	0.061	5.15
Case 2 : $Y \sim \text{Binom}(20, 1/5)$					
0.05	2.321	0.109	2.072	0.049	2.11
0.25	3.360	0.071	3.256	0.032	3.23
0.50	4.086	0.070	4.064	0.032	4
0.75	4.825	0.077	4.869	0.034	4.80
0.95	6.051	0.128	6.294	0.056	6.15

obtained via an empirical estimation in Eq.(10) for different cases. Both Tables 1 and 2 show that the results obtained by the proposed Bayesian inference are reasonably accurate.

5.2 Simulated Example 2

We consider a discrete quantile linear regression:

$$Y_i = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} + \varepsilon_i, \ i = 1, \cdots, n; k = 1, \cdots, p, \tag{15}$$

where n and p denote the number of observations and independent variables, respectively. $\beta_k, k = 1, \dots, p$ are the regression parameters. Let the random item ε_i follows a Poisson distribution with parameter 3. 500 simulations for each case of $\tau \in \{0.25, 0.50, 0.75\}$ and $n \in \{300, 1500\}$ are performed.

Without loss of generality, let $p = 2$ in Eq.(15):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \ \ i = 1, \cdots, n,$$

where covariate $X_{1,i}$ is generated from a Geometric distribution with probability 1/4, and covariate $X_{2,i}$ is generated from a Poisson distribution with parameter 2. We generate the training data with $\beta_i = \{6, 2, -4\}, i = \{0, 1, 2\}$ and $\varepsilon_i \sim \text{Pois}(3)$. 500 simulations for each case of $\tau \in \{0.25, 0.50, 0.75\}$ and $n_1 \in \{200, 1000\}$ are performed.

Therefore, the corresponding discrete quantile function is of the form

$$Q_\tau(Y|X) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2.$$

Although we have chosen improper flat priors in *Simulated example 1* above, one may use other priors for analysis in a relatively straightforward fashion. For example,

along with [32], conditional conjugate prior distribution in the Normal-Gamma Inverse form for the unknown parameters $\boldsymbol{\beta}$ can be obtained. Given $\tau \in (0, 1)$, for any $a > 0$, the prior mean and covariance matrix for $\boldsymbol{\beta}$ are given, respectively, by

$$E(\boldsymbol{\beta}) = \boldsymbol{\beta}_{a\tau}$$
$$Cov(\boldsymbol{\beta}) = 2g(XVX^T)^{-1},$$

where $\boldsymbol{\beta}_a$ is anticipated values, and $g > 0$ is a known scaling factor. Various values of g have been used in the context of variable selection and estimation. [33] performed variable selection using splines and suggested that the value of g is in the range $10 \leq g \leq 1000$. Following the discussions in [32, 34] and among others, we set $g = 100$ in this paper. Thus, given τ and $\boldsymbol{\beta}_{a\tau}$, the conditional prior distribution for $\boldsymbol{\beta}$ is readily available. Here, we suggest a particular form of a conjugate Normal-Inverse Gamma family for $\boldsymbol{\beta}$ given by

$$\boldsymbol{\beta}|V, X \sim N(\boldsymbol{\beta}_a, 2g(XVX^T)^{-1}).$$

For simplicity, let $E(\boldsymbol{\beta})$ and $Cov(\boldsymbol{\beta})$ be the fitted values obtained by the semi-parametric jittering approach ([25]), as presented in Table 3.

Under the proposed Bayesian inference in Sect. 2, Table 4 reports the posterior mean, standard deviation and 95% credible interval for the regression parameters $\beta_0(\tau), \beta_1(\tau)$ and $\beta_2(\tau)$, under 500 simulations with $\varepsilon_i \sim \text{Pois}(3)$, based on a conjugate Normal-Inverse Gamma prior for $\boldsymbol{\beta}$. It can be shown from Table 4 that under different prior settings, the posterior estimates of regression coefficients obtained from the working likelihood analysis are consistent .

