

Article

A New Class of Estimators Based on a General Relative Loss Function

Tao Hu ¹  and Baosheng Liang ^{2,*} ¹ School of Mathematical Sciences, Capital Normal University, Beijing 100048, China; hutao@cnu.edu.cn² Department of Biostatistics, School of Public Health, Peking University, Beijing 100191, China

* Correspondence: liangbs@hsc.pku.edu.cn; Tel.: +86-010-8280-5541

† Current address: Room 617, Yifu Building, Xueyuan Rd., Haidian District, Beijing 100191, China.

Abstract: Motivated by the relative loss estimator of the median, we propose a new class of estimators for linear quantile models using a general relative loss function defined by the Box–Cox transformation function. The proposed method is very flexible. It includes a traditional quantile regression and median regression under the relative loss as special cases. Compared to the traditional linear quantile estimator, the proposed estimator has smaller variance and hence is more efficient in making statistical inferences. We show that, in theory, the proposed estimator is consistent and asymptotically normal under appropriate conditions. Extensive simulation studies were conducted, demonstrating good performance of the proposed method. An application of the proposed method in a prostate cancer study is provided.

Keywords: Box–Cox transformation; quantile regression; relative error



Citation: Hu, T.; Liang, B.S. A New Class of Estimators Based on a General Relative Loss Function. *Mathematics* **2021**, *9*, 1138. <https://doi.org/10.3390/math9101138>

Academic Editor: Filipe J. Marques

Received: 31 March 2021

Accepted: 14 May 2021

Published: 18 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In contrast to the mean-based regression that mainly gives an overall quantification for the central covariate effect, quantile regression can directly model a series of quantiles (from lower to higher) of the response variable to deliver a global evaluation of the covariate effect [1,2]. A major advantage of quantile regression is that no assumptions about the distribution of the response are required, which makes it practical, robust and amenable to skewed response distributions [3]. Additionally, quantile regression methods can help to handle the cases of heteroscedasticity [4]. Nowadays, quantile regressions have been widely used in many fields, established numerous methodologies covering linear, nonlinear and longitudinal quantile regressions [5,6], as well as applications in survival analysis [7]. Application studies show that quantile regression allows adjustment for potential confounders and calculation of interaction terms and variable selection, while being more robust to statistical outliers and yielding much more information about the underlying associations [1,3]. However, the computation of quantile regressions is relatively complex and somewhat unique, especially compared with ordinary least squares for mean-based linear (or nonlinear) regressions. Take a 0.5 quantile estimation as an example: the quantile regression minimizes the sum of weighted absolute residuals instead of squared residuals [8]. One drawback of quantile regressions is that the estimation efficiency fluctuates a lot at different quantiles and is relatively low at the tails [9]. A traditional quantile regression is typically based on minimizing a check loss function [1], but often the relative quantile loss could be more relevant than the check loss function and hence might be used to gain more efficiency for inference. So far, to the best of our knowledge, there have been no consistent studies on other type of loss functions, such as the relative error loss for quantile regression.

In many practical applications, the magnitude of the relative error, rather than the absolute error, is of the major concern. In general, the relative error is more relevant when the range of predicted values is large and that of predictors is small. Narula and Wellington

(1977) proposed an estimation approach for linear models by minimizing the sum of relative errors [10], without any theoretical results. Khoshgoftaar et al. (1992) studied the asymptotic properties of the estimators by minimizing both the squared relative loss and the absolute relative loss under nonlinear regression models [11], and made a great comparative study on them. Later, Chen et al. (2010) applied the least absolute relative error loss [12] for the linear regression model

$$Y_i^* = X_i^T \beta + \varepsilon_i^*,$$

and proposed to estimate the model parameters by minimizing the sum of the absolute relative error

$$\sum_{i=1}^n \left\{ \left| \frac{Y_i - \exp(X_i^T \beta)}{Y_i} \right| + \left| \frac{Y_i - \exp(X_i^T \beta)}{\exp(X_i^T \beta)} \right| \right\},$$

where $Y_i = \log(Y_i^*)$ and they assumed $\log(\varepsilon_i^*)$ has a mean of 1. Motivated by [12], Yang and Ye (2013) established the connection between relative error estimators and the M-estimation under a linear model [13]. However, their works only consider the absolute relative loss, not related to any quantile estimates of model distribution. None of these studies discussed the way to apply the relative error for a general quantile regression.

