

# Bayesian Multiple Quantile Regression for Linear Models Using a Score Likelihood\*

Teng Wu<sup>†</sup> and Naveen N. Narisetty<sup>‡</sup>

**Abstract.** We propose the use of a score based working likelihood function for quantile regression which can perform inference for multiple conditional quantiles of an arbitrary number. We show that the proposed likelihood can be used in a Bayesian framework leading to valid frequentist inference, whereas the commonly used asymmetric Laplace working likelihood leads to invalid interval estimations and requires further correction. For computation, we propose a novel adaptive importance sampling algorithm to compute important posterior summaries such as the posterior mean and the covariance matrix. Our proposed approach makes it feasible to perform valid inference for parameters such as the slope differences at different quantile levels, which is either not possible or cumbersome using existing Bayesian approaches. Empirical results demonstrate that the proposed likelihood has good estimation and inferential properties and that the proposed computational algorithm is more efficient than its competitors.

**Keywords:** Bayesian quantile regression, multiple quantile regression, working likelihood, adaptive importance sampling.

## 1 Introduction

Quantile regression (QR), first introduced by Koenker and Bassett (1978), provides a more comprehensive description of the relationship between an outcome and a set of covariates of interest compared to the more commonly used conditional mean regression. Moreover, it is also more robust against outliers. Suppose  $Y$  denotes the continuous response variable of interest and  $\mathbf{X}$  denotes the  $p$ -dimensional covariate vector, with its first element equals to 1. For a given quantile level  $\tau \in (0, 1)$ , the linear quantile regression model assumes the  $\tau$ th conditional quantile function of  $Y$  given  $\mathbf{X}$  to be

$$Q_Y(\tau \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}(\tau),$$

where  $\boldsymbol{\beta}(\tau)$  is the regression quantile vector of length  $p$  which depends on the quantile level of interest  $\tau$ . By allowing the slopes to change with  $\tau$ , quantile regression allows for a much more general conditional relationship between  $Y$  and  $\mathbf{X}$  (Koenker, 2005).

For a given sample  $\{(y_i, \mathbf{x}_i)\}_{i=1, \dots, n}$ , the standard quantile regression estimator (Koenker and Bassett, 1978) is given by

$$\hat{\boldsymbol{\beta}}(\tau) = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i^T \boldsymbol{\beta}), \quad (1.1)$$

---

\*The authors gratefully acknowledge funding support from NSF grants DMS-1811768 and CAREER-1943500.

<sup>†</sup>University of Illinois at Urbana-Champaign

<sup>‡</sup>University of Illinois at Urbana-Champaign, [naveen@illinois.edu](mailto:naveen@illinois.edu)

where  $\rho_\tau(u) = u(\tau - I(u < 0))$  is the asymmetric check loss function. Inference on the quantile regression parameter can be carried out based on the asymptotic normality results in Koenker and Bassett (1978); Koenker (2005). This requires estimating the joint covariance matrix of two quantile levels which involves density estimation that can be sensitive to bandwidth selection. Bayesian approaches have an advantage as they can automatically estimate the variance and more generally the posterior distribution of the parameters without having to estimate densities based on which credible intervals can be constructed (Yu and Moyeed, 2001; Yang and He, 2012; Yang et al., 2016).

The main challenge for Bayesian methods is that the quantile regression model does not specify a parametric likelihood, which is required in the Bayesian framework. Yu and Moyeed (2001) proposed to use a working likelihood based on the Asymmetric Laplace Distribution (ALD) given by

$$ALD(Y | \mathbf{X}, \beta) = \frac{\tau^n (1 - \tau)^n}{\sigma^n} \exp \left\{ -\frac{\sum_{i=1}^n \rho_\tau(y_i - \mathbf{x}_i^T \beta)}{\sigma} \right\},$$

where  $\sigma$  is a fixed scale parameter. The intuition behind this likelihood is that it incorporates the check loss into the likelihood function and is maximized at the standard QR estimator  $\hat{\beta}(\tau)$ . This is called a working likelihood since quantile regression problem does not specify any particular data generating process and there is no “true” likelihood function. Computation of the posterior with the ALD likelihood is easy to implement using Gibbs sampling (Kozumi and Kobayashi, 2011) or Metropolis-Hastings algorithms. Sriram et al. (2013) showed that the resultant posterior exhibits posterior consistency properties. However, Sriram (2015); Yang et al. (2016) noticed that the asymptotic variance of the posterior distribution based on ALD likelihood is incorrect, implying invalidity of inference from a frequentist viewpoint. To remedy this, the authors proposed a correction to the posterior variance to make valid inference. It is worth noting that if one takes a subjective Bayesian perspective and truly believes that the data is generated from a Laplace distribution, the inference using the ALD likelihood will be valid. However, that is a very strong assumption and is likely to be violated in most applications where quantile regression is of interest. Therefore, we take an objective Bayesian perspective where it is desirable to have valid frequentist properties for the Bayesian procedures.

Estimation and especially inference for regression quantiles at multiple quantile levels is more challenging. For two different quantile levels  $\tau_1, \tau_2$  and  $\tau_1 > \tau_2$ , suppose we are interested in testing the following hypothesis,

$$H_0 : \beta_1(\tau_1) = \beta_1(\tau_2) \text{ vs } H_1 : \beta_1(\tau_1) \neq \beta_1(\tau_2),$$

i.e. whether the first covariate  $X_1$  has a different effects on  $Y$  at quantiles  $\tau_1$  and  $\tau_2$ . In the frequentist context, hypothesis testing can be performed based on the asymptotic normality results for multiple quantiles as in Koenker and Bassett (1978); Koenker (2005) or several other simultaneous estimation approaches proposed in the literature such as He (1997); Liu and Wu (2009); Zou et al. (2008); Bondell et al. (2010).

To test this in the Bayesian framework, we can reparametrize the quantile regression model with  $\theta = \beta_1(\tau_1) - \beta_1(\tau_2)$  after which testing the slope difference is transformed to testing whether  $\theta = 0$ . If zero is not included in the credible interval based on the

posterior distribution for  $\theta$ , we would reject the null hypothesis. For this procedure, we require a valid posterior distribution that can automatically capture the joint uncertainty of the regression quantiles at multiple levels. Sriram et al. (2016) proposed a working likelihood to model multiple quantile levels as an extension of the ALD likelihood and showed it provides consistent estimation of the regression quantiles. However, it cannot quantify the joint uncertainty for multiple quantile levels and the inference based on this posterior is not valid as discussed earlier. Although the correction proposed by Yang et al. (2016) can be generalized for this case, it requires fitting two different likelihoods that are not compatible with each other and is not preferable also since it would not utilize the information between two nearby quantile levels.

Bayesian methods for simultaneously modeling conditional quantiles at multiple levels  $\tau_1, \tau_2, \dots, \tau_m$  have also been considered. For estimating multiple unconditional quantiles in a Bayesian framework, Lavine (1995) proposed a multinomial substitution likelihood. Dunson and Taylor (2005) extended Lavine (1995)'s idea and proposed a generalization of this substitution likelihood for the regression case with categorical covariates having a limited number of categories. However, this likelihood does not generalize to the usual quantile regression setting with continuous predictors. There are several existing Bayesian approaches which accommodate continuous predictors but a majority of them have computational or inferential challenges in modeling a finite collection of quantile levels which forms a major motivation for our current work.

Lancaster and Jae Jun (2010) used Bayesian exponentially tilted empirical likelihood and provided an explicit form for the posterior density. Yang and He (2012) proposed a Bayesian empirical likelihood method which also allows a simultaneous estimation of multiple quantiles. This approach provides valid posterior inference unlike the ALD likelihood based approach and it is applicable to model quantile slope differences. However, the empirical likelihood based methods are computationally demanding even under modest dimensions. Feng et al. (2015) proposed an approximation likelihood based on linearly interpolated density targeting multiple quantile regression but they require the number of quantiles to go to infinity which is not efficient when we are only interested in a finite collection of quantile levels. If quantile regression model is assumed for all the quantile levels, a global likelihood for all the regression quantiles can be specified because the conditional density can be written as a function of the conditional quantiles (Reich et al., 2011; Yang and Tokdar, 2017). Reich et al. (2011) proposed a model based spatial quantile regression method within the Bayesian framework by assuming a global quantile regression model at all the quantile levels. Using a Bernstein basis polynomial representation, their approach is to simultaneously estimate the regression quantile process, which is computationally expensive especially when we are only interested in estimating a few quantile levels. Moreover, unlike our proposed approach in the paper, this approach does not ensure validity of the coverage based on credible intervals. Yang and Tokdar (2017) proposed a semiparametric Bayesian quantile regression method by using a Gaussian process prior distributions on the function-valued parameters. However, the resultant estimators are not necessarily  $\sqrt{n}$ -consistent due to the semiparametric nature of the priors. The likelihood evaluation is computationally intensive and not suitable if the interest lies in only a few quantile levels. Also, it does not guarantee the frequentist validity of the resultant posterior intervals.

In this paper, we propose the use of a working likelihood based on the score function for implementing quantile regression at multiple quantile levels and propose a new adaptive importance sampling algorithm for its computation. The proposed likelihood can be used for Bayesian settings and we provide theoretical justification for this approach by showing that the posterior corresponding to the proposed likelihood provides valid frequentist inference for multiple quantile levels. With the ability to incorporate multiple quantiles into one single likelihood, we can perform inference on parameters such as the slope differences directly, which is not easy to perform using existing Bayesian approaches. Compared to the frequentist approaches, our formulation allows automatic quantification of uncertainty of multiple regression quantiles, the flexibility of incorporating prior information, additional shrinkage or smoothness in the regression quantile parameters by using an appropriate prior distribution. Computationally, we propose an adaptive importance sampling algorithm, which does not require any optimization steps and is suitable for estimating multiple quantile levels together efficiently. Our empirical results show that the proposed importance sampling algorithm is very efficient and is more efficient than the computational algorithms typically used for the ALD likelihood.