5.3 Analysis of Length of Stay (LoS) in Days

The data are extracted from the Worcester Heart Attack Study with 500 observations ([35]), which describes factors associated with trends over time in the incidence and survival rates following hospital admission for acute myocardial infarction. We aim to explore the relationship between the LoS and its associated factors such as age, gender (years), hr (initial heart rate by beats per minute), BMI (body mass index by

Table 3 The prior mean and covariance matrix for $\boldsymbol{\beta}$

τ	$\boldsymbol{\beta}_a$	$Cov(\boldsymbol{\beta})$
0.25	[1.882, 0.489, −3.344]	$\begin{bmatrix} 28.696 & -27.642 & 224.511 \\ -27.642 & 27.209 & -221.267 \\ 224.512 & -221.267 & 1802.613 \end{bmatrix}$
0.50	[2.691, 0.349, −1.482]	$\begin{bmatrix} 0.107 & 0.007 & -0.114 \\ 0.007 & 0.003 & 0.003 \\ -0.114 & -0.013 & 0.136 \end{bmatrix}$
0.75	[2.320, 0.271, −0.674]	$\begin{bmatrix} 0.001 & 0.000 & -0.0010 \\ 0.000 & 0.000 & -0.001 \\ -0.001 & -0.000 & 0.002 \end{bmatrix}$

Table 4 Posterior mean, standard deviation and 95% credible interval of $\beta_k(\tau), k = 0, 1, 2$ from simulated example 2 based on a conjugate Normal-Inverse Gamma prior for $\boldsymbol{\beta}$

$\beta_k(\tau)$	$n_1 = 200$			$n_1 = 1000$			
	CI	Mean	S.D	CI	Mean	S.D	value
$\beta_0(.25)$	(7.894, 8.308)	8.075	0.111	(7.678, 8.131)	7.733	0.111	8
$\beta_0(.50)$	(8.881, 9.277)	9.061	0.098	(9.009, 9.051)	9.025	0.016	9
$\beta_0(.75)$	(9.697, 10.483)	10.086	0.186	(9.980, 10.116)	10.031	0.036	10
$\beta_1(.25)$	(1.976, 2.050)	2.007	0.020	(1.897, 2.016)	2.007	0.030	2
$\beta_1(.50)$	(1.926, 2.036)	1.996	0.025	(1.998, 2.003)	2.000	0.002	2
$\beta_1(.75)$	(1.926, 2.070)	1.999	0.034	(1.995, 2.018)	2.004	0.006	2
$\beta_2(.25)$	(−4.043, −3.867)	−3.977	0.042	(−4.006, −3.999)	−4.004	0.004	−4
$\beta_2(.50)$	(−4.032, −3.881)	−3.979	0.042	(−4.001, −3.983)	−3.998	0.003	−4
$\beta_2(.75)$	(−4.072, −3.679)	−3.930	0.101	(−4.014, −3.970)	−3.996	0.010	−4

kg/m^2), av3 (complete heart block), cvd (history of cardiovscular disease), sysbp (initial systolic blood pressure by mmHg) and diasbp (initial diastolic blood pressure by mmHg). Among covariates, gender (0=Male, 1=Female), av3 (0=No, 1=Yes) and cvd (0=No, 1=Yes) are binary variables. Age, hr, BMI, sysbp and diasbp as the continuous covariates are detailed in Table 5. The distribution of LoS is skewed and one is usually more interested in long stay or short stay than an average stay ([36, 37] and among others). We aim to investigate how these factors affect the LoS, from short LoS, to middle LoS and long LoS, so that we carry out the analysis in a complete range of the quantiles $\tau \in \{0.05, 0.1, 0.25, 0.50, 0.75, 0.95\}$.

Therefore, we fit a quantile regression model for LoS of the form:

$$Q_\tau(Y|X) = \beta_0(\tau) + \beta_1(\tau)\text{Age} + \beta_2(\tau)\text{Gender} + \beta_3(\tau)\text{hr} + \beta_4(\tau)\text{BMI}$$
$$+ \beta_5(\tau)\text{av3} + \beta_6(\tau)\text{cvd} + \beta_7(\tau)\text{sysbp} + \beta_8(\tau)\text{diasbp}.$$

Table 6 shows the posterior mean of all regression parameters under selected quantiles. The boxplots in Fig. 1 also display the posterior mean of these regression parameters across τs.