In this article, we propose a general class of relative loss functions via a Box–Cox transformation [14,15]. The proposed loss function includes the absolute loss and the relative loss as special cases. We also show that the proposed loss function is convex, scale-free and able to be elicited. We apply the proposed loss function for a linear quantile regression [1] and prove that the estimates of the regression coefficients are consistent and asymptotically normal. Through numerical studies on two concrete examples that have well-derived theoretical solutions, we show that the proposed method is feasible and verify that the numerical results have the expected theoretical properties obtained from the theoretical study. We also apply the proposed method to a prostate cancer study, showing that our method provides more accurate statistical inferences on quantile estimates compared to the regular quantile regression, especially at the region of tail quantiles.

The rest of this article is organized as follows. In Section 2, we introduce the model and propose the estimation procedure. In Section 3, we establish the consistency and asymptotic normality for the parameter estimates under certain regularity conditions. In Section 4, we examine the finite-sample properties using two simple simulations. An application example is given in Section 5 with a prostate cancer study. We conclude with a brief summary and remarks in Section 6.

2. Model and Methods

Let Y_i be the response of interest and X_i be a $(p + 1)$ -dimensional covariate with the first element being 1. Consider a linear quantile regression model

$$Y_i = X_i^T \beta_0(\tau) + \varepsilon_i(\tau), \quad i = 1, \dots, n, \tag{1}$$

where $\beta_0(\tau)$ is a $(p + 1)$ -dimensional coefficient for some $\tau \in (0, 1)$, and $\varepsilon_i(\tau)$ is a random error with the τ th quantile being equal to 0 conditional on X_i . In model (1), Y_i could also be replaced by any other reasonable monotone transformation, but considering that a linear relationship in the transformed model may not be linear in the original scale, one may need to transform the result back to the original measurement scale for the interpretation.

Through an exponential transformation, model (1) can be rewritten as

$$T_i = \exp\{X_i^T \beta_0(\tau)\} \varepsilon_i(\tau), \quad i = 1, \dots, n, \tag{2}$$

where $T_i = \exp(Y_i) > 0$, $\varepsilon_i(\tau) = \exp\{\varepsilon_i(\tau)\} = \exp\{Y_i - X_i^T \beta_0(\tau)\}$. Denote $H(\cdot)$ as a monotone function of real value, then according to the equivariance property of quantiles, $Q_\tau\{H(\varepsilon_i)|X_i\} = H\{Q_\tau(\varepsilon_i|X_i)\}$ holds for any random error ε_i . Here, $Q_\tau(\varepsilon|X)$ is the τ th conditional quantile of ε given X . This characteristic enables us to consider quantile

regressions under a proper transformation of the random error so as to enhance the estimation efficiency.

To this end, we propose to conduct the model regression using a general class of relative loss functions, which leads to the objective function in the form

$$\frac{1}{n} \sum_{i=1}^n \left\{ \rho_{\tau} [H_{\gamma} \{T_i \exp(-X_i^T \beta)\}] + \rho_{\tau} [H_{-\gamma} \{T_i \exp(-X_i^T \beta)\}] \right\}, \tag{3}$$

where $\rho_{\tau}(u) = u\{\tau - I(u < 0)\}$ is the traditional quantile loss function and $H_{\gamma}(\cdot)$ takes the Box–Cox transformation with $H_{\gamma}(u) = (u^{\gamma} - 1)/\gamma$ for $\gamma \neq 0$ and $H_{\gamma}(u) = \log(u)$ for $\gamma = 0$ (see Figure 1).

