International Statistical Review

International Statistical Review (2016), 84, 3, 327–344 doi:10.1111/insr.12114

Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood

Yunwen Yang[1], Huixia Judy Wang[2] and Xuming He[3]

[1]*Google Inc., Seattle, WA 98103*
E-mail: yunweny@google.com
[2]*Department of Statistics, George Washington University, Washington, DC 20052, USA*
E-mail: judywang@gwu.edu
[3]*Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA*
E-mail: xmhe@umich.edu

Summary

The paper discusses the asymptotic validity of posterior inference of pseudo-Bayesian quantile regression methods with complete or censored data when an asymmetric Laplace likelihood is used. The asymmetric Laplace likelihood has a special place in the Bayesian quantile regression framework because the usual quantile regression estimator can be derived as the maximum likelihood estimator under such a model, and this working likelihood enables highly efficient Markov chain Monte Carlo algorithms for posterior sampling. However, it seems to be under-recognised that the stationary distribution for the resulting posterior does not provide valid posterior inference directly. We demonstrate that a simple adjustment to the covariance matrix of the posterior chain leads to asymptotically valid posterior inference. Our simulation results confirm that the posterior inference, when appropriately adjusted, is an attractive alternative to other asymptotic approximations in quantile regression, especially in the presence of censored data.

Key words: Bayesian; censoring; posterior; quantile regression.

1 Introduction

Quantile regression, first introduced by Koenker & Bassett (1978), has been widely used as a valuable tool for analysing the conditional quantile functions of a response variable given covariates. In contrast to the ordinary least squares regression that focuses on the conditional mean function, quantile regression provides a more comprehensive analysis on how covariates may influence different aspects of the conditional distributions of the response. Quantile regression model also enhances the flexibility of parametric regression models by allowing error heteroscedasticity. Compared with nonparametric regression, quantile regression has a direct target on a quantile level of interest without modelling the whole conditional distribution and avoids the curse of dimensionality by assuming a parametric quantile function.

In recent years, Bayesian quantile regression has attracted attention because of some of its distinctive properties. For example, Bayesian quantile regression methods make use of Markov chain Monte Carlo (MCMC) algorithms to sample the parameter values from the posterior distribution, and the resultant estimator is as efficient as the classical estimator directly calculated through numerical optimisation. In some cases, the MCMC computation alleviates the computational curse of dimensionality in the optimisation of non-convex objective functions such as the one used in Powell (1986) for censored quantile regression. Moreover, uncertainty estimates or interval estimates can be calculated easily from a posterior sequence of MCMC draws. In contrast, the asymptotic variance–covariance of the conventional quantile estimator involves unknown conditional density functions, which are often difficult to estimate reliably.

The quantile regression models are most helpful when a parametric likelihood cannot be specified, and thus, a working likelihood is needed for the Bayesian approach to work. Some researchers considered non-parametric working likelihoods, for example, the Dirichlet process mixture models in Gelfand & Kottas (2002) and Kottas & Krnjajić (2009), an infinite mixture of normals in Reich *et al.* (2010) and Jeffreys' substitution likelihood in Dunson & Taylor (2005). Reich *et al.* (2011), Reich (2012) and Reich & Smith (2013) proposed semiparametric models on the entire quantile process. Lancaster & Jun (2010) considered exponential tilted empirical likelihood, while Otsu (2008) and Yang & He (2012) considered empirical likelihood. Arguably, the most popular choice of the working likelihood is the asymmetric Laplace (AL) distribution. A Bayesian approach based on the AL likelihood was formally discussed in Yu & Moyeed (2001) for linear quantile regression. In recent years, the AL likelihood has been adopted for Bayesian quantile regression in different contexts and applications, for instance, quantile regression with random effects (Geraci & Bottai, 2007; Yuan & Yin, 2009; Geraci & Bottai, 2013; Yue & Rue, 2011; Luo *et al.*, 2014; Wang, 2012), variable selection for quantile regression (Li *et al.*, 2010; Alhamzawi *et al.*, 2012; Alhamzawi & Yu, 2013; 2012), spatial quantile regression (Lum & Gelfand, 2012), quantile regression for count data with application to environmental epidemiology (Lee & Neocleous, 2010), non-parametric and semiparametric quantile regression models (Chen & Yu, 2009; Thompson *et al.*, 2010; Hu *et al.*, 2013; Waldmann *et al.*, 2013; Zhu *et al.*, 2013; Hu *et al.*, 2014), quantile regression with fixed censoring (Yu & Stander, 2007; Kozumi & Kobayashi, 2011; Kobayashi & Kozumi, 2012; Yue & Hong, 2014; Alhamazawi & Yu, 2015; Zhao & Lian, 2015), and binary quantile regression (Benoit & Poel, 2012; Benoit *et al.*, 2013; Miguéis *et al.*, 2013).

Whatever is chosen as the working likelihood, it is generally not the true data generating likelihood. Therefore, the validity of the posterior inference based on the working likelihood does not follow automatically from the Bayes formula. Yang & He (2012) established the asymptotic validity of the posterior inference based on the empirical likelihood. However, the general validity of the posterior inference is questionable. We wish to emphasise here that a direct use of the posterior interval based on a misspecified likelihood can be misleading.

In this article, we focus on Bayesian quantile regression using the AL working likelihood, as this is widely used in the literature. As we pointed out earlier, the posterior from this working likelihood is not the conditional distribution of the parameter given the data so the credible intervals obtained from the posterior do not generally have the right Bayesian confidence level. Furthermore, we show that the posterior based on the misspeficied AL likelihood does not approximate the sampling distribution of the parameter estimates even as the sample size increases, but asymptotically valid posterior inference can be achieved with a simple adjustment. We present posterior variance adjustments in two cases, quantile regression for complete data and for data with fixed censoring. The adjustments enable effective and valid posterior inference without requiring the estimation of the unknown conditional density functions as in the frequentist inference methods, and they are shown through simulation to have advantages over other inference methods, especially for the cases of fixed censoring.

The rest of the paper is organised as follows. In Section 2, we review AL-based Bayesian quantile regression methods including the recent development in computation. In Section 3, we discuss the asymptotic properties of the posterior from the AL-based Bayesian quantile regression and present adjustments on the posterior variance for complete data and for data with fixed censoring. To assess the finite sample performance, we evaluate the interval estimates from the posterior with and without adjustments and compare them with the usual large sample approximations under the frequentist approach through simulation studies and the analysis of a women's labour force data in Section 4. While the example on the women's labour force data is not meant to be comprehensive in any way, the alarming difference in the interval estimates between the AL-based Bayesian quantle regression with and without adjustments shows that we must remain vigilant in the interpretation of any posterior obtained from a working likelihood. We hope that this paper plays a positive role in the promotion of Bayesian inference for quantile regression in a wide variety of applications. Some concluding remarks are given in Section 5. The technical details to support the asymptotic validity of the adjusted posterior inference are provided in the Appendix.

2 Bayesian Quantile Regression with Asymmetric Laplace Likelihood

2.1 Basic Setup

Suppose that Y is the continuous response variable of interest and $\mathbf{X}$ is the p-dimensional vector of covariates with the first element equal to one. At a given quantile level $\tau \in (0, 1)$, we consider the following linear quantile regression model

$$Q_\tau(Y \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta}(\tau), \tag{2.1}$$

where $Q_\tau(Y \mid \mathbf{X} = \mathbf{x})$ denotes the τ-th conditional quantile of Y given $\mathbf{X} = \mathbf{x}$ and $\boldsymbol{\beta}(\tau)$ is the quantile coefficient vector.

Based on a random sample $\mathcal{D} = \{(y_i, \mathbf{x}_i), i = 1, \ldots, n\}$ of $(Y, \mathbf{X})$, the unknown parameters $\boldsymbol{\beta}(\tau)$ can be estimated by $\hat{\boldsymbol{\beta}}(\tau)$, which minimises

$$R_n(\boldsymbol{\beta}, \mathcal{D}) = \sum_{i=1}^{n} \rho_\tau\left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}\right), \tag{2.2}$$

where $\rho_\tau(\mu) = \mu\{\tau - I(\mu < 0)\}^*$ is the quantile loss function given in Koenker & Bassett (1978). In the rest of the paper, we omit the τ in various expressions such as $\boldsymbol{\beta}(\tau)$ for the sake of simplicity.