For the purpose of computing generalized method of moment (GMM) estimators, Chernozhukov and Hong (2003) proposed using Markov Chain Monte Carlo (MCMC) algorithms based on likelihoods constructed by exponentiation of the objective function of interest. While the score likelihood advocated in this paper is considered by Chernozhukov and Hong (2003), the focus of our current work is to generalize the score likelihood to perform valid inference for parameters involving multiple quantile levels which is not feasible using existing Bayesian approaches and to demonstrate that our proposed computational algorithm is more efficient in terms of effective sample size than the Metropolis-Hasting algorithm proposed by Chernozhukov and Hong (2003) for generic likelihoods.

The rest of the paper is structured as follows: Section 2 will first present the formulation of our likelihood. Section 3 provides our proposed computational methods which can be used to compute the posterior and make inference based on the proposed likelihood. Section 4 demonstrates the performance of the proposed method and computational algorithms via simulation studies. Section 5 provides an application to a real dataset. Finally, Section 6 provides a conclusion.

## 2 Methodology and Theoretical Results

### 2.1 Score Based Likelihood for Quantile Regression

We first focus on the case for modeling a single quantile level. The score function for the quantile regression objective function given in (1.1) is a  $p$  dimensional vector, which takes the following form:

$$s_\tau(\beta) = \sum_{i=1}^n \mathbf{x}_i \psi_\tau(y_i - \mathbf{x}_i^T \beta),$$

where  $\psi_\tau(u) = \tau - I(u < 0)$ . Since the quantile regression estimator  $\hat{\beta}(\tau)$  minimizes the quantile loss function,  $s_\tau(\hat{\beta}(\tau)) \approx 0$ . Consider the following working likelihood:

$$L(Y | \mathbf{X}, \beta) = C \exp \left\{ -\frac{1}{2n} s_\tau(\beta)^T W s_\tau(\beta) \right\}, \quad (2.1)$$

where  $W$  is a  $p \times p$  positive definite weight matrix, and  $C$  is a constant free of  $\beta$ . We shall also use the short notation  $L(\beta)$  for  $L(Y | \mathbf{X}, \beta)$  as it is viewed as a function of  $\beta$ . Since the exponent in the posterior is a quadratic form, values of  $\beta$  for which  $s_\tau(\beta)$  is close to zero will correspond to high posterior density. As the expectation of the score function equals zero uniquely at the true regression quantile  $\beta_0(\tau)$ , we expect that this leads to posterior concentration around the truth. Our Theorem 1 makes this intuition more formal under regularity conditions. Quasi-posteriors for Bayesian inference have been considered in the literature, for example, we refer to Li and Jiang (2016) for quasi-posterior inference based on Bayesian generalized method of moments. Chernozhukov and Hong (2003) considered such working likelihoods in the context of computation of frequentist estimators using MCMC algorithms.

The weight matrix we use in the likelihood is:

$$W = \frac{n}{\tau(1-\tau)} \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right)^{-1}.$$

Our theoretical results from Section 2.4 show that this choice for  $W$  will lead to valid inference based on the posterior. Compared with the ALD likelihood, this will allow us to carry out posterior inference directly without having to correct the posterior covariance as suggested by (Sriram, 2015; Yang et al., 2016).

The likelihood (2.1) does not arise from a distributional specification for  $Y$  given  $(\mathbf{X}, \beta)$ . Therefore, this can be viewed as a working likelihood. To visualize the score likelihood in (2.1), we provide a simple illustration using the univariate case without covariates. We generate samples from  $N(0, 1)$  and plot the score likelihood for the median in Figure 1. In Figure 1, we take numerical integration of the likelihood over the interval  $[-0.5, 0.5]$ , which is large enough, to obtain a good approximation the normalizing constant  $C$ . The ALD likelihood with  $\sigma = 1$  is also provided for comparison. The dashed black curve in Figure 1 is the true asymptotic distribution for the median. Since all the likelihoods provide consistent estimation, we centered all of them to zero to provide a better comparison of their variances. It can be seen that the variance of the ALD likelihood is different from the true asymptotic distribution as it is affected by the scale parameter  $\sigma$ . On the other hand, the score based likelihood is quite close to the asymptotic normal distribution, especially as the sample size increases.

The score likelihood (2.1) can also be generalized to model multiple conditional quantiles. Consider modeling  $m$  quantiles  $\tau_1, \tau_2, \dots, \tau_m$  at the same time, with the true quantile regression parameters  $\beta_0 = (\beta(\tau_1), \beta(\tau_2), \dots, \beta(\tau_m))$ . For any vector  $\beta = (\beta_1, \beta_2, \dots, \beta_m)$ , define the score function to be

$$s(\beta) = (s_{\tau_1}(\beta_1), s_{\tau_2}(\beta_2), \dots, s_{\tau_m}(\beta_m)),$$

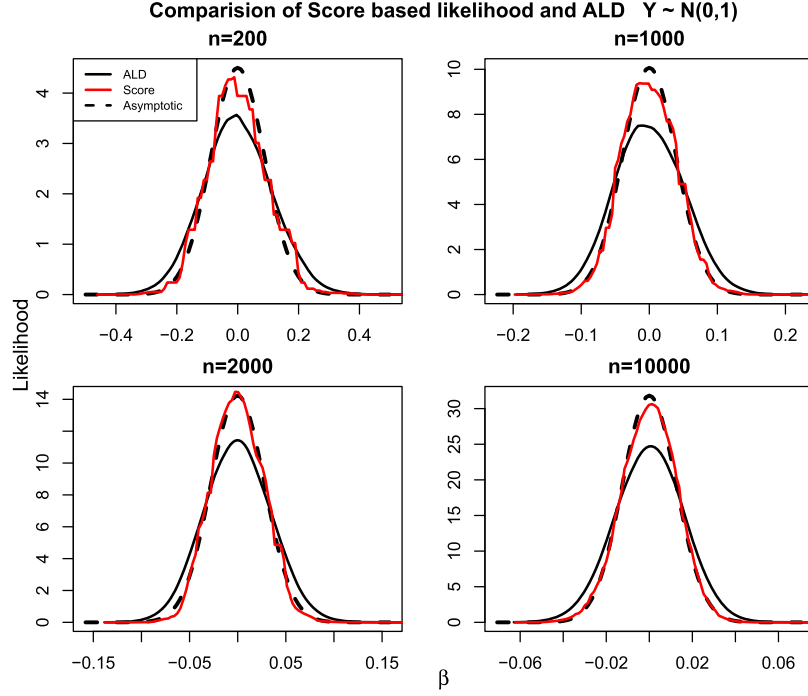


Figure 1: Comparison of the score based likelihood (solid red) and the asymmetric Laplace (ALD) likelihood (solid black) with the asymptotic distribution (dashed black). All the likelihoods are centered at zero.

where  $s_{\tau_i}(\cdot)$  is the score function correspond to the quantile level  $\tau_i$  for  $i = 1, \dots, m$ . The score based likelihood for multiple quantile levels takes the following form:

$$L(\beta) \propto \exp \left\{ -\frac{1}{2n} s(\beta)^T W s(\beta) \right\}, \quad (2.2)$$

where  $W$  is an  $mp \times mp$  positive definite weight matrix. The weight matrix in this case is given by

$$W = (Q \otimes G)^{-1}, \text{ where } Q = (\tau_i \wedge \tau_j - \tau_i \tau_j)_{ij}, \quad G = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T,$$

$\otimes$  denotes the Kronecker product, and  $x \wedge y = \min(x, y)$ . Our theoretical results show that this  $W$  will lead to correct posterior inference. It is also natural to incorporate monotonicity of the conditional quantiles so that the proposed likelihood becomes:

$$L(\beta) \propto \exp \left\{ -\frac{1}{2n} s(\beta)^T W s(\beta) \right\} \mathbb{1} \{ \mathbf{x}_i \beta_1 \leq \mathbf{x}_i \beta_2 \leq \dots \leq \mathbf{x}_i \beta_m \text{ for all } i \}. \quad (2.3)$$

We place the monotonicity constraints on the estimated quantiles only at the observed  $x$ -values. This implies monotonicity of the conditional quantiles at all the covariate values

belonging to the convex hull formed by the observed  $x$ 's. In the literature, alternative approaches for achieving monotone quantile curves in the Bayesian context have been considered. For example, Rodrigues and Fan (2017) used a two stage approach where posterior samples from Bayesian quantile regression are made to be monotone using a post-hoc adjustment procedure. Yang and Tokdar (2017) proposed a more elegant way to impose monotonicity for all the quantile levels  $\tau \in (0, 1)$  by reparametrizing the quantile regression model. Both these approaches, however, can be too restrictive in terms of their constraints if we are only interested in a few quantile levels.

## 2.2 Prior Distributions

An advantage of the proposed Bayesian framework for quantile regression is that prior information or smoothness in the regression quantiles can be induced with appropriately specified prior distributions. When there is no additional information available on the regression quantile parameters, a bounded uniform prior such as  $\beta \in [-n, n]^{mp}$  can be used. Such a prior is imposed to ensure a proper posterior distribution. In practice, an improper uniform prior may also be used for simplicity. If the regression quantiles are expected to vary smoothly as a function of the quantile level  $\tau$ , this can be induced by using prior distributions with high correlations between regression parameters corresponding to nearby quantile levels. For example, we can use a multivariate normal distribution with the covariance between  $\beta_1(\tau_1)$  and  $\beta_1(\tau_2)$  to be a smooth covariance kernel function such as the squared exponential kernel given by

$$k(\tau_1, \tau_2) = \sigma^2 \exp \left\{ -\frac{1}{2b^2} (\tau_1 - \tau_2)^2 \right\}, \quad (2.4)$$

where  $\sigma^2$  and  $b$  are hyperparameters. For simplicity, the prior covariance between parameters of different predictors can be assumed to be zero as is common in the linear regression context (George and McCulloch, 1993; Narisetty and He, 2014). Such Gaussian priors have been used in quantile regression settings (Yang and Tokdar, 2017; Rodrigues and Fan, 2017) to induce smooth conditional quantiles. With the non-informative uniform prior or the Gaussian prior, the posterior will be a proper distribution and can be used for inference.