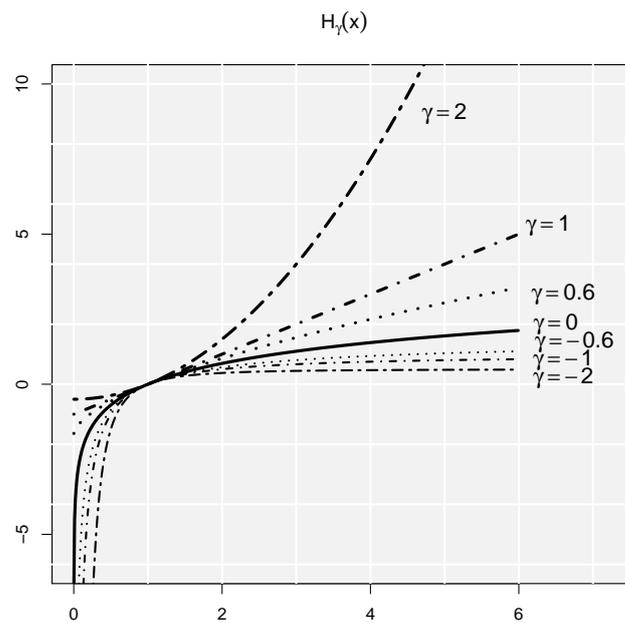


Figure 1. Plots of $H_{\gamma}(x)$ at different values of γ .

The objective function, (3), can be further simplified as

$$W_n(\beta; \gamma, \tau) = \frac{1}{n} \sum_{i=1}^n V_{\gamma} \left(\exp(Y_i - X_i^T \beta) \right) \left\{ \tau - I(Y_i < X_i^T \beta) \right\}, \tag{4}$$

where $V_{\gamma}(u) = (u^{\gamma} - u^{-\gamma})/\gamma$ if $\gamma \neq 0$, otherwise $V_{\gamma}(u) = 2 \log(u)$ is a general relative loss function. In particular, if $\gamma = 0$, (4) is reduced to the objective function of traditional quantile regression, and if $\gamma = 1$ and $\tau = 0.5$, (4) is reduced to

$$\frac{1}{n} \sum_{i=1}^n \left\{ \left| \frac{T_i - \exp(X_i^T \beta)}{T_i} \right| + \left| \frac{T_i - \exp(X_i^T \beta)}{\exp(X_i^T \beta)} \right| \right\},$$

which is exactly the objective function of the least absolute relative error in [12].

The proposed framework is very flexible—it allows us to adapt the quantile regression to either the absolute or relative loss or somewhere in between by tuning the parameter γ . Furthermore, the function $V_{\gamma}(u)$ in (4) can guarantee the proposed criterion function to be convex (see Lemma A1), scale-free, and able to be elicited (see Definition 2 in [16]). Therefore, given γ , the minimizer of $W_n(\beta; \gamma, \tau)$ with respect to β , denoted as $\hat{\beta}_n(\gamma; \tau)$, can be obtained conveniently using classical algorithms, such as the Nelder–Mead simplex method recommended by Yin and Cai [17].

3. Asymptotic Properties

3.1. Conditions and Main Results

We assumed that $(Y_i, X_i, \epsilon_i(\tau)), i = 1, \dots, n$, are independent. Let F_ϵ and F_ε be the cumulative distribution functions of ϵ and ε , and $F(\cdot|X)$ be the conditional distribution function of ε given X . We considered a random design for X and used $f(\cdot|X)$ to denote the corresponding density function that is almost definitely continuous in the neighborhood of 1. To establish the asymptotic properties of the estimators under a certain metric, we imposed regularity conditions as follows:

- (C1) Covariate X is bounded and does not concentrate on any hyperplane of p dimensions.
- (C2) For any fixed γ , it holds that $E\{(\varepsilon^\gamma + \varepsilon^{-\gamma})^2|X\} < \infty$ and $E[(\varepsilon^\gamma + \varepsilon^{-\gamma})\{\tau - I(\varepsilon < 1)\}|X] = 0$.
- (C3) If $\beta(\gamma; \tau) = \beta(\gamma^*; \tau)$ for any fixed τ , then $\gamma(\tau) = \gamma^*(\tau)$.
- (C4) $D = E(XX^T)$ is a positive definite.

Condition (C1) is regular, and condition (C2) ensures the consistency and asymptotic normality, which can be treated as a generalized version of the zero mean and zero median assumptions for the least square estimation and least absolute deviation regression methods, respectively. Condition (C3) guarantees the identifiability of the adaptive parameter and regression parameters. Condition (C4) is to ensure the asymptotic normality of the estimates of the regression coefficients, similar to the finite second moment condition for the least square estimator in linear regressions.

Theorem 1. Under conditions (C1)–(C4), for any $\tau \in (0, 1)$ and any finite $\gamma \in [0, \Gamma]$, Γ is a constant

$$\|\hat{\beta}_n(\gamma; \tau) - \beta_0(\tau)\| \rightarrow 0$$

in probability as $n \rightarrow \infty$, where $\|\cdot\|$ indicates the Euclidean norm.