The values of the posterior mean of β_1, β_3, β_7 and β_8 in Table 6 clearly indicate that *age* and all *initial states of heart rate, systolic blood pressure and initial diastolic blood pressure*, have little effect on LoS (days), particularly on the low quantiles of

Table 5 Statistical description of the continuous covariates in data

	Max.	Min.	Mean	S.D.
Age	104	30	69.850	14.491
hr	186	35	87.018	23.586
BMI	44.839	13.045	26.614	5.406
sysbp	244	57	144.7	32.295
diasbp	198	6	78.270	21.545

Table 6 Posterior mean of regression parameters with different quantiles under the proposed method

τ	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.05	−0.382	0.007	0.751	0	0.021	0.072	−0.187	0.005	−0.001
0.1	−0.803	0.024	0.337	−0.002	0.036	−0.445	−0.224	0.004	−0.001
0.25	2.023	0.022	0.283	−0.004	0.012	−0.934	−0.309	−0.005	0.003
0.5	2.472	0.02	0.393	0.006	−0.006	0.12	0.133	−0.004	0.008
0.75	2.383	0.014	2.437	0.02	0.029	2.319	0.649	−0.004	0.01
0.9	1.221	0.023	1.785	0.051	0.023	3.171	1.73	−0.011	0.031
0.95	3.85	0.028	1.507	0.06	−0.051	4.341	2.181	−0.015	0.034

its distribution, whereas the values of the posterior mean of β_2, β_5 and β_6 show that *gender*, *complete heart block* and *history of cardiovscular disease* are those factors to affect the LoS most. Specifically, female patients tend to stay longer than male patients generally once they got those health problems and admitted into hospitals. Similarly, patients suffering *complete heart block* or *history of cardiovscular disease* stay much longer than patients without these problems, particularly for very long stay needed.

Based on the posterior mean of β_4 in Table 6, the effect on LoS from *BMI* is not big but generally negative except on the median and 95% quantile of the distribution of LoS.

If we compare the fitted *median* quantiel regression of LoS with a Poisson mean regression below,

$$Mean/LoS = 1.388 + 0.001(\tau)\text{Age} + 0.140(\tau)\text{Gender} + 0.003(\tau)\text{hr}$$
$$- 0.005(\tau)\text{BMI} + 0.113(\tau)\text{av3} + 0.071(\tau)\text{cvd}$$
$$- 0.001(\tau)\text{sysbp} + 0.003(\tau)\text{diasbp}.$$

We can see that both the proposed median model and Poisson mean model provide consistent conclusion, but Poisson regression can not explored the short and long LoS.

Unlike the jittering method for count (Machado and Silva, 2005), the proposed method in this paper is density function based Bayesian inference. However, if we compare our findings to the results from the count model of Machado and Silva (2005) in Table 7, we can draw similar conclusions to those from Table 6. The difference is that the proposed method can show that *complete heart block* or *history of cardiovscular disease* increase the long LoS significantly, which is true in real situations and much significant than what explored in the count model of Machado and Silva (2005).
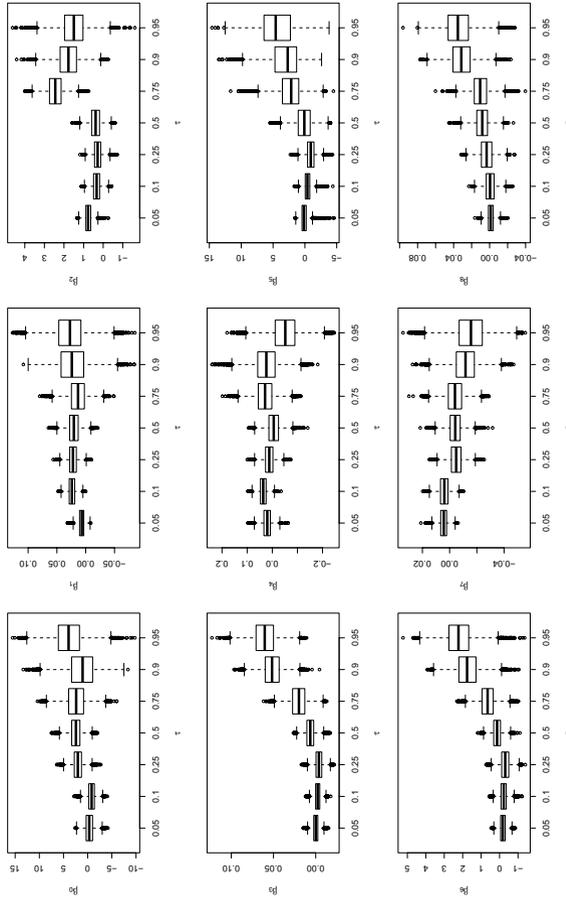
Fig. 1 Boxplots of posterior mean of regression parameters