## 2.3 Estimation of Contrasts in Quantile Regression

Using scored-based likelihood and a prior distribution of  $\beta$ , we can obtain the posterior distribution of any linear combination of the quantile regression parameters, and in particular for any contrasts. For simplicity, consider a simple quantile regression model

$$Q_Y(\tau \mid X = x) = \beta_0(\tau) + \beta_1(\tau)x,$$

and we are interested in testing slope differences  $\theta = \beta_1(\tau_1) - \beta_1(\tau_2)$ . If we assume an bounded uniform prior, the posterior of  $\theta$  can be simply calculated from the posterior samples of  $\beta_1(\tau_1)$  and  $\beta_1(\tau_2)$ . In general, if prior information is available only on the

parameter  $\theta$ , we can place a prior  $\pi(\theta)$  on  $\theta$  and uniform prior on the other parameters, the posterior distribution of  $\theta$  would take the following form:

$$f(\theta, \beta_1(\tau_2), \beta_0(\tau_1), \beta_0(\tau_2) \mid Y, \mathbf{X}) \propto L(\beta_0(\tau_1), \theta + \beta_1(\tau_2), \beta_0(\tau_2), \beta_1(\tau_2))\pi(\theta).$$

Similarly, the proposed framework can be used for obtaining posterior samples for any linear combinations of the quantile regression coefficients across multiple quantile levels.

## 2.4 Posterior Concentration Results

In this section, we shall show that the score-based likelihood together with either of the priors discussed will lead to a posterior that is asymptotically similar to a normal distribution with the posterior variance matching the frequentist variance of the quantile regression estimator. Before we state the assumption and theorems, we define some notations used in this paper.

**Notation.** For a sequence of random variable  $X_n$  and sequence of constant  $a_n$ , the notation  $X_n = O_p(a_n)$  is defined by: For any  $\epsilon > 0$ , there exist an  $M > 0$  and  $N > 0$ , such that

$$P(|X_n/a_n| > M) < \epsilon, \text{ when } n > N,$$

and  $X_n = o_p(a_n)$  is defined by: For any  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n/a_n| > \epsilon) = 0.$$

$X_n = o_p(a_n)$  is equivalent to  $|X_n/a_n| \xrightarrow{P} 0$ , where  $\xrightarrow{P}$  denotes convergence in probability.

To obtain the asymptotic result, the following assumptions are made.

**Assumption 1** (On the distribution of  $X$ ).  $\max \|\mathbf{x}_i\| = O_p(n^{1/4} (\log n)^{-1/2})$  and  $E(\|\mathbf{x}_i\|^4) < B$  for all  $i$  and some constant  $B$ .

**Assumption 2** (On the data generation mechanism). *The conditional quantile function of  $Y$  given  $\mathbf{X}$  is linear,*

$$Q_Y(\tau \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}(\tau).$$

*Furthermore, the conditional distribution of  $Y_i$  given  $\mathbf{x}_i$ ,  $F_i(\cdot \mid \mathbf{x}_i)$ , is absolutely continuous with continuous density  $f_i(\cdot \mid \mathbf{x}_i)$  bounded away from 0, and uniformly bounded away from  $\infty$  in a neighborhood of  $\xi_i(\tau) = Q_{Y_i}(\tau \mid \mathbf{x}_i)$ , for  $i = 1, \dots, n$ . In addition, the first derivative of  $f_i(\cdot)$  is uniformly bounded in a neighborhood of  $\xi_i(\tau)$ .*

**Assumption 3** (On the existence of asymptotic limits).

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(\mathbf{x}_i \mathbf{x}_i^T) = D_0$$

for some matrix  $D_0$  and

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(f_i(\mathbf{x}_i^T \boldsymbol{\beta}(\tau_j)) \mathbf{x}_i \mathbf{x}_i^T) = D_1(\tau_j),$$



for some matrix  $D_1(\tau_j)$ , for  $j = 1, 2, \dots, m$ . The sample version of  $D_0$  converges,

$$\left\| n^{-1} \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \right) - D_0 \right\| \xrightarrow{p} 0.$$

We would like to point out that these are standard assumptions for quantile regression and for instance similar conditions have been used by He et al. (1996). We slightly modify the last two assumptions to accommodate our random design setting.

We further impose the following additional assumptions which appeared in Example 3 in Chernozhukov and Hong (2003).

**Assumption A1.** The expectation of score function is uniquely minimized at  $\beta_0$ , i.e.  $E[s_\tau(\beta)] = 0$  if and only if  $\beta = \beta_0$ .

**Assumption A2.** The function  $\mathbf{x}_i(\tau - I(Y_i \leq \mathbf{x}_i^T \beta))$  is a Donsker class as a function in  $\beta$ ,  $E[\sup_\beta \|\mathbf{x}_i(\tau - I(Y_i < \mathbf{x}_i^T \beta))\|^2] < \infty$ .

**Assumption A3.**  $G(\beta) = \nabla_\beta E[\mathbf{x}_i(\tau - I(y_i < \mathbf{x}_i^T \beta))]$  is continuous in  $\beta$ .  $G(\beta)^T W G(\beta)$  is positive definite and continuous in an open ball at  $\beta_0$ .

Assumptions A1–A3 are technical assumptions to ensure posterior concentration for score likelihood. Assumption A1 is regarding the identifiability of  $\beta_0$  and Assumptions A2–A3 guarantee regularity of the score function as a function of  $\beta$ . Heuristically speaking, these conditions will be satisfied if the underlying conditional distribution of the response is sufficiently smooth which is usually the case in most practical applications.

The following theorem shows that our proposed likelihood leads to asymptotically valid frequentist inference for multiple quantile levels  $\tau_1, \dots, \tau_m$  at the same time.

**Theorem 1.** a) (Asymptotic Normality of Posterior Distribution) Under Assumptions 1–3, consider the proposed likelihood in (2.2) with the bounded uniform prior and the following weight matrix

$$W = (Q \otimes G)^{-1}, \text{ where } Q = (\tau_i \wedge \tau_j - \tau_i \tau_j)_{ij} \text{ and } G = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T.$$

Consider  $\beta$  such that  $\|\beta - \beta_0\| = O(n^{-1/2})$ . The posterior density for  $\beta$  satisfies

$$p_n(\beta \mid \mathcal{D}) \propto \exp \left\{ -\frac{1}{2} n (\beta - \beta_0)^T \Sigma^{-1} (\beta - \beta_0) + o_p(1) \right\}, \quad (2.5)$$

where  $\Sigma$  is an  $mp \times mp$  matrix with the  $ij^{th}$  block  $\Sigma_{ij} = (\tau_i \wedge \tau_j - \tau_i \tau_j) D_1(\tau_i)^{-1} D_0 D_1(\tau_j)^{-1}$ . b) (Posterior concentration) Assume the true parameter  $\beta_0$  belongs to the interior of a compact convex set  $\Theta$ , and the prior distribution restricted to  $\Theta$ . Under the additional Assumptions A1–A3, the posterior distribution of  $\beta$  satisfies

$$\Pi(\|\beta - \beta_0\| > M_n n^{-1/2} \mid \mathcal{D}) \xrightarrow{p} 0$$

where  $M_n$  is any sequence diverging to infinity with  $n$ .

The first part of the Theorem is regarding the shape of the posterior around the true parameter. This result is analogous to the asymptotic results for the frequentist quantile regression estimators (Koenker and Bassett, 1978; Koenker, 2005). The second part of the theorem ensures posterior concentration, which requires stronger technical conditions as indicated by the results of Chernozhukov and Hong (2003) which are the basis for this part of the theorem. We defer those assumptions to the supplementary material (Wu and Narisetty, 2020). For the second part, the prior needs to be supported on a fixed compact set, which is for instance satisfied by a Uniform prior on a large fixed interval. One possibility to ensure this condition is to truncate any prior on a compact set. While we expect that this is not a necessary condition, we require it for our theoretical results. In practice, priors with unbounded support such as the Gaussian prior and a Uniform prior on  $[-n, n]$  also work well as observed in our simulation studies.

Notice that since we are in the Bayesian framework, it is possible to impose monotonicity constraints to avoid the crossing problem. Therefore, we only consider  $\beta_i$ 's such that  $\mathbf{x}_i\beta_1 \leq \mathbf{x}_i\beta_2 \leq \dots \leq \mathbf{x}_i\beta_m$  for all  $\mathbf{x}_i, i = 1, 2, \dots, n$ . Heuristically, since the crossing probability is small, the posterior distribution would not be affected by these constraints. However, in small samples, our simulation results show that the proposed joint likelihood leads to smaller mean squared error compared with estimating them separately using the frequentist approach.

## 2.5 Extension to Censored Quantile Regression

For censored data, quantile regression is preferred over conditional mean regression since the latter is not necessarily identifiable without making stringent distributional assumptions. Motivated by the Powell's estimator (Powell, 1984, 1986), the proposed likelihood can be extended to censored quantile regression with fixed censoring. Consider a simple case with fixed left censoring at 0,

$$y_i = \max(0, \mathbf{x}_i^T \beta_0 + u_i), \quad (2.6)$$

where  $u_i$  is the error term. The Powell's estimator is defined as

$$\hat{\beta}_P = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\tau}(y_i - \max(\mathbf{x}_i^T \beta, 0)). \quad (2.7)$$

The Powell's estimator is consistent and has asymptotic normality with covariance matrix  $\tau(1-\tau)M_1^{-1}M_0M_1^{-1}$ , where

$$M_0 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(\mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta_0 > 0)),$$

$$M_1 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(f_i(\mathbf{x}_i^T \beta_0) \mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta_0 > 0)).$$

Define the score function for censored quantile regression to be

$$c_{\tau}(\beta) = \sum_{i=1}^n \mathbf{x}_i \psi_{\tau}(y_i - \mathbf{x}_i^T \beta) I(\mathbf{x}_i^T \beta > 0). \quad (2.8)$$

We propose the following likelihood to be used for Bayesian censored quantile regression

$$cL(\beta) \propto \exp \left\{ -\frac{1}{2n} c_\tau(\beta)^T W(\beta) c_\tau(\beta) \right\}, \quad (2.9)$$

where the weight matrix  $W(\beta)$  is given by

$$W(\beta) = \left( \tau(1-\tau)n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta > 0) \right)^{-1}. \quad (2.10)$$

The motivation of our likelihood is that the likelihood is maximized at  $\hat{\beta}_P$  because  $c_\tau(\hat{\beta}_P) = o_p(\sqrt{n})$ , as discussed in Powell (1984) so that the posterior is consistent. We shall also show that with our choice of the weight matrix, the resultant posterior is approximately normal with the correct variance. Unlike the uncensored case, notice that the weight matrix now depends on the parameter  $\beta$  which makes the theoretical analysis more challenging. To obtain the asymptotic posterior concentration results for censored quantile regression, we need the following assumptions similar to the ones used by Powell (1986).