Proof of Theorem 1. By the results of Lemma A1 in Appendix A, we know that $W_n(\beta; \gamma, \tau)$ is a convex function with respect to β . According to the Convexity Lemma [18], it holds that

$$\sup_{\beta \in \mathcal{B}} |W_n(\beta; \gamma, \tau) - E\{W_n(\beta; \gamma, \tau)\}| \rightarrow 0 \tag{5}$$

with the probability as $n \rightarrow +\infty$ and the function $E\{W_n(\beta; \gamma, \tau)\}$ is necessarily convex on \mathcal{B} . Through an algebraic calculation, we obtain

$$\begin{aligned} & E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} \\ &= \frac{1}{n} \sum_{i=1}^n E\left\{\rho_\tau\left(\frac{T_i^\gamma \exp(-\gamma X_i^T \beta) - 1}{\gamma}\right) + \rho_\tau\left(\frac{T_i^{-\gamma} \exp(\gamma X_i^T \beta) - 1}{-\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^\gamma - 1}{\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^{-\gamma} - 1}{-\gamma}\right)\right\} \\ &= \frac{1}{n} \sum_{i=1}^n E\left[\rho_\tau\left(\frac{T_i^\gamma \exp(-\gamma X_i^T \beta) - 1}{\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^\gamma - 1}{\gamma}\right)\right] \\ & \quad + \frac{1}{n} \sum_{i=1}^n E\left[\rho_\tau\left(\frac{T_i^{-\gamma} \exp(\gamma X_i^T \beta) - 1}{-\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^{-\gamma} - 1}{-\gamma}\right)\right]. \end{aligned} \tag{6}$$

According to the Knight’s identity [19], it holds that

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\{\tau - I(u < 0)\} + \int_0^v \{I(u \leq s) - I(u \leq 0)\} ds,$$

Then, by using the Knight’s identity we obtain

$$\begin{aligned} & \rho_\tau\left(\frac{T_i^\gamma \exp(-\gamma X_i^T \beta) - 1}{\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^\gamma - 1}{\gamma}\right) \\ &= \frac{\varepsilon_i^\gamma [\exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\} - 1]}{\gamma} \{\tau - I(\varepsilon_i < 1)\} \\ & \quad + \int_0^{\varepsilon_i^\gamma [1 - \exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\}]/\gamma} \{I(H_\gamma(\varepsilon_i) \leq s) - I(H_\gamma(\varepsilon_i) \leq 0)\} ds, \end{aligned}$$

and similarly we obtain

$$\begin{aligned} & \rho_\tau\left(\frac{T_i^{-\gamma} \exp(\gamma X_i^T \beta) - 1}{-\gamma}\right) - \rho_\tau\left(\frac{\varepsilon_i^{-\gamma} - 1}{-\gamma}\right) \\ &= \frac{\varepsilon_i^{-\gamma} [\exp\{\gamma X_i^T (\beta - \beta_0(\tau))\} - 1]}{-\gamma} \{\tau - I(\varepsilon_i < 1)\} \\ & \quad + \int_0^{\varepsilon_i^{-\gamma} [1 - \exp(\gamma X_i^T \{\beta - \beta_0(\tau)\})]/(-\gamma)} \{I(H_{-\gamma}(\varepsilon_i) \leq s) - I(H_{-\gamma}(\varepsilon_i) \leq 0)\} ds. \end{aligned}$$

Therefore,

$$\begin{aligned} & E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ E\left[-\frac{\varepsilon_i^\gamma + \varepsilon_i^{-\gamma}}{\gamma} \{\tau - I(\varepsilon_i < 1)\} \left(\exp\{\gamma X_i^T (\beta - \beta_0(\tau))\} - 1 \right) \right] \right. \\ & \quad + E\left[\frac{\varepsilon_i^\gamma}{\gamma} \{\tau - I(\varepsilon_i < 1)\} \left(\exp\{\gamma X_i^T (\beta - \beta_0(\tau))\} + \exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\} - 2 \right) \right] \\ & \quad + E\left[\int_0^{\varepsilon_i^\gamma [1 - \exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\}]/\gamma} \{I(H_\gamma(\varepsilon_i) \leq s) - I(H_\gamma(\varepsilon_i) \leq 0)\} ds \right. \\ & \quad \left. \left. + \int_0^{\varepsilon_i^{-\gamma} [1 - \exp\{\gamma X_i^T (\beta - \beta_0(\tau))\}]/(-\gamma)} \{I(H_{-\gamma}(\varepsilon_i) \leq s) - I(H_{-\gamma}(\varepsilon_i) \leq 0)\} ds \right] \right\}. \end{aligned} \tag{7}$$