To incorporate quantile regression models into a Bayesian framework, we consider the AL working likelihood

$$L(\boldsymbol{\beta}; \mathcal{D}) = \frac{\tau^n (1-\tau)^n}{\sigma^n} \exp\left\{-\frac{\sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma}\right\},$$

where σ is a fixed scale parameter. With a prior specified as $p_0(\boldsymbol{\beta})$ on $\boldsymbol{\beta}$, the posterior of $\boldsymbol{\beta}$ can be formally written as

$$p_n(\boldsymbol{\beta} \mid \mathcal{D}) \propto p_0(\boldsymbol{\beta}) \exp\left\{-\frac{\sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})}{\sigma}\right\}. \tag{2.3}$$

Any reasonable choice of the prior, including the flat prior, leads to a proper posterior under some mild conditions (Yu & Moyeed, 2001; Tsionas, 2003; Choi & Hobert, 2013). For any fixed prior, the asymptotic properties of the posterior are independent of the prior choices, even though the computational algorithms may have to adapt to the choice of the prior.

* Correction added on 21 December 2015, after first online publication: "τ" in equation corrected to "μ".

The priors can also be used to enable a Lasso-type regularisation in quantile regression from a Bayesian perspective. As proposed in Koenker (2004) and Li & Zhu (2008), the Lasso-regularised quantile regression is given by

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right) + \lambda \sum_{j=1}^{p} |\beta_j|,$$

where β_j is the j-th element of $\boldsymbol{\beta}$. To accommodate the penalty term $\lambda \sum_{j=1}^{p} |\beta_j|$, Li *et al.* (2010) proposed a Laplace prior $p_0(\boldsymbol{\beta} \mid \lambda) = (\tau\lambda/2)^p \exp\left\{-\tau\lambda \sum_{j=1}^{p} |\beta_j|\right\}$ and suggested a Gibbs sampling algorithm. Alhamzawi *et al.* (2012) further extended it to Bayesian adaptive Lasso quantile regression.

2.2 Computation and Properties

It is natural to use MCMC methods for posterior sampling. The posterior of $\boldsymbol{\beta}(\tau)$ is in general intractable, but sampling can be simplified by using a mixture representation of the AL distribution.

Consider the working model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau) + \epsilon_i, \; i = 1, \ldots, n, \tag{2.4}$$

where ϵ_i are independent random variables following the AL distribution $AL(0, \tau, \sigma)$ with density

$$f(\epsilon_i) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{\rho_\tau(\epsilon_i)}{\sigma}\right\}.$$

Here, σ can either be fixed or considered as part of the parameter to be assigned a prior distribution. By Kotz *et al.* (2001), ϵ_i can be represented by a scale mixture of normals,

$$\epsilon_i = \sigma \left(\theta_1 v_i + \theta_2 z_i \sqrt{v_i} \right), \tag{2.5}$$

where $\theta_1 = (1-2\tau)/\{\tau(1-\tau)\}$, $\theta_2^2 = 2/\{\tau(1-\tau)\}$, $z_i \sim N(0,1)$, v_i follows the standard exponential distribution, and z_i and v_i are independent.

By using the representation (2.5) and assuming a proper Gaussian-Inverse Gamma prior on $(\boldsymbol{\beta}(\tau), \sigma)$, Kozumi & Kobayashi (2011) proposed a three-variable Gibbs sampling algorithm. More specifically, Model (2.4) can be rewritten as

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\tau) + \sigma \left(\theta_1 v_i + \theta_2 z_i \sqrt{v_i} \right), \; i = 1, \ldots, n.$$

With a normal prior on $\boldsymbol{\beta}(\tau)$, the full conditional density of $\boldsymbol{\beta}(\tau)$ given $\mathbf{y} = (y_1, \ldots, y_n)^T$ and $\mathbf{v} = (v_1, \cdots, v_n)^\top$ is normal, and the conditional density of v_i is a generalised inverse Gaussian distribution. Consequently, a Gibbs sampler based on standard distributions can be applied. If σ is assigned a prior as inverse Gamma that is independent of $\boldsymbol{\beta}(\tau)$, then the conditional distribution of σ given the other quantities remains in the family of inverse Gamma, as detailed in Kozumi & Kobayashi (2011). Khare & Hobert (2012) showed that the Markov chain underlying this three-variable Gibbs sampling algorithm converges at a geometric rate.

If an improper prior on $(\boldsymbol{\beta}(\tau), \sigma)$ is used, Choi & Hobert (2013) used the same mixture representation (2.5) to propose a data augmentation algorithm and showed that the Markov chains associated with the algorithms are geometrically ergodic. In addition, Choi & Hobert (2013) showed that when the prior on $(\boldsymbol{\beta}(\tau), \sigma)$ takes the form $p_0(\boldsymbol{\beta}(\tau), \sigma^2) \propto (\sigma^2)^{-(a+1)/2}$ on $\sigma^2 > 0$, where a is a hyper-parameter, the posterior is proper if and only if (i) $a > -n + p + 1$, (ii) the design matrix is of full column rank and (iii) $\mathbf{y}$ is not in the column space of the design matrix. Currently, there are two R (R Core Team, 2014) packages, *brq* (Alhamzawi, 2012) and *bayesQR* (Benoit *et al.*, 2014), that utilise efficient Gibbs sampling algorithms for Bayesian quantile regression.

Even though a prior distribution on σ can be used in Bayesian computation, we find through our empirical studies that fixing σ at a pre-estimated value often makes the MCMC algorithm more efficient. As we shall show in the following section, the proposed adjustments to the posterior variance make the results asymptotically invariant to the choice of any fixed σ. For a specific choice of σ to reflect the scale of the conditional distributions, we refer to Remark 1 in Section 3.1.

3 Posterior Variance Adjustment

Based on the AL working likelihood, the posterior mean and variance of $\boldsymbol{\beta}(\tau)$ can be computed directly from the MCMC chains. Based on empirical evidence, Yu & Moyeed (2001) argued that the use of the AL likelihood is satisfactory for quantile regression, even when the likelihood is misspecified. Sriram *et al.* (2013) established sufficient conditions for the posterior consistency of model parameters in Bayesian quantile regression with the AL likelihood. However, the posterior consistency results do not imply that the interval estimates constructed from the posterior are automatically valid. It is tempting to construct interval estimates, whether they are called credible intervals or confidence intervals, from the quantiles of the posterior or by normal approximations using the variance–covariance matrix of the posterior sequence, as reported in Yu & Moyeed (2001), Li *et al.* (2010), Alhamzawi *et al.* (2012), Yue & Hong (2014) and Lum & Gelfand (2012), among others. Here, we argue that the posterior variance–covariance must be adjusted for the interval estimates to be asymptotically valid. We will present the proposed adjustments for linear quantile regression with complete data in Section 3.1 and with fixed censored data in Section 3.2. The basic idea of asymptotically valid posterior inference goes back to Chernozhukov & Hong (2003) and Yang & He (2012), but the specific results presented in this section are new.