**Assumption 4.** For  $r = 0, 1, 2$  and  $\|\beta - \beta_0\| = O(n^{-1/2})$ ,

$$E(I(|\mathbf{x}_i^T \beta| \leq \|\mathbf{x}_i\| \cdot \|\beta - \beta_0\|) \|\mathbf{x}_i\|^r) \leq K \|\beta - \beta_0\|.$$

**Assumption 5.**

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(\mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta_0 > 0)) = M_0,$$

and

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n E(f_i(\mathbf{x}_i^T \beta_0) \mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta_0 > 0)) = M_1.$$

We also assume that the finite sample version of  $M_0$  converges to the true value.

**Assumption 6.**

$$\left\| n^{-1} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta_0 > 0)) - M_0 \right\| \xrightarrow{p} 0.$$

We further impose the following technical assumptions, which are also made for Proposition 1 in Chernozhukov and Hong (2003).

**Assumption A4.** The expectation of score function is uniquely minimized at  $\beta_0$ , i.e.  $E[c_\tau(\beta)] = 0$  if and only if  $\beta = \beta_0$ .

**Assumption A5.**

$$\sup_{\beta \in \Theta} \left\| n^{-1} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta > 0)) - M_0(\beta) \right\| \xrightarrow{p} 0,$$

for some  $M_0(\beta)$  and  $M_0(\beta)$  is continuous uniformly and positive semidefinite for all  $\beta \in \Theta$ .

**Assumption A6.**  $G(\beta) = \nabla_{\beta} E[\mathbf{x}_i(\tau - I(y_i < \mathbf{x}_i^T \beta))I(\mathbf{x}_i^T \beta > 0)]$  is continuous in  $\beta$ , and

$$-1/\sqrt{n}c_{\tau}(\beta_0)^T M_0(\beta_0)G(\beta_0) \xrightarrow{d} N(0, G(\beta_0)M_0(\beta_0)G(\beta_0)).$$

These are technical conditions to ensure the posterior concentration using score likelihood for censored quantile regression. Assumption A4 is regarding the identifiability of  $\beta_0$  and Assumptions A5–A6 ensure regularity of the score function. Assumption A6 also imposes a distribution assumption on  $c_{\tau}(\beta)$ , which is the same as the condition (iii) in Proposition 1 of Chernozhukov and Hong (2003).

The following theorem shows that our proposed likelihood asymptotically leads to valid frequentist inference for censored quantile regression.

**Theorem 2.** *a) (Asymptotic Normality of Posterior Distribution) Consider the proposed likelihood in (2.9) with the bounded uniform prior and the following weight matrix*

$$W(\beta) = \left( \tau(1 - \tau)n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T I(\mathbf{x}_i^T \beta > 0) \right)^{-1}.$$

*Under Assumptions 1, 2, and 4–6, for  $\beta$  such that  $\|\beta - \beta_0\| = O(n^{-1/2})$ , the posterior of  $\beta$  satisfies*

$$p_n(\beta \mid \mathcal{D}) \propto \exp \left\{ -\frac{1}{2}n(\beta - \beta_0)^T \Sigma_c^{-1}(\beta - \beta_0) + o_p(1) \right\},$$

*where  $\Sigma_c = \tau(1 - \tau)M_1^{-1}M_0M_1^{-1}$ .*

*b) (Posterior concentration) Assume the true parameter  $\beta_0$  belongs to the interior of a compact convex set  $\Theta$ , and the prior distribution restricted to  $\Theta$ . Under the additional Assumptions A4–A6, the posterior distribution of  $\beta$  satisfies*

$$\Pi(\|\beta - \beta_0\| > M_n n^{-1/2} \mid \mathcal{D}) \xrightarrow{p} 0,$$

*where  $M_n$  is any sequence diverging to infinity with  $n$ .*

For the censored case as well, the posterior distribution concentrates and is asymptotically normal, centered at the true regression parameter  $\beta_0$ , with the correct covariance matrix  $\Sigma_c = \tau(1 - \tau)M_1^{-1}M_0M_1^{-1}$ . The asymptotic normal expansion of the posterior distribution is same as the asymptotic distribution of the Powell's estimator. While both approaches result in similar behavior asymptotically, the score likelihood based approach is better suited for computations since the Powell's objective function is highly non-convex (Womersley, 1986; Chernozhukov and Hong, 2003). Note that Chernozhukov and Hong (2003)'s approach is different from ours as they used the Powell's objective function in place of the score function in the likelihood we proposed. In the next section, we will discuss our proposed computational strategies.

### 3 Computational Methods

We propose an efficient computational method for computing the posterior corresponding to the proposed likelihood function. Since the posterior distribution is asymptotically normal, the mean and the variance of the posterior provide a good description of the distribution. We propose an adaptive importance sampling algorithm for computing the posterior mean and the variance and the same approach can be used if higher moments of the posterior distribution need to be obtained. To ensure efficiency in terms of a good effective sample size, we propose an adaptive approach to update the proposal distribution of the importance sampling algorithm in an iterative manner. Importance sampling has been widely used as an alternative to Markov Chain Monte Carlo and the performance of the importance sampling algorithm relies on a good proposal distribution (Robert and Casella, 2013). The essential idea of adaptive importance sampling is to update the proposal distribution based on an iterative process to gain efficiency (Bugallo et al., 2017; Cornuet et al., 2012). Adaptive importance sampling methods are often used for computation in the contexts of mixture models (Cappé et al., 2008) and missing data models (Celeux et al., 2006).

#### 3.1 Adaptive Importance Sampling Algorithm (Ada-IS) for a Single Quantile Level

We now describe our algorithm to compute the posterior mean and the covariance matrix. As the posterior distribution is asymptotically normal, it is reasonable to use a normal distribution as the proposal distribution. We use the results from a linear regression (or median regression) as an initialization, and use importance sampling to update the mean and the covariance matrix of the proposal distribution. The resultant estimators are expected to be better than the initialization and will be used as the new proposal to improve the efficiency of the algorithm. Updating the proposal just a few times turns out to be helpful in making the final proposal closer to the posterior distribution. The details of the algorithm are stated as follows.

**Ada-IS Algorithm for a single quantile level:**

- (1) Fit a linear mean (or median) regression model, and use  $\mathbf{a}_1$  to denote the fitted slope parameter. Let  $a_0(\tau)$  be the  $\tau$ th sample quantile of  $y_i - \mathbf{x}_i^T \mathbf{a}_1$ ,  $i = 1, 2, \dots, n$  and  $\mathbf{a} = (a_0(\tau), \mathbf{a}_1)$ . Use  $q(\mathbf{b}) = N(\mathbf{a}, \hat{\Sigma})$  as the initial proposal, where

$$\hat{\Sigma} = c\tau(1 - \tau) \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T / n \right)^{-1},$$

and  $c$  is a constant to be selected.

- (2) Generate samples  $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \dots, \mathbf{b}^{(M)}$  from  $q(\mathbf{b})$  for some large  $M$  and calculate the importance weights  $w^{(r)} = L(\mathbf{b}^{(r)})\pi(\mathbf{b}^{(r)})/q(\mathbf{b}^{(r)})$ , where  $\pi$  is the prior distribution for  $\beta$ . Effective sample size (ESS) is calculated based on

$$ESS = \frac{M}{1 + cv^2}, \text{ where } cv^2 = \frac{(M-1)^{-1} \sum_{r=1}^M (w^{(r)} - \bar{w})^2}{\bar{w}^2}, \text{ and } \bar{w} = \frac{1}{n} \sum_{r=1}^M w^{(r)}.$$

- (3) The posterior mean can be estimated as

$$\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_p) = \frac{\sum_{i=1}^M w^{(r)} \mathbf{b}^{(r)}}{\sum_{i=1}^M w^{(r)}},$$

and the  $j^{th}$  entry of the posterior covariance matrix can be estimated by

$$\hat{\sigma}_{jk} = \frac{\sum_{r=1}^M w^{(r)} \mathbf{b}_j^{(r)} \mathbf{b}_k^{(r)}}{\sum_{r=1}^M w^{(r)}} - \hat{\mu}_j \hat{\mu}_k,$$

where  $\mathbf{b}_j^{(r)}$  is the  $j^{th}$  coordinate of  $\mathbf{b}^{(r)}$ .

- (4) Use the mean and the covariance matrix from Step 3 to form a new normal proposal distribution and repeat Steps 2–4 for a pre-specified number of times.
- (5) Estimate the posterior mean and the covariance matrix by using importance sampling from the final proposal distribution obtained from Step 4.

In Step 1, the motivation for the initialization for the covariance matrix is that when the densities  $f(\mathbf{x}_i | \boldsymbol{\beta}(\tau))$  are approximately equal to the constant  $k$ , then the true covariance matrix would take the following form

$$\tau(1 - \tau) D_1(\tau)^{-1} D_0 D_1(\tau)^{-1} \approx \frac{1}{k^2} \tau(1 - \tau) \left( \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T / n \right)^{-1}.$$

This is equivalent to find a proper constant  $c$ . A larger value of  $c$  will inflate the variance and make the tails of the proposal distribution heavy, while a smaller  $c$  will make the tail lighter. In our empirical studies, we used a common value of  $c = 1$  and did not optimize it. However, it is possible to try a few different values on a pilot run and choose the one providing the largest effective sample size in practice. It is also possible to use the estimates from **quantreg** as initialization, which is expected to perform better than the least squares initialization. However, even using the least squares initialization, in most cases, the proposed method performs well in terms of obtaining a good effective sample size. For Step 4, it is also possible to make the algorithm adaptive by continuing to update the proposal until a desired amount of effective sample size is reached. In our implementation, we will update the proposal for a pre-specified number of times rather than aiming for a target effective sample size.

For small sample sizes, the posterior distribution could be heavy tailed, causing the posterior variance to be potentially over-estimated. To improve the finite sample performance, we can estimate the posterior variance based on the posterior interquartile range, which is more robust. Simulation examples show that the finite sample performance of the estimation based on the posterior interquartile range is slightly better compared to the one based on the posterior variance. The importance sampling algorithm can be used to estimate the posterior quantiles and the interquartile range. We now present the algorithm to estimate the standard deviation  $\sigma_1$  of the first slope coefficient  $\beta_1$  after obtaining samples from the proposal.