The first term in the summand of right-hand side is equal to 0 by condition (C2). By condition (C2) and the fact that

$$\begin{aligned} 2E\{\varepsilon^\gamma I(\varepsilon^\gamma \geq 1) | X\} &> E\{(\varepsilon^\gamma + \varepsilon^{-\gamma}) I(\varepsilon^\gamma \geq 1) | X\} \\ &= \frac{\tau}{1 - \tau} E\{(\varepsilon^\gamma + \varepsilon^{-\gamma}) I(\varepsilon^\gamma < 1) | X\} \\ &> \frac{2\tau}{1 - \tau} E\{\varepsilon^\gamma I(\varepsilon^\gamma < 1) | X\}, \end{aligned} \tag{8}$$

which implies $E\{\varepsilon^\gamma [\tau - I(\varepsilon < 1)] | X\} > 0$. We know the second term in (7) is non-negative. The third term in (7) is equivalent to

$$\begin{aligned} & \sum_{i=1}^n E\left[\int_0^{[1 - \exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\}]/\gamma} \varepsilon_i^\gamma \{I(\varepsilon_i^\gamma \leq (1 - \gamma s)^{-1}) - I(\varepsilon_i^\gamma \leq 1)\} ds \right. \\ & \quad \left. + \int_0^{[1 - \exp\{\gamma X_i^T (\beta - \beta_0(\tau))\}]/(-\gamma)} \varepsilon_i^{-\gamma} \{I(\varepsilon_i^\gamma \leq 1 + \gamma s) - I(\varepsilon_i^\gamma \leq 1)\} ds \right]. \end{aligned} \tag{9}$$

As the upper limits of both $[1 - \exp\{-\gamma X_i^T (\beta - \beta_0(\tau))\}]/\gamma$ and $[1 - \exp\{\gamma X_i^T (\beta - \beta_0(\tau))\}]/(-\gamma)$ always have same signs, plus the fact that the integrated functions $I(\varepsilon_i^\gamma \leq (1 - \gamma s)^{-1}) - I(\varepsilon_i^\gamma \leq 1)$ and $I(\varepsilon_i^\gamma \leq 1 + \gamma s) - I(\varepsilon_i^\gamma \leq 1)$ are all monotone functions with respect to s and the signs of the integrated functions are also consistent with those

of the upper limits of the integrations, we know that (9) is also non-negative. Thus, $E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} = 0$ ensures

$$E\left\{\varepsilon^\gamma[\tau - I(\varepsilon < 1)]\left(\exp\{\gamma X_i^T(\beta - \beta_0(\tau))\} + \exp\{-\gamma X_i^T(\beta - \beta_0(\tau))\} - 2\right)\right\} = 0.$$

As $\beta = \beta_0(\tau)$ is the unique minimizer of $\exp\{\gamma X_i^T(\beta - \beta_0(\tau))\} + \exp\{-\gamma X_i^T(\beta - \beta_0(\tau))\} - 2$, it follows from conditions (C1), (C3) and $E\{\varepsilon^\gamma[\tau - I(\varepsilon < 1)]|X\} > 0$ that $\beta = \beta_0(\tau)$ is the unique minimizer of $E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\}$.