3.1 Quantile Regression with Complete Data

Consider the linear quantile regression model (2.1) with the true parameter $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0$. Under the conditions that guarantee the asymptotic normality of the conventional quantile regression estimator $\hat{\boldsymbol{\beta}}(\tau)$ that minimises (2.2), we show in the Appendix that for $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| = O(n^{-1/2})$, the posterior density (assuming any fixed σ) is

$$p_n(\boldsymbol{\beta} \mid \mathcal{D}) \propto p_0(\boldsymbol{\beta}) \exp\left\{-\frac{n\{\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\tau)\}^\top D_1\{\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\tau)\} + o_p(1)}{2\sigma}\right\}, \tag{3.1}$$

where $D_1 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n f_{Y_i}(\mathbf{x}_i^\top \boldsymbol{\beta}_0 \mid \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^\top$ with $f_{Y_i}(\cdot \mid \mathbf{x})$ as the conditional density of the response Y_i given covariates $\mathbf{x}$. The aforementioned expansion suggests that for large n and $p_0(\boldsymbol{\beta}) \propto 1$, $p_n(\boldsymbol{\beta}|D)$ is approximately a normal density with mean $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\tau)$ and variance–covariance $\sigma D_1^{-1}/n$. Let $\hat{\Sigma}(\sigma)$ denote the posterior variance–covariance matrix with $n\hat{\Sigma}(\sigma) \approx \sigma D_1^{-1}$. On the other hand, it is known (for instance, Koenker 2005, Chapter 3) that the asymptotic variance of $n^{1/2}\hat{\boldsymbol{\beta}}(\tau)$ is $\tau(1-\tau)D_1^{-1}D_0D_1^{-1}$, where $D_0 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$.

This explains why the posterior variance is not the right approximation to the sampling variance of $\hat{\boldsymbol{\beta}}(\tau)$.

The good news is that we can use a simple adjustment to $\hat{\Sigma}(\sigma)$ to obtain asymptotically valid posterior inference,

$$\hat{\Sigma}_{\text{adj}} = \frac{n\tau(1-\tau)}{\sigma^2}\hat{\Sigma}(\sigma)\hat{D}_0\hat{\Sigma}(\sigma),$$

where $\hat{D}_0 = n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$ provides a consistent estimator of D_0. A posterior interval can be constructed based on a normal approximation using $\hat{\Sigma}_{\text{adj}}$.

Remark 1. *The unadjusted posterior variance $\hat{\Sigma}(\sigma)$ depends, in a rather complicated way, on the specification of σ, the scale parameter in the AL distribution. However, the adjusted posterior variance, $\hat{\Sigma}_{adj}$, is asymptotically invariant in the value of σ. In practice, to adjust for the scale of the data, we recommend fixing σ at $n^{-1}R_n(\hat{\boldsymbol{\beta}}(0.5), \mathcal{D})$, which is the maximum likelihood estimator of σ under the AL working likelihood at the median.*

The proposed adjustment not only applies to the posterior with the prior $p_0(\boldsymbol{\beta}) \propto 1$ but also applies to other proper priors. For asymptotic analysis about a more general class of priors, including sample-size-dependent priors, we refer to Yang & He (2012).

3.2 Quantile Regression with Fixed Censoring

Quantile regression is especially appealing for censored data, because many of the quantiles are identifiable under censoring when the conditional mean is not identifiable without additional (and often not verifiable) model assumptions. In this subsection, we discuss an important application of quantile regression when the response variable is subject to fixed censoring because of, for example, top or bottom coding.

Suppose that T is a latent continuous response variable of interest. Because of left censoring, we only observe $Y = \max(T, C)$ and the censoring indicator $\delta = I(T > C)$, where C is a known censoring point. Without loss of generality, we assume $C = 0$. Let $\mathcal{D} = \{(y_i, \mathbf{x}_i, \delta_i), i = 1, \ldots, n\}$ be a random sample of $(Y, \mathbf{X}, \delta)$. We consider the following linear quantile regression model,

$$Q_\tau(T \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}^\top\boldsymbol{\beta}(\tau). \tag{3.2}$$

Various estimation methods have been developed for censored quantile regression, including Portnoy (2003), Ying *et al.* (1995), Peng & Huang (2008) and Wang & Wang (2009) for random censoring, Lin *et al.* (2012) for double censoring and Powell (1986) and Tang *et al.* (2012) for fixed censoring. In this paper, we focus on the estimator of Powell (1986).

Because $Y = \max(T, 0)$, by the equivariance property of quantiles to monotone transformations, model (3.2) implies that the τ-th conditional quantile of the observed response Y is $Q_\tau(Y \mid \mathbf{X} = \mathbf{x}) = \max(\mathbf{x}^\top\boldsymbol{\beta}(\tau), 0)$. Motivated by this, Powell (1986) proposed to estimate $\boldsymbol{\beta}(\tau)$ by $\hat{\boldsymbol{\beta}}(\tau)$, the minimiser of the following objective function

$$R_n(\boldsymbol{\beta}, \mathcal{D}) = \sum_{i=1}^n \rho_\tau\left\{y_i - \max(\mathbf{x}_i^\top\boldsymbol{\beta}, 0)\right\}. \tag{3.3}$$

As pointed out in Womersley (1986) and Buchinsky & Hahn (1998), the objective function (3.3) is highly non-convex in $\boldsymbol{\beta}$, and the optimisation is computationally challenging. Particularly, the existing computing methods become unstable when heavy censoring is present. More details related to the computational issues can be referred to Buchinsky (1994), Fitzenberger (1997) and Chernozhukov & Hong (2002) and Portnoy (2010).

Yu & Stander (2007) proposed a Bayesian Tobit quantile regression model, which employs the AL likelihood based on the objective function in (3.3),

$$L(\boldsymbol{\beta};\mathcal{D}) = \frac{\tau^n(1-\tau)^n}{\sigma^n}\exp\left\{-\frac{\sum_{i=1}^n \rho_\tau\left\{y_i - \max\left(\mathbf{x}_i^\top\boldsymbol{\beta},0\right)\right\}}{\sigma}\right\}.$$

The resultant posterior can be written as

$$p_n(\boldsymbol{\beta}\mid\mathcal{D}) \propto p_0(\boldsymbol{\beta})\exp\left\{-\frac{\sum_{i=1}^n \rho_\tau\left\{y_i - \max\left(\mathbf{x}_i^\top\boldsymbol{\beta},0\right)\right\}}{\sigma}\right\}. \tag{3.4}$$

This is a direct extension of the AL working likelihood discussed in Section 3.1. Because the optimisation of (3.3) is far more difficult than the quantile regression problem without censoring, the Bayesian computation for the censored quantile regression is attractive from the computational perspective. A Gibbs algorithm was described in Yu & Stander (2007) with $\sigma = 1$ and $p_0(\boldsymbol{\beta}) \propto 1$.

Assume the censored quantile regression model (3.2) with the true parameter $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0$. Under the assumptions of Powell (1986), for any $\boldsymbol{\beta}$ such that $||\boldsymbol{\beta}-\boldsymbol{\beta}_0|| = O(n^{-1/2})$, we have the following quadratic expansion,

$$p_n(\boldsymbol{\beta}\mid\mathcal{D}) \propto p_0(\boldsymbol{\beta})\exp\left\{-\frac{n\{\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}(\tau)\}^\top D_1\{\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}(\tau)\} + o_p(1)}{2\sigma}\right\}, \tag{3.5}$$

$$\text{where } D_1 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n f_T\left(\mathbf{x}_i^\top\boldsymbol{\beta}_0 \mid \mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^\top I\left(\mathbf{x}_i^\top\boldsymbol{\beta}_0 > 0\right),$$

and $f_T(\cdot|\mathbf{x}_i)$ denotes the conditional density of the latent response T given covariates $\mathbf{x}_i$; see the verification of this result in the Appendix.

The expansion (3.5) suggests that for large n and $p_0(\boldsymbol{\beta}) \propto 1$, the posterior is approximately normal with mean $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\tau)$ and variance $\sigma D_1^{-1}/n$, which is different from the asymptotic variance of $\hat{\boldsymbol{\beta}}(\tau)$, that is, $\tau(1-\tau)D_1^{-1}D_0D_1^{-1}/n$, where $D_0 = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top I\left(\mathbf{x}_i^\top\boldsymbol{\beta}_0 > 0\right)$. Therefore, an interval estimate directly from the posterior is not asymptotically valid, and a simple adjustment based on the posterior variance $\hat{\Sigma}(\sigma)$ is needed,

$$\hat{\Sigma}_{\text{adj}} = \frac{n\tau(1-\tau)}{\sigma^2}\hat{\Sigma}(\sigma)\hat{D}_0\hat{\Sigma}(\sigma), \text{where } \hat{D}_0 = n^{-1}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top I\left(\mathbf{x}_i^\top\hat{\boldsymbol{\beta}}(\tau) > 0\right).$$

Similarly as in Section 3.1, the adjusted posterior variance $\hat{\Sigma}_{\text{adj}}$ is asymptotically invariant with respect to the value of the scale parameter σ, and it can be used to construct an asymptotically valid interval estimate for $\boldsymbol{\beta}(\tau)$ using normal approximations.