- Sort the samples corresponding to the parameter  $\beta_1$  obtained from the importance sampling algorithm  $b_1^{(1)}, b_1^{(2)}, \dots, b_1^{(M)}$  in increasing order. For simplicity, denote the resultant values by  $b_{(1)} < b_{(2)} < \dots < b_{(M)}$  and the corresponding importance weights by  $w_{(1)}, \dots, w_{(M)}$ .
- Let  $k_1 = \inf\{k \mid \sum_{i=1}^k w_{(i)} \geq 0.25\}$  and  $k_2 = \inf\{k \mid \sum_{i=1}^k w_{(i)} \geq 0.75\}$ . Then, an estimator of  $\sigma_1$  by matching the estimated interquartile range (IQR) with the asymptotic interquartile range is given by

$$\hat{\sigma}_1 = (b_{(k_2)} - b_{(k_1)})/1.349,$$

where the factor 1.349 is the interquartile range of the standard normal distribution.

As suggested by a reviewer, another possibility to improve the effective sample size of the importance sampling algorithm in finite samples is to utilize a heavy tailed distribution such as the  $t$  distribution as proposal. We have observed that this strategy also works equally well in practice and difference between these two strategies is not substantial as the Ada-IS algorithm updates the proposal distribution along the path.

### 3.2 Ada-IS Algorithm for Multiple Quantile Levels

For multiple quantile regression, it is more challenging to devise a proposal distribution which will yield good effective sample size. The difficulty is that it is not straightforward to obtain a good initialization for the joint covariance matrix of the multiple regression quantiles. To address this, we first obtain good estimators of the mean and the covariance matrix for each of the individual quantile levels using the Ada-IS algorithm for a single quantile level. This will give a good estimation for the block diagonal part for the joint covariance matrix. We propose a way to impute the off-diagonal blocks, which correspond to the covariances at different quantile levels, based on the diagonal blocks as described in the following algorithm.

#### Ada-IS Algorithm for multiple quantiles:

- (1) For each individual quantile level in  $\{\tau_1, \tau_2, \dots, \tau_m\}$ , implement the Ada-IS Algorithm for a single quantile level separately. Use  $(\hat{\mu}_1, \dots, \hat{\mu}_m)$  to denote the estimated means, and  $(\hat{\Sigma}_1, \dots, \hat{\Sigma}_m)$  to denote the estimated covariance matrices.
- (2) Use an  $mp$ -dimensional multivariate normal proposal, with  $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_m)$  as the initialization for the mean and  $(\hat{\Sigma}_1, \dots, \hat{\Sigma}_m)$  as the block diagonal part for the joint covariance matrix. For the covariance matrix between the quantile levels  $\tau_i, \tau_j$ , we use

$$\hat{\Sigma}_{ij} = \gamma(\tau_i \wedge \tau_j - \tau_i \tau_j) \left( (\tau_i(1 - \tau_i)\hat{\Sigma}_i^{-1} + \tau_j(1 - \tau_j)\hat{\Sigma}_j^{-1})/2 \right)^{-1}, \quad (3.1)$$

where  $\gamma \leq 1$  is a value close to 1 chosen to make the covariance matrix  $\Sigma$  positive definite.

- (3) Generate samples from the proposal distribution and estimate the mean and covariance matrix as described in the Ada-IS Algorithm for a single quantile. Use these estimates to form a new normal proposal distribution as described in as described in Step 3 of the Ada-IS Algorithm for a single quantile.
- (4) Update the proposal for a pre-specified number of times. Estimate the posterior mean and the covariance matrix using importance samples from the final proposal distribution from Step 3.

We will now provide a heuristic intuition behind the imputation of the off-diagonal blocks of the covariance matrix in Step 2. Ideally, we would like to have

$$\hat{\Sigma}_{ij} \approx (\tau_i \wedge \tau_j - \tau_i \tau_j) D_1(\tau_i)^{-1} D_0 D_1(\tau_j)^{-1}.$$

We aim to approximate  $D_1(\tau_i)^{-1} D_0 D_1(\tau_j)^{-1}$  based on the approximation of the diagonal blocks:

$$\hat{\Sigma}_i \approx \tau_i(1 - \tau_i) D_1(\tau_i)^{-1} D_0 D_1(\tau_i)^{-1} \text{ and } \hat{\Sigma}_j \approx \tau_j(1 - \tau_j) D_1(\tau_j)^{-1} D_0 D_1(\tau_j)^{-1}.$$

If two quantile levels are close, we would assume that  $\|D_1(\tau_i) - D_1(\tau_j)\|$  is small. Then

$$\begin{aligned} & \|D_1(\tau_i) D_0^{-1} D_1(\tau_i) + D_1(\tau_j) D_0^{-1} D_1(\tau_j) - 2D_1(\tau_j) D_0^{-1} D_1(\tau_i)\| \\ &= \|(D_1(\tau_i) - D_1(\tau_j)) D_0^{-1} (D_1(\tau_i) - D_1(\tau_j))\| \approx 0. \end{aligned}$$

Therefore, we can obtain the following approximation,

$$\begin{aligned} \frac{1}{2} \left( \tau_i(1 - \tau_i) \hat{\Sigma}_i^{-1} + \tau_j(1 - \tau_j) \hat{\Sigma}_j^{-1} \right) &\approx (D_1(\tau_i) D_0^{-1} D_1(\tau_i) + D_1(\tau_j) D_0^{-1} D_1(\tau_j)) / 2 \\ &\approx D_1(\tau_j) D_0^{-1} D_1(\tau_i). \end{aligned}$$

The proposed imputation for  $\Sigma_{ij}$  given by (3.1) follows from this approximation. Notice that for modeling multiple quantile levels, if a sample from the normal proposal does not satisfy the monotonicity constraint, its weight will be set to zero. This will make sure that the conditional quantiles are monotone. However, this may cause the algorithm to be less efficient, especially when  $m \times p$ , the number of coefficients to be estimated, is large and the quantile levels are very close to each other. This constraint can be easily removed from the algorithm if higher efficiency is desired.

## 4 Simulation Study

In this section, we use simulated datasets to study the performance of our proposed approach. The following model is used for generating the data. For  $i = 1, \dots, n$ :

$$y_i = 5 + 2x_i + \epsilon_i, \quad \epsilon_i \sim N(0, (1 + 0.5x_i)^2), \quad (4.1)$$

where  $\epsilon_i$ 's are independent and the  $x_i$ 's are  $n$  grid points equally spaced between 0 and 20. The conditions assumed for the theoretical results are satisfied by this data generating process used for our simulation studies. We choose  $n = 2000$  to evaluate the asymptotic behavior while  $n = 100$  and  $200$  are used to evaluate the small sample size performance. Each simulation study uses 1000 Monte Carlo replications to aggregate the results.



## 4.1 Single Quantile Regression

We consider the median quantile regression with  $\tau = 0.5$ . We consider the following methods for comparing their performance in terms of parameter estimation and computational efficiency.

- (1) **Importance sampling algorithm (IS):** Use the initialization described in Step (1) of Ada-IS Algorithm for single quantile with  $c = 1$ . In principle, one can choose a better  $c$  by searching over a small grid of values but we use this as a default. Implement an importance sampling algorithm without updating the proposal.
- (2) **Adaptive importance sampling algorithm (Ada-IS):** Implement the Ada-IS algorithm for single quantile. Start with the same initialization in (IS) and update the proposal three times with 2000 samples generated each time to get an informative proposal.
- (3) **Metropolis-Hastings algorithm (MH):** This is an Metropolis-Hastings algorithm as proposed by Chernozhukov and Hong (2003) and with a good proposal distribution already obtained from the Ada-IS algorithm. That is, use a multivariate normal distribution with mean and covariance matrix the same as the informative proposal in (2) as the proposal distribution.
- (4) **Asymmetric Laplace Likelihood (ALD):** This one uses the asymmetric Laplace likelihood and implements a Metropolis-Hastings Algorithm using the same proposal distribution as in (3). For estimating the variance, the correction proposed by Yang et al. (2016) is used.

We first consider the case with  $n = 2000$  observations generated from the above model. For each of the algorithms, 10000 final samples are obtained to estimate the parameters and the variances. In Table 1, we report the mean squared error (MSE), bias, and variance of the estimators for different methods mentioned above. The bias is calculated based on the average of the 1000 Monte Carlo replications. The average of the estimated variances  $\hat{\sigma}$  is compared with the true asymptotic variance, which is given by  $\tau(1 - \tau)D_1^{-1}D_0D_1^{-1}$ . We also report the coverage of the 90% credible/confidence intervals, the effective sample size for all the methods. Quantile regression estimators using the R package **quantreg** are also provided here for comparison.

The simulation results in Table 1 demonstrate the asymptotic validity of the score based likelihood and efficiency of the proposed algorithm. The proposed score based likelihood provides good point estimation and has better mean squared error compared with the other methods. However, the estimated variances tend to be slightly larger than the asymptotic variance of the estimator, which also causes the coverage to be slightly larger than the nominal level. In terms of the effective sample size, the proposed adaptive importance sampling algorithm has a clearly superior performance with an effective sample size of 9485 which substantially improves upon the one-step importance sampling algorithm which has an effective sample size of 1089. Moreover, the Ada-IS algorithm for the score likelihood has much better efficiency even compared to the commonly used Metropolis-Hastings algorithm with the ALD likelihood. Also, since importance sampling is a non-stochastic algorithm, it does not suffer from the potential convergence

	Quantreg		IS		MH		Ada-IS		ALD	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE( $10^{-2}$ )	2.059	0.067	1.901	0.063	1.900	0.062	1.895	0.062	1.970	0.065
Bias( $10^2$ )	-0.217	0.057	-0.032	0.053	-0.041	0.054	-0.044	0.055	-0.113	0.061
Var( $\hat{\beta}$ )( $10^{-2}$ )	2.059	0.067	1.901	0.063	1.900	0.062	1.895	0.062	1.970	0.065
$n \times \hat{\sigma}^2$	41.400	1.387	43.317	1.424	43.237	1.421	43.235	1.422	41.934	1.392
Asym Var	40.731	1.368	40.731	1.368	40.731	1.368	40.731	1.368	40.731	1.368
Coverage	0.902	0.901	0.921	0.915	0.922	0.914	0.922	0.919	0.910	0.892
ESS	—	—	1089	1089	5447	5447	9485	9485	3376	3376

<sup>1</sup> Quantreg corresponds to the quantile regression estimator using R package **quantreg**.