When $Q(\beta; \gamma, \tau) = E\{W_n(\beta; \gamma, \tau)\}$, then for every $\delta > 0, \eta > 0$, such that $Q(\beta; \gamma, \tau) > Q(\beta_0(\tau); \gamma, \tau) + \eta$ for $\|\beta - \beta_0(\tau)\| \geq \delta$. For any $\delta > 0$ and a constant $D > 0$, suppose the minimizer of $W_n(\beta; \gamma, \tau)$, i.e. $\hat{\beta}_n(\gamma; \tau)$, is achieved in $\{\beta : \delta \leq \|\beta - \beta_0\| \leq D\} \subset \mathcal{B}$. Following (5), we know that $W_n(\hat{\beta}_n(\gamma; \tau); \gamma, \tau) \rightarrow Q(\hat{\beta}_n(\gamma; \tau); \gamma, \tau)$ in probability as $n \rightarrow \infty$, and $Q(\hat{\beta}_n(\gamma; \tau); \gamma, \tau) > Q(\beta_0(\tau); \gamma, \tau) + \eta$. So $\lim_{n \rightarrow \infty} \inf_{\|\beta - \beta_0(\tau)\| \geq \delta} W_n(\beta; \gamma, \tau) \geq Q(\hat{\beta}_n(\gamma; \tau); \gamma, \tau) > Q(\beta_0(\tau); \gamma, \tau) + \eta$ holds in probability for any constant $\delta > 0$. This is contradictory to (5). Hence, the minimum of $W_n(\beta; \gamma, \tau)$ can only be achieved in $\|\beta - \beta_0(\tau)\| < \delta$. By the randomness of δ , we know $\hat{\beta}_n(\gamma; \tau) \rightarrow \beta_0(\tau)$ in probability as $n \rightarrow \infty$. \square

Theorem 2. Under conditions (C1)–(C4), for any $\tau \in (0, 1)$ and any finite $\gamma \in [0, \Gamma]$,

$$\sqrt{n}\{\hat{\beta}_n(\gamma; \tau) - \beta_0(\tau)\} \xrightarrow{d} N\left(0, [\gamma\{J_\gamma(\tau) + f_\varepsilon(1)\}]^{-2}B(\gamma; \tau)D^{-1}\right),$$

where $J_\gamma(\tau) = E[\varepsilon^\gamma\{\tau - I(\varepsilon < 1)\}|X]$, $B(\gamma; \tau) = E[(\varepsilon^\gamma + \varepsilon^{-\gamma})\{\tau - I(\varepsilon < 1)\}|X]^2$ and $D = E(XX^T)$. In particular, if ε follows a symmetric distribution with the symmetrical axis $\varepsilon = 1$, then $B(\gamma; \tau = 1/2) = E\{(\varepsilon^\gamma + \varepsilon^{-\gamma})^2|X\}/4$.

Proof of Theorem 2. To prove the asymptotic normality, we approximate $E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\}$ for every β in a neighborhood of $\beta_0(\tau)$ first. By the proof in Theorem 1, we know

$$E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} = \frac{1}{n} \sum_{i=1}^n E\{G_i(\beta)\} + \frac{1}{n} \sum_{i=1}^n E\{R_i(\beta)\},$$

where $G_i(\beta) = \gamma^{-1}\varepsilon_i^\gamma\{\tau - I(\varepsilon_i < 1)\}(\exp\{\gamma X_i^T(\beta - \beta_0(\tau))\} + \exp\{-\gamma X_i^T(\beta - \beta_0(\tau))\} - 2)$ and

$$R_i(\beta) = \int_0^{\varepsilon_i^\gamma[1 - \exp\{-\gamma X_i^T(\beta - \beta_0(\tau))\}]/\gamma} \{I(H_\gamma(\varepsilon_i) \leq s) - I(H_\gamma(\varepsilon_i) \leq 0)\} ds + \int_0^{\varepsilon_i^{-\gamma}[1 - \exp\{\gamma X_i^T(\beta - \beta_0(\tau))\}]/(-\gamma)} \{I(H_{-\gamma}(\varepsilon_i) \leq s) - I(H_{-\gamma}(\varepsilon_i) \leq 0)\} ds.$$

The item $E\{G_i(\beta)\}$ can be expanded near $\beta_0(\tau)$ as

$$\gamma E[\varepsilon_i^\gamma\{\tau - I(\varepsilon_i < 1)\}](\beta - \beta_0(\tau))^T E(X_i X_i^T)(\beta - \beta_0(\tau)) + O(\|\beta - \beta_0(\tau)\|^3).$$