6 Discussion

Discrete responses or count data are common in many disciplines. Regression analysis of discrete responses has been an active and promising area of research. Data with discrete responses may present features of skewness, fat-tailed and lepto-kurtic. Quantile regression is a more suitable tool to analyse this type of data than mean regression. We propose Bayesian quantile regression and Bayesian expectile regression for discrete responses. This is achieved by using a discrete asymmetric Laplace distribution and discrete asymmetric normal distribution to form the likelihood function, respectively. The method is shown robust numerically and coherent theoretically. The Bayesian approach which is fairly easy to implement and provides complete univariate and joint posterior distributions of parameters of interest. The posterior distributions of the unknown model parameters are obtained by using M-H algorithm implemented in R. We have shown promising results through Monte Carlo simulation studies and one real data analysis.

Appendix A. Proof of Theorem 3.1

A parametrization of the DALD in Eq.(5) leads to the following alternative form,

$$\phi(y;\mu,p,q) = \begin{cases} p^{y-\mu}(1-p)\log q, & y \geq \mu, y \in \mathbb{Z} \\ q^{y-\mu}(1-q)\log p, & y < \mu, y \in \mathbb{Z} \end{cases} \tag{A1}$$

where the parameters p and q ($0 < p, q < 1$) are related to τ via the relationships $p = \exp\{-\tau\}$ and $q = \exp\{1 - \tau\}$.

Lemma A.1 *The p.m.f. $\phi(t)$ defined in Eq.(A1) is bounded by $p^{|t|}(1-q)\log p$ and $q^{|t|}(1-p)\log q$.*

Proof of Lemma A.1 Expand $\phi(t)$ as a mixture of g, consider $0 < q \leq p < 1$,

Table 7 Posterior mean of regression parameters with different quantiles under Machado and Silva (2005)

τ	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0.05	−0.667	0.007	0.302	0.001	0.015	0.044	−0.130	0.003	−0.000
0.1	−0.177	0.010	0.132	−0.001	0.012	−0.080	−0.080	0.012	−0.000
0.25	0.999	0.007	0.068	−0.001	0.003	−0.301	−0.059	−0.002	0.002
0.5	1.144	0.004	0.105	0.002	−0.002	0.121	0.045	−0.001	0.002
0.75	1.201	0.002	0.348	0.003	0.004	0.364	0.079	−0.001	0.002
0.9	1.230	0.005	0.1816	0.006	0.001	0.107	0.210	−0.002	0.004
0.95	2.560	−0.001	0.095	0.007	−0.020	0.453	0.261	−0.001	0.001

$$\phi(t) = p^{|t|}(1-p)\log q \mathbb{I}(t \geq 0) + q^{|t|}(1-q)\log p \mathbb{I}(t < 0)$$
$$\leq (1-q)\log p \left(p^{|t|}\mathbb{I}(t \geq 0) + q^{|t|}\mathbb{I}(t < 0) \right)$$
$$\leq p^{|t|}(1-q)\log p.$$

Also,

$$\phi(t) = p^{|t|}(1-p)\log q \mathbb{I}(t \geq 0) + q^{|t|}(1-q)\log p \mathbb{I}(t < 0)$$
$$\geq (1-p)\log q \left(p^{|t|}\mathbb{I}(t \geq 0) + q^{|t|}\mathbb{I}(t < 0) \right)$$
$$\geq q^{|t|}(1-p)\log q.$$

Now, it is known that $g(t;a) = a^{|t|}(a-1)\log a$ with $0 < a < 1$ is a increasing function of t. Therefore, $\phi(t)$ has upper bound $h(p,q)q^{|t|}$ and lower bound $h(q,p)p^{|t|}$. The same procedure may be easily adapted to $q \geq p$. □