<sup>2</sup>  $\hat{\sigma}$  is average of the estimated variance. For the Bayesian approach, this is the posterior variance. For Quantreg, this is computed based on a Huber's sandwich estimator.

<sup>3</sup> Var( $\hat{\beta}$ ) is the variance of the estimator and is calculated by the sample variance of the estimators in 1000 Monte Carlo replications.

<sup>4</sup> Asym Var is the true asymptotic variance calculated based on the underlying model.

<sup>5</sup> Coverage is the empirical coverage for 90% confidence/credible interval.

Table 1: Single quantile regression ( $\tau = 0.5$ ).

issues associated with the Metropolis-Hastings algorithm. In the later simulation studies, we use the adaptive importance sampling algorithm as the default algorithm for the proposed score likelihood.

We investigate the finite sample performance of the proposed algorithm for smaller sample sizes  $n = 100$  and  $n = 200$ . As mentioned earlier, the posterior distribution could be heavy tailed for small sample sizes so we use the interquartile range approach discussed in Section 3.1 to estimate the posterior variance. Table 2 compares the finite sample performance of our approach with the frequentist approach and the corrected ALD likelihood approach. The performance of the estimated variances and the corresponding coverage probabilities are reported. Notice that for small sample sizes, our score likelihood based approach tends to have coverage probabilities larger than the nominal level without the finite sample correction while **quantreg** and the corrected ALD likelihood tend to have under-coverage. With the finite sample correction, the score likelihood provides better empirical coverage compared to the ALD likelihood. We provided further simulation results in the supplementary material (Wu and Narisetty, 2020) based on heavy tailed  $t$  distributions as proposals which also perform well for small sample sizes (see the results presented in Table A1 in the supplementary material).

## 4.2 Multiple Quantile Regression

To evaluate the performance of our method for multiple quantile regression, we choose five quantile levels which are relatively close to each other,  $\tau_1 = 0.4$ ,  $\tau_2 = 0.45$ ,  $\tau_3 = 0.5$ ,  $\tau_4 = 0.55$  and  $\tau_5 = 0.6$ . This will help us evaluate if the imposed monotonicity constraint will result in higher efficiency. We use the same model given by (4.1) to

n = 100	Quantreg		Ada-IS		ALD	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE	0.404	0.014	0.384	0.012	0.391	0.013
Bias	0.002	-0.002	-0.005	0.001	0.009	-0.002
$Var(\hat{\beta})$	0.404	0.013	0.384	0.012	0.391	0.013
$n \times \hat{\sigma}^2$	43.396	1.438	58.070	1.695	43.678	1.512
$n \times \hat{\sigma}_{FS}^2$	-	-	43.383	1.437	-	-
Asym Var	40.731	1.368	40.731	1.368	40.731	1.368
Coverage (90%)	0.859	0.897	0.944	0.926	0.829	0.840
Coverage- $\hat{\sigma}_{FS}$ (90%)	-	-	0.869	0.906	-	-
n = 200	Quantreg		Ada-IS		ALD	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE	0.196	0.007	0.181	0.006	0.185	0.006
Bias	-0.016	0.004	-0.014	0.003	-0.009	0.003
$Var(\hat{\beta})$	0.196	0.007	0.181	0.006	0.184	0.006
$n \times \hat{\sigma}^2$	41.541	1.409	50.317	1.563	43.759	1.533
$n \times \hat{\sigma}_{FS}^2$	-	-	41.439	1.403	-	-
Asym Var	40.731	1.368	40.731	1.368	40.731	1.368
Coverage 90%	0.863	0.902	0.930	0.916	0.873	0.869
Coverage- $\hat{\sigma}_{FS}$ (90%)	-	-	0.873	0.914	-	-

<sup>1</sup>  $\hat{\sigma}_{FS}^2$  is the average posterior variance computed based on the IQR approach described in Section 3.1.

<sup>2</sup> Coverage- $\hat{\sigma}_{FS}^2$  (90%) is the empirical coverage probability based on  $\hat{\sigma}_{FS}^2$ .

Table 2: Comparison of finite sample performance.

generate  $n = 2000$  samples and implement our Ada-IS algorithm for multiple quantiles. We use  $\gamma = 0.9$  when imputing the off-diagonal blocks for the joint covariance matrix as a default but a better choice can be made in practice by choosing from a grid of values. Results from **quantreg** obtained separately for different quantile levels and Ada-IS without monotonicity constraints are also presented.

The simulation results for the multiple quantiles are shown in Table 3. For estimating the conditional median, the average posterior variance based on the joint estimation is smaller than the one from Table 1, which is based on the single quantile level. Also, the MSE's obtained using the default Ada-IS algorithm are smaller than those without imposing monotonicity constraints. These provide empirical evidence that estimating multiple quantile levels jointly is better owing to the monotonicity constraints, at least when the model is correctly specified. On the other hand, the effective sample size is 1745 out of 10000 samples from the final proposal. The effective sample size loss compared with separate estimation is expected due to the increased dimension and the monotonicity constraints. However, one should not be discouraged, since when estimating them separately we still need 5 times the computational time. Moreover, by jointly modeling multiple quantile levels, the joint covariance matrix at different quantile levels is obtained which is useful for making inference about parameters such as the slope differences at different quantile levels as demonstrated in Section 4.4.

Quantreg	0.4		0.45		0.5		0.55		0.6	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE( $10^{-2}$ )	2.150	0.067	2.076	0.065	2.059	0.067	2.025	0.069	2.113	0.069
Bias( $10^{-2}$ )	0.211	0.036	-0.314	0.069	-0.217	0.057	0.095	0.051	0.359	0.043
Var( $\hat{\beta}$ )( $10^{-2}$ )	2.149	0.067	2.075	0.065	2.059	0.067	2.025	0.069	2.112	0.069
$n \times \hat{\sigma}^2$	41.530	1.407	41.236	1.389	41.400	1.387	41.848	1.396	42.248	1.415
Asym Var	41.693	1.401	40.965	1.376	40.731	1.368	40.965	1.376	41.693	1.401
Coverage	0.899	0.907	0.891	0.904	0.902	0.901	0.907	0.900	0.901	0.900
Ada-IS	0.4		0.45		0.5		0.55		0.6	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE( $10^{-2}$ )	2.073	0.065	1.971	0.064	1.894	0.064	1.953	0.067	2.071	0.068
Bias( $10^{-2}$ )	-1.540	0.237	-0.797	0.130	0.068	0.043	1.093	-0.069	2.209	-0.190
Var( $\hat{\beta}$ )( $10^{-2}$ )	2.049	0.064	1.965	0.063	1.894	0.064	1.941	0.067	2.022	0.067
$n \times \hat{\sigma}^2$	43.142	1.422	41.993	1.404	41.841	1.373	42.366	1.391	43.756	1.431
Asym Var	41.693	1.401	40.965	1.376	40.731	1.368	40.965	1.376	41.693	1.401
Coverage	0.910	0.898	0.914	0.908	0.908	0.902	0.904	0.899	0.899	0.903
Ada-IS No Const	0.4		0.45		0.5		0.55		0.6	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE( $10^{-2}$ )	2.094	0.067	2.025	0.065	1.943	0.065	1.988	0.069	2.077	0.069
Bias( $10^{-2}$ )	-0.086	0.023	-0.153	0.032	0.060	0.039	0.510	0.009	0.921	-0.010
Var( $\hat{\beta}$ )( $10^{-2}$ )	2.094	0.067	2.024	0.065	1.943	0.065	1.985	0.069	2.068	0.069
$n \times \hat{\sigma}^2$	44.243	1.451	43.344	1.440	43.205	1.403	43.837	1.430	44.970	1.462
Asym Var	41.693	1.401	40.965	1.376	40.731	1.368	40.965	1.376	41.693	1.401
Coverage	0.904	0.896	0.912	0.907	0.915	0.897	0.908	0.895	0.904	0.899

Table 3: Multiple quantile regression ( $\tau = 0.4, 0.45, 0.5, 0.55, 0.6$ ).

C = 10	Powell		Ada-IS	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE	0.099	0.001	0.092	0.001
Bias	-0.017	0.002	-0.040	0.004
$\text{Var}(\hat{\beta})$	0.098	0.001	0.091	0.001
$n \times \hat{\sigma}^2$	147.474	2.471	208.812	2.935
Asym Var	190.578	2.766	190.578	2.766
Coverage	0.839	0.876	0.919	0.917
C = 15	Powell		Ada-IS	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
MSE	0.329	0.003	0.309	0.003
Bias	-0.026	0.003	-0.089	0.007
$\text{Var}(\hat{\beta})$	0.329	0.003	0.301	0.003
$n \times \hat{\sigma}^2$	443.224	4.596	685.547	5.944
Asym Var	617.867	5.522	617.867	5.522
Coverage	0.824	0.854	0.920	0.908

Table 4: Censored quantile regression ( $\tau = 0.5$ ).

### 4.3 Simulation Results for Censored Quantile Regression

We evaluate the performance of the score based likelihood for censored quantile regression. Consider median regression with a fixed left censoring for the model (4.1). We consider a sample size of  $n = 2000$  and censoring times at  $C = 10$  and  $C = 15$  corresponding to censoring rates around 15% and 25%. We implement the Ada-IS algorithm for censored quantile regression and compare the results with the Powell's estimator implemented using **quantreg**. The results are summarized in Table 4. We can see that the proposed likelihood provides good point estimation with a smaller mean squared error compared with the Powell's estimator. Also, the coverage probabilities for the Powell's estimator are much lower the nominal 90% level, while the coverage probability using the score based likelihood is much closer to 90%. This suggests that the proposed method using the score likelihood has a clear advantage over the Powell's approach for censored quantile regression.