Note that the item $E\{R_i(\beta)\}$ equals

$$\begin{aligned} & E_X \left[\int_0^{[1-\exp\{-\gamma X_i^T(\beta-\beta_0(\tau))\}]/\gamma} \left\{ \int_0^{(1-\gamma s)^{-1}} x f_\varepsilon(x) dx - \int_0^1 x f_\varepsilon(x) dx \right\} ds \right] \\ & + E_X \left[\int_0^{[1-\exp\{\gamma X_i^T(\beta-\beta_0(\tau))\}]/(-\gamma)} \left\{ \int_0^{1+\gamma s} x^{-1} f_\varepsilon(x) dx - \int_0^1 x^{-1} f_\varepsilon(x) dx \right\} ds \right] \\ = & E_X \left[\int_0^{[1-\exp\{-\gamma X_i^T(\beta-\beta_0(\tau))\}]/\gamma} \left\{ \gamma f_\varepsilon(1)s + O(s^2) \right\} ds \right] \\ & + E_X \left[\int_0^{[1-\exp\{\gamma X_i^T(\beta-\beta_0(\tau))\}]/(-\gamma)} \left\{ \gamma f_\varepsilon(1)s + O(s^2) \right\} ds \right] \\ = & E_X \left(\frac{\gamma f_\varepsilon(1)}{2} \frac{[1 - \exp\{-\gamma X_i^T(\beta - \beta_0(\tau))\}]^2}{\gamma^2} \right) \\ & + E_X \left(\frac{\gamma f_\varepsilon(1)}{2} \frac{[1 - \exp\{\gamma X_i^T(\beta - \beta_0(\tau))\}]^2}{\gamma^2} \right) + O(\|\beta - \beta_0(\tau)\|^4) \\ = & \frac{f_\varepsilon(1)}{2\gamma} \left[2\gamma^2(\beta - \beta_0(\tau))^T E(X_i X_i^T)(\beta - \beta_0(\tau)) \right] + O(\|\beta - \beta_0(\tau)\|^4) \\ = & \gamma f_\varepsilon(1)(\beta - \beta_0(\tau))^T E(X_i X_i^T)(\beta - \beta_0(\tau)) + O(\|\beta - \beta_0(\tau)\|^4). \end{aligned}$$

Thus,

$$\begin{aligned} & E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} \\ = & \{\gamma J_\gamma(\tau) + \gamma f_\varepsilon(1)\}(\beta - \beta_0(\tau))^T D(\beta - \beta_0(\tau)) + O(\|\beta - \beta_0(\tau)\|^3), \end{aligned}$$

where $J_\gamma(\tau) = E[\varepsilon_i^\gamma \{\tau - I(\varepsilon_i < 1)\} | X_i]$ and $D = E(X_i X_i^T)$.

Let $V_n = n^{-1} \sum_{i=1}^n (\varepsilon_i^\gamma + \varepsilon_i^{-\gamma}) \{\tau - I(\varepsilon_i < 1)\} X_i$. Observe that

$$\begin{aligned} & W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau) - E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\} \\ = & -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\varepsilon_i^\gamma + \varepsilon_i^{-\gamma}}{\gamma} \{\tau - I(\varepsilon_i < 1)\} \left(\exp\{\gamma X_i^T(\beta - \beta_0(\tau))\} - 1 \right) \right\} \tag{10} \\ & + \frac{1}{n} \sum_{i=1}^n [G_i(\beta) - E\{G_i(\beta)\}] + \frac{1}{n} \sum_{i=1}^n [R_i(\beta) - E\{R_i(\beta)\}] + O(\|\beta - \beta_0(\tau)\|^3), \end{aligned}$$

we next claim that

$$\begin{aligned} & \sup_{\|\beta - \beta_0(\tau)\| \leq Cn^{-1/2}} |W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau) + V_n^T(\beta - \beta_0(\tau)) \\ & \quad - E\{W_n(\beta; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\}| \rightarrow 0 \tag{11} \end{aligned}$$

in probability as $n \rightarrow \infty$. Let $\theta = \sqrt{n}(\beta - \beta_0(\tau))$, then the above equation is equivalent to

$$\begin{aligned} & \sup_{\|\theta\| \leq C} |W_n(\beta_0(\tau) + \theta/\sqrt{n}; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau) + n^{-1/2} V_n^T \theta \\ & \quad - E\{W_n(\beta_0(\tau) + \theta/\sqrt{n}; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau)\}| \rightarrow 0. \tag{12} \end{aligned}$$

Similar to the proof of Lemma B.4. in [20], using the arguments of VC-subgraph classes we can show

$$\sup_{\|\beta - \beta_0(\tau)\| \leq Cn^{-1/2}} \left| \frac{1}{n} \sum_{i=1}^n [G_i(\beta) - E\{G_i(\beta)\}] + \frac{1}{n} \sum_{i=1}^n [R_i(\beta) - E\{R_i(\beta)\}] \right| \rightarrow 0$$

with probability as $n \rightarrow \infty$. Then, according to (10) and the Taylor expansion in the first term, (11) holds.