Lemma A.2 *For any constant* $a(0 < a < 1)$ *and sample size* $n > m$,

$$\int \prod_{k=0}^{m} |\beta_k|^{r_k} \prod_{i=1}^{n} \exp\{(\log a)|Y_i - X_i^T \beta|\}d\beta < \infty.$$

Proof of Lemma A.2 Without loss of generality and consider $m = 1$ for simplicity, then $X_i^T \beta = \beta_0 + \beta_1 X_{1i}$,

$$\int_{\mathbb{R}^2} |\beta_0|^{r_0}|\beta_1|^{r_1} \exp\left\{ (\log a) \sum_{i=1}^{n} |Y_i - X_i^T \beta| \right\} d\beta_0 d\beta_1$$
$$\leq \int_{\mathbb{R}^2} |\beta_0|^{r_0} \exp\left\{ (\log a)|\beta_0 + \beta_1 X_{11} - Y_1| \right\} |\beta_1|^{r_1}$$
$$\times \exp\left\{ (\log a)|\beta_0 + \beta_1 X_{12} - Y_2| \right\} d\beta_0 d\beta_1.$$

Since the double-integration $\int_{\mathbb{R}^2} |U|^{r_0} \exp(-|U+V+c_1|)|V|^{r_1}\exp(-|U+V+c_2|)dUdV$ is finite for any constants $c_1, c_2, r_0 \geq$ and $r_1 \geq 0$, Lemma 2 is proved. □

Theorem 3.1 below establishes that in the absence of any realistic prior information we could legitimately use an improper uniform prior distribution for all the components of β.

Theorem 3.1 *Assume the posterior is given by Eq.(8) and* $\pi(\beta) \propto 1$, *then all posterior moments of* β *in Eq.(9) exist.*

Proof of Theorem 3.1 We need to prove that

$$\int_{\mathbb{R}^{m+1}} \prod_{k=0}^{m} |\beta_k|^{r_k} \exp\left\{ (\log a) \sum_{i=1}^{n} |Y_i - X_i^T \beta| \right\} d\beta,$$

is finite. According to Lemma A.1 and Lemma A.2, it suffices to be proved. □

Appendix B. Property in Sect. 4

Under Bayesian inference of expectile μ for the discrete random variable Y, it can be proved that the posterior distribution is proper with regards to improper priors for the unknown parameters.

Similar to Lemma A.1, the p.m.f $\phi^{(E)}(t)$ defined in Eq.(13) is bounded. Consider $\theta \leq 0.5$,

$$
\begin{aligned}
\phi^{(E)}(t) &= k\sqrt{\frac{\pi}{\theta}}\left[\Phi\left(\sqrt{2\theta}(t+1)\right) - \Phi\left(\sqrt{2\theta}t\right)\right]\mathbb{1}_{t>0} \\
&\quad + k\sqrt{\frac{\pi}{1-\theta}}\left[\Phi\left(\sqrt{2(1-\theta)}(t+1)\right) - \Phi\left(\sqrt{2(1-\theta)}t\right)\right]\mathbb{1}_{t\leq 0} \\
&\leq k\sqrt{\frac{\pi}{\theta}}\left(\left[\Phi\left(\sqrt{2\theta}(t+1)\right) - \Phi\left(\sqrt{2\theta}t\right)\right]\right)\mathbb{1}_{t>0} \\
&\quad + k\sqrt{\frac{\pi}{\theta}}\left(\left[\Phi\left(\sqrt{2(1-\theta)}(t+1)\right) - \Phi\left(\sqrt{2(1-\theta)}t\right)\right]\right)\mathbb{1}_{t\leq 0} \\
&\leq k\sqrt{\frac{\pi}{\theta}}\left[\Phi\left(\sqrt{2(1-\theta)}(t+1)\right) - \Phi\left(\sqrt{2(1-\theta)}t\right)\right] \\
&= k\sqrt{\frac{2\pi(1-\theta)}{\theta}}\Phi'(t).
\end{aligned}
$$