4 Numerical Studies

We carry out two simulation studies to assess the finite sample performance of the proposed posterior variance adjustments, one for complete data and the other for fixed censored data. Each simulation study uses 1000 Monte Carlo replications, and the Bayesian methods use MCMC chains of length 20 000 with a burn-in of 4000.

4.1 Simulation for Complete Data

In this study, we focus on two sample sizes $n = 200$ and 500. Three data-generating models are specified below.

Case 1: $y_i = 2x_{1i} + 2x_{2i} + e_i$, where x_{1i} and x_{2i} are independent standard normal variables, and $e_i \sim t_3$, Student's t-distribution with three degrees of freedom, is independent of (x_{1i}, x_{2i});
Case 2: $y_i = 2/3 + 4x_{1i} + 4x_{2i} + (1 + 0.6x_{i1}^2)e_i$, where x_{1i} and x_{2i} are independent standard normal variables, and $e_i \sim N(0, 1)$ is independent of (x_{1i}, x_{2i});
Case 3: $y_i = a(u_i) + b_1(u_i)x_{i1} + b_2(u_i)x_{i2}$, where $u_i \sim \mathit{Unif}(0, 1)$ is independent of (x_{1i}, x_{2i}), $a(u) = 0.5 + \Phi^{-1}(u)$, $b_1(u) = 2 + u^2$, $b_2(u) = 2$, $x_{i1} \sim \chi_2^2/2$ and $x_{i2} \sim N(0, 1)$ are independent, where χ_2^2 denotes the chi-square distribution with two degrees of freedom.

Case 1 represents a homoscedastic error model, and *Cases 2 and 3* represent heteroscedastic error models. In both *Cases* 1 and 3, the conditional quantile function of Y takes the form $Q_\tau(Y|x_1, x_2) = a(\tau) + b_1(\tau)x_1 + b_2(\tau)x_2$ for any $\tau \in (0, 1)$, where $a(\tau) = F_{t_3}^{-1}(\tau)$ and $b_1(\tau) = b_2(\tau) = 2$ in Case 1. In Case 2, the linear conditional quantile function $Q_\tau(Y|x_1, x_2)$ holds only at $\tau = 0.5$ with $a(0.5) = 2/3$ and $b_1(0.5) = b_2(0.5) = 4$.

Based on normal approximations, we construct confidence intervals for $a(\tau)$, $b_1(\tau)$, $b_2(\tau)$ using the unadjusted posterior variance for the Bayesian quantile regression with the AL likelihood and using the proposed adjustments to the posterior variance. For comparison, we include two forms of confidence intervals from the conventional quantile regression obtained, respectively, by using the default rank score method and by the Wald method based on the asymptotic approximation to the variance–covariance matrix for models with heteroscedastic errors in the R package *quantreg* (Koenker, 2015). Table 1 summarises the coverage probabilities and the average lengths of confidence intervals of the competing methods in Cases 1–3 at $\tau = 0.5$, and Table 2 summarises the results at two tail quantiles in Case 1 ($\tau = 0.1$) and Case 3 ($\tau = 0.9$). The standard error of the coverage probabilities with a nominal level 90% is 1%. The results confirm that the Bayesian intervals from the AL likelihood, if unadjusted, have poor coverage, which is mainly due to the misspecification of the likelihood. In contrast, the intervals with the proposed variance adjustments have coverage close to the nominal level 90% in all the scenarios considered. Although the two non-Bayesian methods also produce asymptotically valid interval estimates, their performances are less stable, even at $n = 500$, because of the difficulty in approximating the variance–covariance matrices of the quantile estimates.

4.2 Simulation for Data with Fixed Censoring

For censored quantile regression, we consider the following two data-generating models,

Case 4: $T_i = 2.5 + 5x_i + \{1 + (x_i - 0.5)^2\}e_i$, $i = 1, \ldots, n$, where $e_i \sim t_3$ is independent of $x_i \sim N(0, 1)$;

Table 1. *Empirical coverage probabilities and empirical mean lengths of different confidence intervals with nominal level 90% in Cases 1–3 at $\tau = 0.5$ with $n = 200, 500$. The standard errors for EML are no more than 0.005 in all entries.*

			100×ECP			EML		
Case	n	Method	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$
1	$n = 200$	BAL_{adj}	90	91	93	0.33	0.33	0.33
		BAL	86	88	89	0.29	0.29	0.29
		RQ_{rank}	87	89	90	0.31	0.31	0.31
		RQ_{nid}	92	90	90	0.34	0.32	0.32
	$n = 500$	BAL_{adj}	91	90	93	0.20	0.21	0.21
		BAL	86	86	89	0.18	0.18	0.18
		RQ_{rank}	89	89	90	0.20	0.20	0.20
		RQ_{nid}	91	89	92	0.21	0.20	0.20
2	$n = 200$	BAL_{adj}	90	91	92	0.41	0.66	0.42
		BAL	86	77	87	0.34	0.44	0.35
		RQ_{rank}	89	90	89	0.39	0.64	0.40
		RQ_{nid}	91	70	90	0.42	0.40	0.40
	$n = 500$	BAL_{adj}	92	90	93	0.26	0.42	0.26
		BAL	87	75	88	0.22	0.28	0.22
		RQ_{rank}	91	90	91	0.25	0.41	0.25
		RQ_{nid}	92	68	90	0.26	0.25	0.25
3	$n = 200$	BAL_{adj}	90	88	90	0.55	0.58	0.39
		BAL	81	73	82	0.43	0.37	0.30
		RQ_{rank}	87	87	89	0.51	0.55	0.38
		RQ_{nid}	89	88	88	0.54	0.57	0.38
	$n = 500$	BAL_{adj}	92	90	92	0.34	0.37	0.25
		BAL	84	72	82	0.27	0.23	0.19
		RQ_{rank}	90	91	89	0.32	0.35	0.24
		RQ_{nid}	93	91	90	0.34	0.36	0.24

ECP, empirical coverage probability; EML, empirical mean length; BAL_{adj} and BAL, the Bayesian quantile regression based on asymmetric Laplace likelihood, with and without posterior variance adjustment, respectively; RQ_{rank}, the interval estimates based on rank score test in quantile regression; RQ_{nid}, the Wald-type interval estimates in quantile regression based on the asymptotic approximation to the variance–covariance matrix under heteroscedastic errors.

Case 5: $T_i = 0.2 + 3x_{i1} - 2x_{i2} + (0.5 + 0.5x_{i1} + x_{i2})e_i, i = 1, \ldots, n$, where $x_{i1} \sim \chi_1^2$ and $x_{i2} \sim \text{Bernoulli}(0.5)$ are independent, and $e_i \sim N(0, 1)$ is independent of (x_{i1}, x_{i2}).

Because of censoring, we observe $y_i = \max(T_i, 0)$ instead of T_i, and in both cases, we have the censoring proportions around 30%. In *Case 4*, the conditional quantile of T given x is $a(\tau) + b(\tau)x$ only at $\tau = 0.5$ with $a(0.5) = 2.5$ and $b(0.5) = 5$ and is nonlinear in x at the other quantiles. In *Case 5*, the conditional quantiles are $a(\tau) + b_1(\tau)x_1 + b_2(\tau)x_2$ for all $\tau \in (0, 1)$ with $a(\tau) = 0.2 + 0.5\Phi^{-1}(\tau)$, $b_1(\tau) = 3 + 0.5\Phi^{-1}(\tau)$ and $b_2(\tau) = -2 + \Phi^{-1}(\tau)$. We present results for $\tau = 0.5$ in Case 4 at two sample sizes $n = 200, 500$, but for $\tau = 0.25$ and 0.5 with $n = 500$ in Case 5.