### 4.4 Inference for Slope Differences

One advantage of the proposed likelihood is the ability to provide inference for functions of quantile regression coefficients corresponding to multiple quantile levels such as the slope difference  $\theta = \beta_1(0.75) - \beta_1(0.25)$ , which is the difference of the slope at the 75% and 25% levels. Existing methods based on the ALD likelihood do not provide valid inference on  $\theta$  directly. Yang et al. (2016) requires two separate models to be fitted to estimate  $D_1(0.25)$  and  $D_1(0.75)$  which may potentially be used to obtain corrected inference for  $\theta$  but this does not allow placing a prior distribution directly on  $\theta$ . Yang and Tokdar (2017)'s method may be used for inference on  $\theta$  but it requires modeling of all the quantile levels globally which is computationally intensive and also does not guarantee the frequentist validity of the resultant posterior intervals.

On the other hand, the score based likelihood we propose provides the flexibility to place prior distributions on  $\theta$  directly. This allows both the options of placing a bounded uniform prior if we do not have any prior information or an informative prior. In our simulation study, along with the uniform prior we consider two informative priors as follows:  $\theta \sim N(\mu, \sigma^2)$  where  $\mu = 0, \sigma^2 = 1$  and  $\mu = 1, \sigma^2 = 1$ . Denote the prior density of  $\theta$  by  $\pi(\theta)$  and uniform priors on the other parameters  $\beta_1(0.25), \beta_0(0.75), \beta_0(0.25)$ . Then the likelihood should be modified with a Jacobian transformation.

$$f(\mathbf{X} \mid \theta, \beta_0(0.75), \beta_1(0.25), \beta_0(0.25)) \propto L(\theta + \beta_1(0.25), \beta_0(0.75), \beta_1(0.25), \beta_0(0.25)).$$

The posterior distribution of the parameters is given by

$$f(\theta, \beta_0(0.75), \beta_1(0.25), \beta_0(0.25) \mid \mathbf{X}) \propto L(\theta + \beta_1(0.25), \beta_0(0.75), \beta_1(0.25), \beta_0(0.25))\pi(\theta).$$

For computation, we can still use the Ada-IS algorithm but we only need to focus on the marginal distribution of  $\theta$  for performing inference on  $\theta$ .

Table 5 presents the simulation results for our approach using three different prior choices along with the results for the standard quantile regression using **quantreg**. Although practically it is not quite possible to have informative prior such as  $N(1, 1)$ , we want to include it in the comparison to see how an informative prior would improve the performance. From the simulation results, we can see that all the methods provide reasonably good estimation while the three score based methods have smaller mean squared error and smaller variance compared with **quantreg**. We see slightly smaller mean squared error for the  $N(1, 1)$  informative prior but the difference is insignificant when compared with other prior choices. Our approach provides smaller mean squared error and larger coverage.

	Quantreg	Score-Uniform	Score-N(0,1)	Score-N(1,1)
True $\theta$	0.6745	0.6745	0.6745	0.6745
Estimation	0.6735	0.6733	0.6725	0.6737
Bias( $10^{-2}$ )	-0.1000	-0.1180	-0.1957	-0.0804
Variance( $10^{-2}$ )	0.1098	0.0988	0.0985	0.0986
MSE( $10^{-2}$ )	0.1099	0.0989	0.0989	0.0986
Coverage (90%)	0.8880	0.9140	0.9140	0.9160

Table 5: Estimation of  $\theta = \beta_1(0.75) - \beta_1(0.25)$ : The proposed Bayesian score approach is performed using three different prior distributions on  $\theta$ : Uniform,  $N(0, 1)$ , and  $N(1, 1)$ .

## 5 Application to Immunoglobulin-G Data

We now apply our methods to an immunoglobulin-G data set analyzed in Yu and Moyeed (2001). 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. Following Yu and Moyeed (2001), we use IgG concentration as the response variable  $Y$  and consider a quadratic quantile regression model with independent variable  $x$  to be the age.

$$Q_Y(\tau \mid X = x) = \beta_0(\tau) + \beta_1(\tau)x + \beta_2(\tau)x^2.$$

		Quantreg		Score-Uniform		Score-MVN	
		Estimation	Variance	Estimation	Variance	Estimation	Variance
$\tau = 0.25$	$\beta_0$	1.468	0.191	1.566	0.158	1.521	0.147
	$\beta_1$	1.335	0.167	1.334	0.120	1.355	0.110
	$\beta_2$	-0.137	0.005	-0.139	0.004	-0.141	0.004
$\tau = 0.5$	$\beta_0$	2.801	0.274	3.026	0.304	2.879	0.251
	$\beta_1$	1.159	0.193	0.997	0.223	1.129	0.173
	$\beta_2$	-0.075	0.006	-0.047	0.007	-0.068	0.005
$\tau = 0.75$	$\beta_0$	4.342	0.222	4.507	0.323	4.004	0.231
	$\beta_1$	0.705	0.167	0.585	0.255	1.007	0.207
	$\beta_2$	0.019	0.005	0.042	0.007	-0.024	0.006

Table 6: Quantile regression inference for the IgG data: The proposed Bayesian score approach is performed using two different prior distributions on  $\theta_i$ : Uniform, Multivariate Normal(MVN).

We estimate the regression parameters for  $\tau = 0.25, 0.5, 0.75$  jointly using bounded uniform prior and a multivariate normal prior with zero mean and covariance matrix generated by the squared exponential kernel in expression (2.4), with  $\sigma^2 = 2$  and  $b = 0.5$ . We also compare the results with Quantreg which are summarized in Tables 6 and 7.

The results in Table 6 suggest that while the point estimators based on both the approaches are quite similar, their variances are different. Our method (Ada-IS) indicates a smaller variance for the quantile level  $\tau = 0.25$  but a larger variance for the other quantile levels compared with **quantreg**. For the case with multivariate normal prior, the corresponding parameter for different quantile levels shrink towards similar values and has smaller variances compared to the uniform prior. The quantile curves corresponding to the score likelihood approach using the multivariate normal prior and those from **quantreg** are plotted in Figure 2 which shows that they are quite similar to each other.

We also use our proposed score likelihood to test the following hypotheses:

$$H_0 : \theta_i = \beta_i(0.75) - \beta_i(0.25) = 0 \text{ vs } H_1 : \theta_i \neq 0,$$

for  $i = 1, 2$ . We try a bounded uniform prior and a  $N(0, 1)$  prior for both  $\theta_1$  and  $\theta_2$ . The estimated parameters, corresponding standard errors and interval estimation for the linear and quadratic terms are shown in Table 7. We report the 90% credible intervals based on the computed posterior distributions for our approach and the 90% confidence intervals for Quantreg. All the interval estimations for  $\theta_1$  contain zero while neither of the intervals for  $\theta_2$  contain zero, which indicates that there is a significant difference in the quadratic effect at different quantiles and no significant difference in the linear effect. This is a simple demonstration that the proposed likelihood can be used in a natural way to carry out inference for quantities such as the slope differences in a Bayesian framework while incorporating prior information.

	Quantreg			Score-Uniform			Score- $N(0, 1)$		
	Est	SE	Cred.Int	Est	SE	Cred.Int	Est	SE	Cred.Int
$\theta_1$	-0.630	0.472	(-1.407, 0.147)	-0.717	0.524	(-1.580, 0.145)	-0.557	0.464	(-1.321, 0.207)
$\theta_2$	0.156	0.084	(0.018, 0.294)	0.177	0.090	(0.029, 0.325)	0.150	0.080	(0.018, 0.281)

Table 7: IgG data – Inference for the slope differences  $\theta_1 := \beta_1(0.75) - \beta_1(0.25)$  (for the linear term) and  $\theta_2 := \beta_2(0.75) - \beta_2(0.25)$  (for the quadratic term). Est denotes the estimated value, SE denotes the standard error of the estimator, Cred.Int denotes an estimated credible interval.



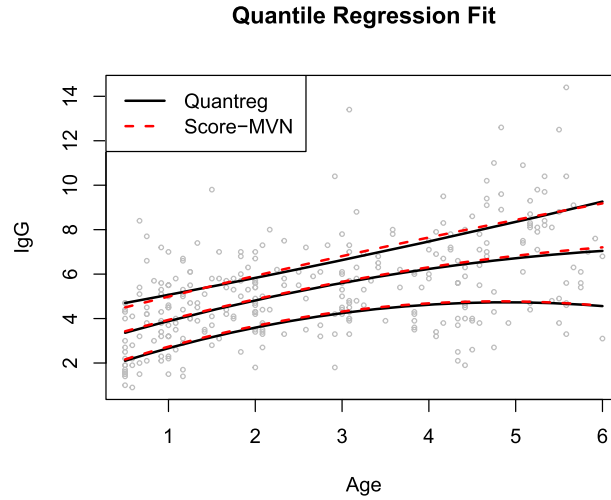


Figure 2: IgG data – Fitted regression curves using the proposed score likelihood with multivariate normal prior (dashed red) and Quantreg (solid black) for  $\tau = 0.25, 0.5, 0.75$ .

## 6 Conclusion

In this paper, we propose a score based likelihood for quantile regression problems which particularly targets modeling multiple quantile levels at the same time. We also propose Ada-IS algorithm for posterior computation which empirically shown to have good efficiency in terms of effective sample size. The score-based likelihood automatically and correctly estimates the variance of the regression quantiles and does not require estimation of the conditional densities as needed by frequentist methods, which is an extremely challenging task especially when the data are not independent and identically distributed. While the ALD likelihood also avoids estimation of the conditional densities, the posterior inference provided based on the score likelihood is valid and does not need further adjustment. The score likelihood can naturally perform inference at multiple quantile levels and deal with the issue of quantile crossing more naturally compared to frequentist methods, which often resort to adhoc strategies (He, 1997; Liu and Wu, 2009; Zou et al., 2008; Bondell et al., 2010). Performing valid Bayesian inference for functions of regression quantiles at multiple quantile levels such as  $\beta(\tau_1) - \beta(\tau_2)$  is possible based on the proposed score likelihood but it is not clear how the ALD based inference can be corrected in such situations. For the censored quantile regression case discussed in Section 2.5, the objective function is highly non-convex. The proposed likelihood provides a more efficient computational technique based on importance sampling to avoid direct optimization of the highly nonconvex objective function that is prone to be stuck in local solutions (Womersley, 1986; Chernozhukov and Hong, 2003).