Also,

$$
\begin{aligned}
\phi^{(E)}(t) &= k\sqrt{\frac{\pi}{\theta}}\left[\Phi\left(\sqrt{2\theta}(t+1)\right) - \Phi\left(\sqrt{2\theta}t\right)\right]\mathbb{1}_{t>0} \\
&\quad + k\sqrt{\frac{\pi}{1-\theta}}\left[\Phi\left(\sqrt{2(1-\theta)}(t+1)\right) - \Phi\left(\sqrt{2(1-\theta)}t\right)\right]\mathbb{1}_{t\leq 0} \\
&\geq k\sqrt{\frac{\pi}{1-\theta}}\left(\left[\Phi\left(\sqrt{2\theta}(t+1)\right) - \Phi\left(\sqrt{2\theta}t\right)\right]\right)\mathbb{1}_{t>0} \\
&\quad + k\sqrt{\frac{\pi}{1-\theta}}\left(\left[\Phi\left(\sqrt{2(1-\theta)}(t+1)\right) - \Phi\left(\sqrt{2(1-\theta)}t\right)\right]\right)\mathbb{1}_{t\leq 0} \\
&\geq k\sqrt{\frac{\pi}{1-\theta}}\left[\Phi\left(\sqrt{2\theta}(t+1)\right) - \Phi\left(\sqrt{2\theta}t\right)\right] \\
&= k\sqrt{\frac{2\pi\theta}{1-\theta}}\Phi'(t),
\end{aligned}
$$

where $k = \frac{2}{\sqrt{\pi}}\frac{\sqrt{\theta(1-\theta)}}{\sqrt{\theta}+\sqrt{1-\theta}}$, $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution.

Now, it is known that $\sqrt{\frac{\theta}{1-\theta}} \leq \sqrt{\frac{1-\theta}{\theta}}$ for $\theta \leq 0.5$. Therefore, $\phi^{(E)}(t)$ has upper bound $k\sqrt{\frac{2\pi(1-\theta)}{\theta}}\Phi'(t)$ and lower bound $k\sqrt{\frac{2\pi\theta}{1-\theta}}\Phi'(t)$. The same procedure may be easily adapted to $\theta \geq 0.5$.

Then, for $n > m$, we also have

$$\int \prod_{k=0}^{m} |\beta_k|^{r_k} \prod_{i=1}^{n} \exp\{-(Y_i - X_i^T \boldsymbol{\beta})^2\} d\boldsymbol{\beta} < \infty.$$

Similarly, consider $m = 1$ for simplicity, then $X_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 X_{1i}$,

$$\int_{\mathbb{R}^2} |\beta_0|^{r_0} \beta_1|^{r_1} \exp\{-(Y_i - X_i^T \boldsymbol{\beta})^2\} d\beta_0 d\beta_1$$

$$\leq \int_{\mathbb{R}^2} |\beta_0|^{r_0} \exp\{-(\beta_0 + \beta_1 X_{11} - Y_1)^2\} |\beta_1|^{r_1}$$

$$\times \exp\{-(\beta_0 + \beta_1 X_{12} - Y_2)^2\} d\beta_0 d\beta_1.$$

As is known to all that the double-integration is finite for any constants.

Therefore, assume the likelihood is given by Eq.(8) and $\pi(\boldsymbol{\beta}) \propto 1$, then it can be proved that all posterior moments of $\boldsymbol{\beta}$ in Eq.(9) exist.

Appendix C. Worcester Heart Attack Study Data Set

Data set used in the illustration of the proposed novel method in Sect. 5.3 is available in R-package *smoothHR*.

References

1. Efron B (1991) Resgression percentiles using asymmetric squared error loss. Stat Sinica 1:93–125
2. Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge
3. Waltrup LS, Sobotka F, Kneib T et al (2015) Expectile and quantile regression-david and goliath? Stat Modell 15(5):433–456
4. Ehm W, Gneiting T, Jordan A et al (2016) Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. J R Stat Soc Ser B 78(3):505–562
5. Yu K, Lu Z, Stander J (2003) Quantile regression: applications and current research areas. J R Stat Soc Ser D (The Statistician) 52(3):331–350
6. Yang Y, Wang HJ, He X (2016) Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. Int Stat Rev 84(3):327–344
7. Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447
8. Chen C, Yu K (2009) Automatic bayesian quantile regression curve fitting. Stat Comput 19(3):271–281
9. Yu K (2002) Quantile regression using rjmcmc algorithm. Comput Stat Data Anal 40(2):303–315
10. Klakattawi HS, Vinciotti V, Yu K (2018) A simple and adaptive dispersion regression model for count data. Entropy 20(2):142