The posterior intervals from the AL working likelihood are constructed, and their performances are compared with the bootstrap-based confidence intervals (based on 100 bootstrap replications) for the Powell's estimator. The simulation results are given in Tables 3 and 4. Clearly the interval estimates from the Bayesian method based on Powell's objective function, referred to as BP, have poor coverage probabilities, but the proposed posterior variance adjustment leads to respectable performance. The frequentist intervals from Powell

Table 2. *Empirical coverage probabilities and empirical mean lengths of different confidence intervals with nominal level* 90% *in Cases 1 and 3 at tail quantiles ($\tau = 0.1$ or 0.9) with $n = 200$ and 500. The standard errors for EML range from 0.004 to 0.008 in this table.*

				100×ECP			EML		
Case	τ	n	Method	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$
1	0.1	200	BAL_{adj}	91	92	92	0.76	0.71	0.73
			BAL	83	85	84	0.55	0.54	0.54
			RQ_{rank}	88	89	88	0.70	0.66	0.66
			RQ_{nid}	95	89	90	0.90	0.72	0.72
		500	BAL_{adj}	92	89	90	0.45	0.44	0.44
			BAL	83	81	83	0.34	0.34	0.34
			RQ_{rank}	90	88	89	0.43	0.42	0.41
			RQ_{nid}	93	89	88	0.48	0.43	0.43
3	0.9	200	BAL_{adj}	93	89	93	0.76	0.74	0.55
			BAL	90	82	91	0.66	0.55	0.47
			RQ_{rank}	87	87	89	0.69	0.70	0.50
			RQ_{nid}	88	83	86	0.73	0.66	0.49
		500	BAL_{adj}	91	88	92	0.46	0.46	0.33
			BAL	88	80	90	0.41	0.35	0.29
			RQ_{rank}	88	88	90	0.43	0.44	0.31
			RQ_{nid}	90	85	89	0.45	0.44	0.31

ECP, empirical coverage probabilities; EML, empirical mean lengths.
The notations follow Table 1.

Table 3. *Empirical coverage probabilities and empirical mean lengths for confidence intervals with nominal level* 90% *for Case 4 at $\tau = 0.5$. The standard errors for EML are in the range of 0.004 to 0.008 in this table.*

	100×ECP		EML		100×ECP		EML	
Method	$a(\tau)$	$b(\tau)$	$a(\tau)$	$b(\tau)$	$a(\tau)$	$b(\tau)$	$a(\tau)$	$b(\tau)$
	$n = 200$				$n = 500$			
BP_{adj}	91	90	0.71	1.16	90	91	0.44	0.72
BP	86	78	0.56	0.78	84	76	0.35	0.49
POWELL	85	87	0.59	1.04	84	85	0.36	0.63

ECP, empirical coverage probability; EML, empirical mean length; BP_{adj} and BP, the Bayesian quantile regression method using Powell's objective function with and without posterior variance adjustment, respectively; POWELL, the bootstrap-based interval estimates of Powell (1986).

(denoted as POWELL in the tables) have undercoverage even with the bootstrap method. Part of the issues with POWELL is that we might not find the right solution through optimisation for every bootstrapped data set. Our empirical work shows that the Bayesian quantile regression method with the proposed posterior variance adjustment is attractive for inference when Powell's estimator is used for censored data.

4.3 Analysis of a Women's Labour Force Data

We demonstrate the effect of the proposed posterior variance adjustment on a women's labour force participation data, which was analysed in Mroz (1987) and Yu & Stander (2007). We aim to investigate the relationship between women's working hours, annual non-wife household income (i.e. the household income excluding the wife's labour income) and education. The

Table 4. *Empirical coverage probabilities and empirical mean lengths for confidence intervals with nominal level* 90% *for Case 5. For* $\tau = 0.25$*, the standard errors for EML are around 0.005 with* $a(\tau)$ *and* $b_1(\tau)$ *and are around 0.015 with* $b_2(\tau)$*; for* $\tau = 0.5$*, they are around 0.003 with* $a(\tau)$ *and* $b_1(\tau)$ *and are around 0.01 with* $b_2(\tau)$.

		100×ECP			EML		
τ	Method	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$	$a(\tau)$	$b_1(\tau)$	$b_2(\tau)$
0.25	BP_{adj}	90	86	88	0.35	0.47	1.64
	BP	80	58	59	0.25	0.22	0.80
	POWELL	89	86	86	0.34	0.45	1.37
0.5	BP_{adj}	90	89	89	0.31	0.40	1.07
	BP	78	61	66	0.21	0.19	0.58
	POWELL	90	89	83	0.31	0.39	0.87

ECP, empirical coverage probabilities; EML, empirical mean lengths.
The notations follow Table 3.

observed outcome variable Y_{hrs} is the wife's hours of work outside home in 1975, and the two covariates are x_{inc} (non-wife household income with units \$1000) and x_{edu} (wife's educational attainment in years). The data set contains a total of 753 observations, among which 325 (43%) have zero work hours and thus are treated as left censored to fit a linear quantile model. The sample means of x_{inc} and x_{edu} are 20.43 and 12.29, respectively. For easier interpretation, we centre both covariates at zero by subtracting their means before carrying out the data analysis.

We consider the following linear quantile regression model:

$$Q_\tau(T_{hrs} \mid x_{inc}, x_{edu}) = a(\tau) + b_1(\tau)x_{inc} + b_2(\tau)x_{edu},$$

where T_{hrs} is women's latent total working hours with $Y_{hrs} = \max(T_{hrs}, 0)$ with left censoring. We focus on quantiles $\tau = 0.5, 0.75$ and 0.9.

The goodness-of-fit test in Wang (2008) suggests that the linear quantile functions are appropriate at the selected quantile levels. We used the Bayesian estimates based on the AL likelihood with fixed σ and find that the unadjusted posterior intervals depend significantly on the value of σ (Figure 1). In fact, when $\sigma = 1$ as used in Yu & Stander (2007), the 95% intervals on the quantile coefficients are remarkably narrow, but they cannot be taken seriously. With the proposed posterior variance adjustments, the intervals are rather stable across different values of σ. Figure 2 shows the interval estimates of $b_2(\tau)$ from the Bayesian methods with a pre-estimated $\sigma = 429$, that is, $\log\sigma = 6.06$, as suggested in Remark 1. The results show that education is not statistically significant at $\tau = 0.9$. In this example, the Bayesian intervals after variance adjustments are not far from the frequentist intervals from bootstrapping Powell's estimate.

5 Discussion

The specification of the AL working likelihood relies on the value of the scale parameter σ. In this paper, we consider a fixed σ for easier computation, instead of involving it in the MCMC iteration, because the estimation and inference of σ itself is not of interest in a quantile regression model. It is shown in the paper that the posterior inference with a fixed σ without any variance adjustment can be misleading. However, when the proposed adjustment is used, the posterior inference is asymptotically independent of the choice of σ. In finite samples, a reasonably chosen σ that adapts to the scale of the residuals could help the mixing property of the MCMC chain.