For the score based likelihood, the contribution of each observation to the likelihood is not multiplicative which is the case for the ALD likelihood. However, the multiplicative property of the likelihood is not necessary for valid inference and commonly used

likelihoods such as empirical likelihood (Owen, 1988, 2001) are also not multiplicative. Yang et al. (2016) developed a Bayesian quantile regression approach based on empirical likelihood and showed valid posterior properties.

One potential limitation of the proposed score likelihood is that it is not as straightforward to evaluate it in an online manner as stream of data come in as compared to evaluating the ALD likelihood. However, it is still possible to devise computationally efficient strategies that keep track of the score function and the weight matrix from the previous batch to reduce the computational time needed for evaluation of the score likelihood in an online manner. Development of such computationally efficient algorithms that can be used to generalize the proposed importance sampling algorithms to online settings is an interesting direction for a future study.

It would be another interesting future research direction to investigate if the proposed techniques can be generalized to high dimensional settings so that problems such as Bayesian variable selection for quantile regression can be addressed using the score likelihood.

## Supplementary Material

Supplementary Material for “Bayesian multiple quantile regression for linear models using a score likelihood” (DOI: [10.1214/20-BA1217SUPP](https://doi.org/10.1214/20-BA1217SUPP); .pdf). The online supplement contains the technical proofs for Theorems 1 and 2, and additional simulation results.

## References

- Bondell, H. D., Reich, B. J., and Wang, H. (2010). “Noncrossing quantile regression curve estimation.” *Biometrika*, 97(4): 825–838. MR2746154. doi: <https://doi.org/10.1093/biomet/asq048>. 876, 899
- Bugallo, M. F., Elvira, V., Martino, L., Luengo, D., Miguez, J., and Djuric, P. M. (2017). “Adaptive importance sampling: the past, the present, and the future.” *IEEE Signal Processing Magazine*, 34(4): 60–79. MR3938972. doi: <https://doi.org/10.1214/18-ST668>. 887
- Cappé, O., Douc, R., Guillin, A., Marin, J.-M., and Robert, C. P. (2008). “Adaptive importance sampling in general mixture classes.” *Statistics and Computing*, 18(4): 447–459. MR2461888. doi: <https://doi.org/10.1007/s11222-008-9059-x>. 887
- Celeux, G., Marin, J.-M., and Robert, C. P. (2006). “Iterated importance sampling in missing data problems.” *Computational statistics & data analysis*, 50(12): 3386–3404. MR2236856. doi: <https://doi.org/10.1016/j.csda.2005.07.018>. 887
- Chernozhukov, V. and Hong, H. (2003). “An MCMC approach to classical estimation.” *Journal of Econometrics*, 115(2): 293–346. MR1984779. doi: [https://doi.org/10.1016/S0304-4076\(03\)00100-3](https://doi.org/10.1016/S0304-4076(03)00100-3). 878, 879, 883, 884, 885, 886, 891, 899
- Cornuet, J.-M., Marin, J.-M., Mira, A., Robert, C. P., et al. (2012). “Adaptive Mul-

- tuple Importance Sampling.” *Scandinavian Journal of Statistics*, 39(4): 798–812. MR3000850. doi: <https://doi.org/10.1111/j.1467-9469.2011.00756.x>. 887
- Dunson, D. B. and Taylor, J. A. (2005). “Approximate Bayesian inference for quantiles.” *Journal of Nonparametric Statistics*, 17(3): 385–400. MR2129840. doi: <https://doi.org/10.1080/10485250500039049>. 877
- Feng, Y., Chen, Y., and He, X. (2015). “Bayesian quantile regression with approximate likelihood.” *Bernoulli*, 21(2): 832–850. MR3338648. doi: <https://doi.org/10.3150/13-BEJ589>. 877
- George, E. I. and McCulloch, R. E. (1993). “Variable selection via Gibbs sampling.” *Journal of the American Statistical Association*, 88: 881–889. 881
- He, X. (1997). “Quantile curves without crossing.” *The American Statistician*, 51(2): 186–192. 876, 899
- He, X., Shao, Q.-M., et al. (1996). “A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs.” *The Annals of Statistics*, 24(6): 2608–2630. MR1425971. doi: <https://doi.org/10.1214/aos/1032181172>. 883
- Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press. MR2268657. doi: <https://doi.org/10.1017/CB09780511754098>. 875, 876, 884
- Koenker, R. and Bassett, G. (1978). “Regression Quantiles.” *Econometrica*, 46(1): 33–50. MR0474644. doi: <https://doi.org/10.2307/1913643>. 875, 876, 884
- Kozumi, H. and Kobayashi, G. (2011). “Gibbs sampling methods for Bayesian quantile regression.” *Journal of Statistical Computation and Simulation*, 81(11): 1565–1578. MR2851270. doi: <https://doi.org/10.1080/00949655.2010.496117>. 876
- Lancaster, T. and Jae Jun, S. (2010). “Bayesian quantile regression methods.” *Journal of Applied Econometrics*, 25(2): 287–307. MR2758636. doi: <https://doi.org/10.1002/jae.1069>. 877
- Lavine, M. (1995). “On an approximate likelihood for quantiles.” *Biometrika*, 82(1): 220–222. MR1332852. doi: <https://doi.org/10.1093/biomet/82.1.220>. 877
- Li, C. and Jiang, W. (2016). “On oracle property and asymptotic validity of Bayesian generalized method of moments.” *Journal of Multivariate Analysis*, 145: 132–147. MR3459943. doi: <https://doi.org/10.1016/j.jmva.2015.12.009>. 879
- Liu, Y. and Wu, Y. (2009). “Stepwise multiple quantile regression estimation using non-crossing constraints.” *Statistics and its Interface*, 2(3): 299–310. MR2540088. doi: <https://doi.org/10.4310/SII.2009.v2.n3.a4>. 876, 899
- Narisetty, N. N. and He, X. (2014). “Bayesian variable selection with shrinking and diffusing priors.” *Annals of Statistics*, 42: 789–817. MR3210987. doi: <https://doi.org/10.1214/14-AOS1207>. 881
- Owen, A. B. (1988). “Empirical likelihood ratio confidence intervals for a single func-

- tional." *Biometrika*, 75(2): 237–249. MR0946049. doi: <https://doi.org/10.1093/biomet/75.2.237>. 900
- Owen, A. B. (2001). *Empirical likelihood*. CRC press. 900
- Powell, J. L. (1984). "Least absolute deviations estimation for the censored regression model." *Journal of Econometrics*, 25(3): 303–325. MR0752444. doi: [https://doi.org/10.1016/0304-4076\(84\)90004-6](https://doi.org/10.1016/0304-4076(84)90004-6). 884, 885
- Powell, J. L. (1986). "Censored regression quantiles." *Journal of Econometrics*, 32(1): 143–155. MR0853049. doi: [https://doi.org/10.1016/0304-4076\(86\)90016-3](https://doi.org/10.1016/0304-4076(86)90016-3). 884, 885
- Reich, B. J., Fuentes, M., and Dunson, D. B. (2011). "Bayesian spatial quantile regression." *Journal of the American Statistical Association*, 106(493): 6–20. MR2816698. doi: <https://doi.org/10.1198/jasa.2010.ap09237>. 877
- Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media. MR1707311. doi: <https://doi.org/10.1007/978-1-4757-3071-5>. 887
- Rodrigues, T. and Fan, Y. (2017). "Regression adjustment for noncrossing Bayesian quantile regression." *Journal of Computational and Graphical Statistics*, 26(2): 275–284. MR3640185. doi: <https://doi.org/10.1080/10618600.2016.1172016>. 881
- Sriram, K. (2015). "A sandwich likelihood correction for Bayesian quantile regression based on the misspecified asymmetric Laplace density." *Statistics & Probability Letters*, 107: 18–26. MR3412751. doi: <https://doi.org/10.1016/j.spl.2015.07.035>. 876, 879
- Sriram, K., Ramamoorthi, R., and Ghosh, P. (2016). "On Bayesian quantile regression using a pseudo-joint asymmetric Laplace likelihood." *Sankhya A*, 78(1): 87–104. MR3494580. doi: <https://doi.org/10.1007/s13171-015-0079-2>. 877
- Sriram, K., Ramamoorthi, R., Ghosh, P., et al. (2013). "Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density." *Bayesian Analysis*, 8(2): 479–504. MR3066950. doi: <https://doi.org/10.1214/13-BA817>. 876
- Womersley, R. S. (1986). "Censored discrete linear  $l_1$  approximation." *SIAM Journal on Scientific and Statistical Computing*, 7(1): 105–122. MR0819461. doi: <https://doi.org/10.1137/0907008>. 886, 899
- Wu, T., Narisetty, N. N. (2020). "Supplementary Material for "Bayesian multiple quantile regression for linear models using a score likelihood"." *Bayesian Analysis*. doi: <https://doi.org/10.1214/20-BA1217SUPP>. 884, 892
- Yang, Y. and He, X. (2012). "Bayesian empirical likelihood for quantile regression." *The Annals of Statistics*, 40(2): 1102–1131. MR2985945. doi: <https://doi.org/10.1214/12-AOS1005>. 876, 877
- Yang, Y. and Tokdar, S. T. (2017). "Joint estimation of quantile planes over arbitrary predictor spaces." *Journal of the American Statistical Association*, 112(519): 1107–

1120. MR3735363. doi: <https://doi.org/10.1080/01621459.2016.1192545>. 877, 881, 895
- Yang, Y., Wang, H. J., and He, X. (2016). “Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood.” *International Statistical Review*, 84(3): 327–344. MR3580414. doi: <https://doi.org/10.1111/insr.12114>. 876, 877, 879, 891, 895, 900
- Yu, K. and Moyeed, R. A. (2001). “Bayesian quantile regression.” *Statistics & Probability Letters*, 54(4): 437–447. MR1861390. doi: [https://doi.org/10.1016/S0167-7152\(01\)00124-9](https://doi.org/10.1016/S0167-7152(01)00124-9). 876, 896
- Zou, H., Yuan, M., et al. (2008). “Composite quantile regression and the oracle model selection theory.” *The Annals of Statistics*, 36(3): 1108–1126. MR2418651. doi: <https://doi.org/10.1214/07-AOS507>. 876, 899