11. Manski CF (1985) Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. J Econ 27(3):313–333
12. Kordas G (2006) Smoothed binary regression quantiles. J Appl Econ 21(3):387–407
13. Horowitz JL (1992) A smoothed maximum score estimator for the binary response model. Econ J Econ Soc 60:505–531
14. Aristodemou K, He J, Yu K (2019) Binary quantile regression and variable selection: a new approach. Econ Rev 38(6):679–694
15. Yu K, Stander J (2007) Bayesian analysis of a tobit quantile regression model. J Econ 137(1):260–276
16. Benoit DF, Van den Poel D (2012) Binary quantile regression: a bayesian approach based on the asymmetric laplace distribution. J Appl Econ 27(7):1174–1188
17. Benoit DF, Alhamzawi R, Yu K (2013) Bayesian lasso binary quantile regression. Comput Stat 28(6):2861–2873
18. Rahman MA (2016) Bayesian quantile regression for ordinal models. Bayesian Anal 11(1):1–24
19. Rahman MA, Karnawat S (2019) Flexible bayesian quantile regression in ordinal models. In: Jeliazkov I, Tobias JL (eds) Topics in identification, limited dependent variables, partial observability, experimentation, and flexible modeling: part b. Emerald Publishing Limited, Bingley
20. Alhamzawi R, Ali HTM (2018) Bayesian quantile regression for ordinal longitudinal data. J Appl Stat 45(5):815–828
21. Rahman MA, Vossmeyer A (2019) Estimation and applications of quantile regression for binary longitudinal data. In: Jeliazkov I, Tobias JL (eds) Topics in identification, limited dependent variables, partial observability, experimentation, and flexible modeling: part b. Emerald Publishing Limited, Bingley
22. Omata Y, Katayama H, Arimura TH (2017) Same concerns, same responses? a bayesian quantile regression analysis of the determinants for supporting nuclear power generation in Japan. Environ Econ Policy Stud 19(3):581–608
23. Ojha M, Rahman MA (2020) Do online courses provide an equal educational value compared to in-person classroom teaching? evidence from us survey data using quantile regression. Evidence from US Survey Data Using Quantile Regression (July 14, 2020)
24. Kneib T (2013) Beyond mean regression. Stat Modell 13(4):275–303
25. Machado JAF, Silva JS (2005) Quantiles for counts. J Am Stat Assoc 100(472):1226–1237
26. Canale A, Dunson DB (2011) Bayesian kernel mixtures for counts. J Am Stat Assoc 106(496):1528–1539
27. Roy D (1993) Reliability measures in the discrete bivariate set-up and related characterization results for a bivariate geometric distribution. J Multivar Anal 46(2):362–373
28. Bissiri PG, Holmes C, Walker SG (2016) A general framework for updating belief distributions. J R Stat Soc Ser B (Stat Methodol) 78(5):1103–1130
29. Fernández C, Steel MF (1998) On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 93(441):359–371
30. Newey WK, Powell JL (1987) Asymmetric least squares estimation and testing. Econometrica 55:819–847
31. Yao Q, Tong H (1996) Asymmetric least squares regression estimation: a nonparametric approach. J Nonparametr Stat 6(2–3):273–292
32. Alhamzawi R, Yu K (2013) Conjugate priors and variable selection for bayesian quantile regression. Comput Stat Data Anal 64:209–219
33. Smith M, Kohn R (1996) Nonparametric regression using Bayesian variable selection. J Econ 75(2):317–343
34. Chen RB, Chu CH, Lai TY et al (2011) Stochastic matching pursuit for Bayesian variable selection. Stat Comput 21(2):247–259
35. Hosmer DW, Lemeshow S, May S (2008) Model development. Regression modeling of time-to-event data, 2 Edition, applied survival analysis, pp 132–168
36. Borghans I, Hekkert KD, den Ouden L et al (2014) Unexpectedly long hospital stays as an indicator of risk of unsafe care: an exploratory study. BMJ Open 4(6):e004773
37. Wolkewitz M, Zortel M, Palomar-Martinez M et al (2017) Landmark prediction of nosocomial infection risk to disentangle short-and long-stay patients. J Hosp Infect 96(1):81–84