The Bayesian quantile regression methods are especially useful when the quantile loss function is non-convex, as in the case of censored data. The basic idea of posterior variance

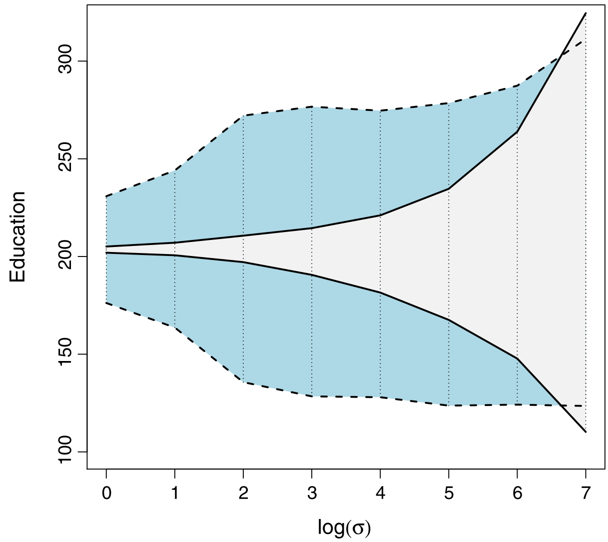

Figure 1. *Ninety-five percent confidence intervals for the coefficients of x_{edu} at $\tau = 0.5$ as σ varies in BP, the Bayesian quantile regression method using Powell's objective function. The dashed lines correspond to the 95% confidence intervals from BP_{adj} with posterior variance adjustment; the solid lines correspond to the 95% confidence intervals from BP without posterior variance adjustment.*

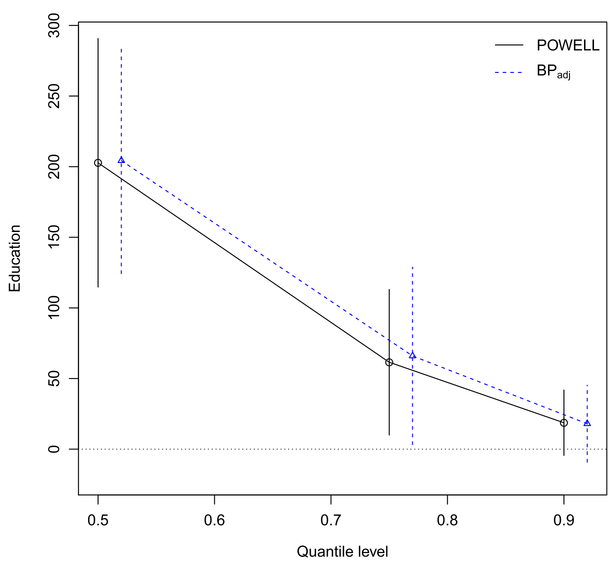

Figure 2. *The coefficient estimates and 95% confidence intervals of x_{edu} at $\tau = 0.5, 0.75, 0.9$. The black circles and solid lines represent the point estimates and the interval estimates from Powell (1986); the blue triangles and dashed lines represent the estimates from the Bayesian method with posterior adjustment.*

adjustments can be attributed to the work of Chernozhukov & Hong (2003), but we feel that the need for adjustments is not widely appreciated yet. We hope that this article helps promote the appropriate use of posterior inference in quantile regression.

When multiple quantiles are of interest, the proposed adjustment to posterior variances can be extended to the Bayesian quantile regression with an AL likelihood employing a combined objective function over multiple quantiles, such as the objective function in the composite quantile regression of Zou & Yuan (2008). The proposed adjustment can also be extended to Bayesian quantile regression with longitudinal data or random censored data. Future research is needed in those directions.

Acknowledgements

The authors are grateful for two anonymous referees whose comments on our earlier draft of the paper helped us improve presentation.

The research of Wang is supported by the National Science Foundation (NSF) CAREER award DMS-1149355. The research of He is partially supported by the NSF DMS-1307566 and the National Natural Science Foundation of China grant 11129101.

References

Alhamazawi, R. & Yu, K. (2015). Bayesian Tobit quantile regression using g-prior distribution with ridge parameter. *J. Statist. Comput. Simulation*, **85**, 2903–2918.

Alhamzawi, R. (2012). Brq: Bayesian analysis of quantile regression models. R package version 1.0.

Alhamzawi, R. & Yu, K. (2012). Bayesian lasso-mixed quantile regression. *J. Statist. Comput. Simulation*, **84**, 868–880.

Alhamzawi, R. & Yu, K. (2013). Conjugate priors and variable selection for bayesian quantile regression. *Comput. Stat. Data Anal.*, **64**, 209–219.

Alhamzawi, R., Yu, K. & Benoit, D. F. (2012). Bayesian adaptive lasso quantile regression. *Statistical Model.*, **12**, 279–297.

Benoit, D. F., Al-Hamzawi, R., Yu, K. & Van den Poel, D. (2014). bayesQR: Bayesian quantile regression. R package version 2.2.

Benoit, D. F., Alhamzawi, R. & Yu, K. (2013). Bayesian lasso binary quantile regression. *Comput. Stat.*, **28**, 2861–2873.

Benoit, D. F. & Poel, D. V. D. (2012). Binary quantile regression: a Bayesian approach based on the asymmetric Laplace distribution. *J. Appl. Econom.*, **27**, 1174–1188.

Buchinsky, M. (1994). Changes in the U.S. wage structure 1963–1987: application of quantile regression. *Econometrica*, **62**, 405–458.

Buchinsky, M. & Hahn, J. (1998). An alternative estimator for the censored quantile regression model. *Econometrica*, **66**, 653–672.

Chen, C. & Yu, K. (2009). Automatic Bayesian quantile regression curve fitting. *Stat. Comput.*, **19**, 271–281.

Chernozhukov, V. & Hong, H. (2002). Three-step censored quantile regression and extramarital affairs. *J. Am. Statist. Assoc.*, **97**, 872–882.

Chernozhukov, V. & Hong, H. (2003). An MCMC approach to classical estimation. *J. Econometrics*, **114**, 293–346.

Choi, H.M. & Hobert, J.P. (2013). Analysis of MCMC algorithms for Bayesian linear regression with Laplace errors. *J. Multivar. Anal.*, **117**, 32–40.

Dunson, T. & Taylor, J. (2005). Approximate Bayesian inference for quantiles. *J. Nonparametr. Statist.*, **17**, 385–400.

Fitzenberger, B. (1997). Computational aspects of censored quantile regression. In *Proceedings of the 3rd International Conference on Statistical Data Analysis Based on the l_1-Norm and Related Methods*, Ed. Y. Dodge, pp. pp. 171-186. CA:IMS.

Gelfand, A.E. & Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.*, **11**, 289–305.

Geraci, M. & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, **8**, 140–154.

Geraci, M. & Bottai, M. (2013). Linear quantile mixed models. *Stat. Comput.*, **24**, 461–479.

Hu, Y., Gramacy, R. & Lian, H. (2013). Bayesian quantile regression for single-index models. *Stat. Comput.*, **23**, 437–454.

Hu, Y., Zhao, K. & Lian, H. (2014). Bayesian quantile regression for partially linear additive models. *Stat. Comput.*, **25**, 651–668.

Khare, K. & Hobert, J.P. (2012). Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression. *J. Multivar. Anal.*, **112**, 108–116.

Knight, K. (1998). Limiting distribution for l_1 regression estimators under general conditions. *Ann. Statist.*, **26**, 755–770.

Kobayashi, G. & Kozumi, H. (2012). Bayesian analysis of quantile regression for censored dynamic panel data. *Comput. Stat.*, **27**, 359–380.

Koenker, R. (2004). Quantile regression for longitudinal data. *J. Multivar. Anal.*, **91**, 74–89.

Koenker, R. (2005). *Quantile Regression.* New York: Cambridge University Press.

Koenker, R. (2015). quantreg: quantile regression. R package version 5.11.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.

Kottas, A. & Krnjajić, M. (2009). Bayesian semiparametric modelling in quantile regression. *Scand. J. Stat.*, **36**, 297–319.

Kotz, S., Kozubowski, T. & Podgorski, K. (2001). *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance.* Boston: Birkhuser.

Kozumi, H. & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *J. Statist. Comput. Simulation.*, **81**, 1565–1578.

Lancaster, T. & Jun, S.J. (2010). Bayesian quantile regression methods. *J. Appl. Econom.*, **25**, 287–307.

Lee, D. & Neocleous, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Statist. Soc. C*, **59**, 905–920.

Li, Q., Xi, R. & Lin, N. (2010). Bayesian regularized quantile regression. *Bayesian Anal.*, **5**, 533–556.

Li, Y. & Zhu, J. (2008). l_1-norm quantile regression. *J. Comp. Graph. Stat.*, **17**, 163–185.

Lin, G., He, X. & Portnoy, S.L. (2012). Quantile regression with doubly censored data. *Comput. Stat. Data Anal.*, **56**, 798–812.

Lum, K. & Gelfand, A.E. (2012). Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Anal.*, **7**, 235–258.

Luo, Y., Lian, H. & Tian, M. (2014). Bayesian quantile regression for longitudinal data models. *J. Statist. Comput. Simulation*, **82**, 1635–1649.

Miguéis, V.L., Benoit, D.F. & Poel, D.V.D. (2013). Enhanced decision support in credit scoring using Bayesian binary quantile regression. *J. Oper. Res. Soc.*, **64**, 1374–1383.

Mroz, T.A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, **55**, 765–799.

Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *J. Econometrics*, **142**, 508–538.

Peng, L. & Huang, Y. (2008). Survival analysis with quantile regression models. *J. Am. Statist. Assoc.*, **103**, 637–649.

Portnoy, S. (2003). Censored quantile regression. *J. Am. Statist. Assoc.*, **98**, 1001–1012.

Portnoy, S. (2010). Is ignorance bliss: fixed vs. random censoring. *Inst Math Stat Collect*, **7**, 215–223.

Powell, J.L. (1986). Censored regression quantiles. *J. Econometrics*, **32**, 143–155.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Reich, B.J. (2012). Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *J. R. Statist. Soc. C*, **61**, 535–553.

Reich, B.J., Bondell, H.D. & Wang, H.J. (2010). Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics*, **11**, 337–352.

Reich, B.J., Fuentesa, M. & Dunsona, D.B. (2011). Bayesian spatial quantile regression. *J. Am. Statist. Assoc.*, **106**, 6–20.

Reich, B.J. & Smith, L.B. (2013). Bayesian quantile regression for censored data. *Biometrics*, **69**, 651–660.

Sriram, K., Ramamoorthi, R.V. & Ghosh, P. (2013). Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density. *Bayesian Anal.*, **8**, 489–504.

Tang, Y., Wang, H.J., He, X. & Zhu, Z. (2012). An informative subset-based estimator for censored quantile regression. *Test*, **21**, 635–655.

Thompson, P., Cai, Y., Moyeed, R., Reeve, D. & Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Comput. Stat. Data Anal.*, **54**, 1138–1150.

Tsionas, E.G. (2003). Bayesian quantile inference. *J. Stat. Comput. Simul.*, **73**, 659–674.

Waldmann, E., Kneib, T., Yue, Y.R., Lang, S. & Flexeder, C. (2013). Bayesian semiparametric additive quantile regression. *Statistical Model.*, **13**, 223–252.

Wang, H.J. & Wang, L. (2009). Locally weighted censored quantile regression. *J. Am. Statist. Assoc.*, **104**, 1117–1128.

Wang, J. (2012). Bayesian quantile regression for parametric nonlinear mixed effects models. *Stat. Methods. Appl.*, **21**, 1635–1649.

Wang, L. (2008). Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canad. J. Statist.*, **36**, 321–336.

Womersley, R.S. (1986). Censored discrete linear $l_1 $ approximation. *SIAM J. Sci. Statist. Comput.*, **7**, 105–122.
Yang, Y. & He, X. (2012). Bayesian empirical likelihood for quantile regression. *Ann. Statist.*, **40**, 1102–1131.
Ying, Z., Jung, S.H. & Wei, L.J. (1995). Survival analysis with median regression models. *J. Am. Statist. Assoc.*, **90**, 178–184.
Yu, K. & Moyeed, R.A. (2001). Bayesian quantile regression. *Stat. Probabil. Lett*, **54**, 437–447.
Yu, K. & Stander, J. (2007). Bayesian analysis of a Tobit quantile regression model. *J. Econometrics*, **137**, 260–276.
Yuan, Y. & Yin, G. (2009). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, **66**, 105–114.
Yue, Y.R. & Hong, H.G. (2014). Bayesian Tobit quantile regression model for medical expenditure panel survey data. *Statistical Model.*, **12**, 323–346.
Yue, Y.R. & Rue, H. (2011). Bayesian inference for additive mixed quantile regression models. *Comput. Stat. Data Anal.*, **55**, 84–96.
Zhao, K. & Lian, H. (2015). Bayesian Tobit quantile regression with single-index models. *J. Statist. Comput. Simulation*, **85**, 1247–1263.
Zhu, H., Zeng, H. & Yu, K. (2013). Bayesian inference for non-parametric quantile regression using a Fourier representation. *IJACT*, **5**, 632–641.
Zou, H. & Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.*, **36**, 1108–1126.

[*Received March 2015, accepted June 2015*]

Appendix

We provide the proofs for the quadratic expansions in (3.1) and (3.5). With these expansions, the validity of the posterior inference with the proposed variance correction can be shown by following Theorem 4 of Chernozhukov & Hong (2003).

A.1 Proof for (3.1)

The following assumptions are made.

A1. The conditional distribution $F_{Y_i}(\cdot|\mathbf{x}_i)$ is absolutely continuous with continuous densities $f_{Y_i}(\cdot|\mathbf{x}_i)$ uniformly bounded away from 0 and ∞ at the points $Q_\tau(Y|\mathbf{x}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}_0$, for $i = 1, \ldots, n$.

A2. There exist positive definite matrices D_0 and D_1 such that

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} \mathbf{x}\mathbf{x}_i^\top = D_0, \quad \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} f_{Y_i}\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0|\mathbf{x}_i\right)\mathbf{x}_i\mathbf{x}_i^\top = D_1.$$

In addition, $\max_{i=1,\ldots,n} n^{-1/2}||\mathbf{x}_i|| \to 0$.

Define $\boldsymbol{\delta} = n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, and

$$Z_n(\boldsymbol{\delta}) = \sum_{i=1}^{n}\left\{\rho_\tau\left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}\right) - \rho_\tau\left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_0\right)\right\},$$

Following Knight (1998), we have

$$Z_n(\boldsymbol{\delta}) = Z_{1n}(\boldsymbol{\delta}) + Z_{2n}(\boldsymbol{\delta}),$$

$$\text{where}\quad Z_{1n}(\boldsymbol{\delta}) = -n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i^\top\boldsymbol{\delta}\phi_\tau\left(y_i - \mathbf{x}_i^\top\boldsymbol{\beta}_0\right),$$

$$\text{and } Z_{2n}(\boldsymbol{\delta}) = \sum_{i=1}^{n} Z_{2ni}(\boldsymbol{\delta}) \text{ with } Z_{2ni}(\boldsymbol{\delta}) = \int_0^{v_{ni}} \left(1_{\{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0 \le s\}} - 1_{\{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0 \le 0\}}\right) ds,$$

and $v_{ni} = n^{-1/2} x_i^\top \boldsymbol{\delta}$. For $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| = O(n^{-1/2})$, from the Bahadur representation in Chapter 4.3 of Koenker (2005), that is,

$$n^{1/2}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0) = D_1^{-1} n^{-1/2} \sum_{i=1}^{n} \mathbf{x}_i \phi_\tau \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0\right) + O\left(n^{-1/4} (\log\log n)^{3/4}\right),$$

we have $Z_{1n}(\boldsymbol{\delta}) = -n^{1/2}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0)^\top D_1 \boldsymbol{\delta} + o_p(1)$.

For $Z_{2n}(\boldsymbol{\delta})$, following the proof of Theorem 4.1 in Koenker (2005), we first have

$$E(Z_{2n}(\boldsymbol{\delta})|\mathbf{x}) = (2n)^{-1} \sum_{i=1}^{n} f_{Y_i}\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0 | \mathbf{x}_i\right) \boldsymbol{\delta}^\top \mathbf{x}_i \mathbf{x}_i^\top \boldsymbol{\delta} + o_p(1) = \frac{1}{2} \boldsymbol{\delta}^\top D_1 \boldsymbol{\delta} + o_p(1).$$

Because

$$var(Z_{2ni}(\boldsymbol{\delta})|\mathbf{x}_i) \le E\left\{Z_{2ni}(\boldsymbol{\delta})^2\right\} \le \max_{1\le i\le n} \{Z_{2ni}(\boldsymbol{\delta})\} E\{|Z_{2ni}(\boldsymbol{\delta})|\} \le \max_{1\le i\le n} \{v_{ni}\} E\{Z_{2ni}(\boldsymbol{\delta})\},$$

we have

$$var(Z_{2n}(\boldsymbol{\delta})|\mathbf{x}) \le n^{-1/2} \max_{1\le i\le n} \{|\mathbf{x}_i \boldsymbol{\delta}|\} E(Z_{2n}(\boldsymbol{\delta})|\mathbf{x}) \to 0$$

with $||\boldsymbol{\delta}|| = O(1)$ under condition A2. Therefore, we have

$$Z_{2n}(\boldsymbol{\delta}) = \frac{1}{2} \boldsymbol{\delta}^\top D_1 \boldsymbol{\delta} + o_p(1),$$

and consequently,

$$\begin{aligned} Z_n(\boldsymbol{\delta}) &= -n^{1/2} \left\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0\right\}^\top D_1 \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^\top D_1 \boldsymbol{\delta} + o_p(1) \\ &= \frac{n}{2} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\tau))^\top D_1 (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\tau)) + C_n + o_p(1), \end{aligned}$$

where $C_n = -\frac{n}{2}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0)^\top D_1 (\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0)$. Because of $\sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) = Z_n(\boldsymbol{\delta}) + \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0)$, we have proved (3.1).

5.1 Proof for (3.5)

Define $\boldsymbol{\delta} = n^{1/2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, $u_i = y_i - \max\{x_i^\top \boldsymbol{\beta}_0, 0\}$, $v_{ni} = \max\left(\mathbf{x}_i^\top \boldsymbol{\beta}, 0\right) - \max(\mathbf{x}_i^\top \boldsymbol{\beta}_0, 0)$, and

$$\begin{aligned} Z_n(\boldsymbol{\delta}) &= \sum_{i=1}^{n} \left\{\rho_\tau\left(y_i - \max\left\{\mathbf{x}_i^\top \boldsymbol{\beta}, 0\right\}\right) - \rho_\tau\left(y_i - \max\left\{\mathbf{x}_i^\top \boldsymbol{\beta}_0, 0\right\}\right)\right\} \\ &= \sum_{i=1}^{n} \left\{\rho_\tau(u_i - v_{ni}) - \rho_\tau(u_i)\right\}. \end{aligned}$$

Following Knight (1998), we have $Z_n(\boldsymbol{\delta}) = Z_{1n}(\boldsymbol{\delta}) + Z_{2n}(\boldsymbol{\delta})$, where

$$Z_{1n}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} v_{ni}\phi_\tau(u_i), \text{ and } Z_{2n}(\boldsymbol{\delta}) = \sum_{i=1}^{n}\int_0^{v_{ni}} [I(u_i \le s) - I(u_i \le 0)]\, ds.$$

By the definition of v_{ni}, we have

$$Z_{1n}(\boldsymbol{\delta}) = -n^{-1/2}\left\{\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0) \right.$$
$$\left. - \sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\beta}\phi_\tau(u_i) I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 \le 0) + \sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\beta}_0\phi_\tau(u_i) I(\mathbf{x}_i^\top \boldsymbol{\beta} \le 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0)\right\}.$$

With bounded $\mathbf{x}_i$, and $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| = O(n^{-1/2})$, we have

$$||n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i)\left[I\left(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) - I\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right)\right]||$$
$$\le ||n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I\left(\mathbf{x}_i^\top \boldsymbol{\beta} \le 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right)||.$$

Because of Assumption R.2. in Powell (1986), $E\left[I\left(\mathbf{x}_i^\top \boldsymbol{\beta} \le 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right)\right]$ can be controlled by $O(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||)$. Noting that $E\left(\phi_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0) I\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right)\right) = 0$, we have

$$||n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I\left(\mathbf{x}_i^\top \boldsymbol{\beta} \le 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right)|| = o_p(1),$$

which leads to

$$-n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I\left(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) = -n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) + o_p(1).$$

The last two terms in $Z_{1n}(\boldsymbol{\delta})$ can be shown of order $o_p(1)$ noting that when $\mathbf{x}_i^\top \boldsymbol{\beta}$ and $\mathbf{x}_i^\top \boldsymbol{\beta}_0$ take different signs, their magnitudes are both controlled by $||n^{-1/2}\mathbf{x}_i^\top \boldsymbol{\delta}||$.

Therefore, we have

$$Z_{1n}(\boldsymbol{\delta}) = -n^{-1/2}\sum_{i=1}^{n} \mathbf{x}_i^\top \boldsymbol{\delta}\phi_\tau(u_i) I(\mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0) + o_p(1).$$

From (3.5) in Powell (1986),

$$Z_{1n}(\boldsymbol{\delta}) = n^{1/2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}(\tau))^\top D_1 \boldsymbol{\delta} + o_p(1). \tag{A.1}$$

In addition, we can partition $Z_{2n}(\boldsymbol{\delta})$ as

$$\begin{aligned}Z_{2n}(\boldsymbol{\delta}) &= \sum_{i=1}^{n} I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0) \int_0^{n^{-1/2}\mathbf{x}_i^\top \boldsymbol{\delta}} [I(u_i \leq s) - I(u_i \leq 0)]\, ds \\ &+ \sum_{i=1}^{n} I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 \leq 0) \int_0^{\mathbf{x}_i^\top \boldsymbol{\beta}} [I(u_i \leq s) - I(u_i \leq 0)]\, ds \\ &+ \sum_{i=1}^{n} I(\mathbf{x}_i^\top \boldsymbol{\beta} \leq 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0) \int_0^{-\mathbf{x}_i^\top \boldsymbol{\beta}_0} [I(u_i \leq s) - I(u_i \leq 0)]\, ds.\end{aligned}$$

Following the same steps as in Section 5, we have

$$\begin{aligned}&E\left\{\sum_{i=1}^{n} I\left(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) \int_0^{n^{-1/2}\mathbf{x}_i^\top \boldsymbol{\delta}} [I(u_i \leq s) - I(u_i \leq 0)]\, ds\right\} \\ &= \frac{1}{2n}\sum_{i=1}^{n}\left\{I\left(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) f_{Y_i}(u_i|\mathbf{x}_i)\boldsymbol{\delta}^\top \mathbf{x}_i \mathbf{x}_i^\top \boldsymbol{\delta}\right\} + o_p(1) \\ &= \frac{1}{2n}\sum_{i=1}^{n}\left\{I\left(\mathbf{x}_i^\top \boldsymbol{\beta}_0 > 0\right) f_{Y_i}(u_i|\mathbf{x}_i)\boldsymbol{\delta}^\top \mathbf{x}_i \mathbf{x}_i^\top \boldsymbol{\delta}\right\} + o_p(1).\end{aligned}$$

Note that under the Assumption R.2 in Powell (1986), $E\left[I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta} \leq 0)\right]$ and $E\left[I(\mathbf{x}_i^\top \boldsymbol{\beta} > 0, \mathbf{x}_i^\top \boldsymbol{\beta} \leq 0)\right]$ can both be controlled by $O(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||)$, and the expectations of the last two terms in $Z_{2n}(\boldsymbol{\delta})$ are of $o_p(1)$ order. Using similar arguments as in Section 5, we can show that

$$var(Z_{2n}(\boldsymbol{\delta})|\mathbf{x}_i) \leq \max_{i=1,\ldots,n}\{v_{ni}\}E\{Z_{2n}(\boldsymbol{\delta})|\mathbf{x}\} \to 0.$$

Therefore, we have

$$Z_{2n}(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^\top D_1 \boldsymbol{\delta} + o_p(1),$$

which together with (A.1) proves the quadratic expansion in (3.5).