Let $W_n(\theta; \tau) = -V_n^T \theta / \sqrt{n} + n^{-1} \{ \gamma J_\gamma(\tau) + \gamma f_\varepsilon(1) \} \theta^T D \theta$, then $W_n(\theta; \tau)$ is convex and thus has a unique minimizer $\theta = \sqrt{n} \{ \gamma J_\gamma(\tau) + \gamma f_\varepsilon(1) \}^{-1} D^{-1} V_n$. By (12), we know that $W_n(\beta_0(\tau) + \theta / \sqrt{n}; \gamma, \tau) - W_n(\beta_0(\tau); \gamma, \tau) \rightarrow W_n(\theta; \tau)$ uniformly holds over $\|\theta\| \leq C$. Thus, $\hat{\theta} = \arg \min W_n(\beta_0(\tau) + \theta / \sqrt{n}; \gamma, \tau) \xrightarrow{d} \arg \min W_n(\theta; \tau)$, that is,

$$\sqrt{n}(\hat{\beta}(\gamma; \tau) - \beta_0(\tau)) \xrightarrow{d} \sqrt{n} \{ \gamma J_\gamma(\tau) + \gamma f_\varepsilon(1) \}^{-1} D^{-1} V_n.$$

Hence, $\sqrt{n}(\hat{\beta}(\gamma; \tau) - \beta_0(\tau)) \xrightarrow{d} N(0, \{ \gamma J_\gamma(\tau) + \gamma f_\varepsilon(1) \}^{-2} B(\gamma; \tau) D^{-1})$, where $B(\gamma; \tau) = E[(\varepsilon^\gamma + \varepsilon^{-\gamma})\{\tau - I(\varepsilon < 1)\} | X]^2$ and $D = E(XX^T)$. \square

3.2. Adaptive Criteria for Choosing γ

The loss function defined in (4) provides a flexible framework that allows us to conduct a quantile regression adaptively to practical scenario. However, how to choose the tuning parameter γ for the proposed loss function adaptively to real data is a challenging but important problem. Reasonable criteria for choosing the tuning parameter need to be explored, so with that we could select the optimal γ based on the criterion using data-driven techniques.

In this article, we investigate two criteria as follows:

- (i) Criterion (I) selects the $\gamma(\tau)$ that minimizes the objective function $W_n(\beta, \gamma; \tau)$ for $\gamma \in [0, \Gamma]$ at each given τ ;
- (ii) Criterion (II) selects the $\gamma(\tau)$ that minimizes the variance of $\hat{\beta}_n(\gamma; \tau)$ at each given τ .

According to Lemma A1, we conclude that $\gamma(\tau) = \arg \min_{\gamma \in [0, \Gamma]} W_n(\beta, \gamma; \tau) \equiv 0$. That is, the proposed quantile regression approach is equivalent to the traditional quantile regression under criterion (I). Thus, the optimal $\gamma(\tau)$ selected by criterion (I) provides the best model fitting at each quantile in the sense of regular quantile regression. Criterion (II) selects $\gamma(\tau)$, which reshapes the distribution of residuals for obtaining the best estimation efficiency. The optimal $\gamma(\tau)$ selected by criterion (II) is denoted as $\gamma^*(\tau)$. According to the derived asymptotic covariance matrix in Theorem 2, we know that the value of γ^* depends on τ as well as the distribution of ε . For a fixed τ and a specified distribution of ε , minimizing the variance of $\hat{\beta}_n(\gamma; \tau)$ is equivalent to maximizing $\{J_\gamma(\tau) + f_\varepsilon(1)\}^2 / B(\gamma; \tau)$, and thus

$$\gamma^*(\tau) = \arg \max_{\gamma \in [0, \Gamma]} \{J_\gamma(\tau) + f_\varepsilon(1)\}^2 / B(\gamma; \tau).$$

In the sequel, we mainly discuss the performance of criterion (II).

Remark 1. To estimate the asymptotic variance for a given γ , we recommend to use the wild bootstrap technique. Define

$$W_n^*(\beta; \gamma, \tau) = \frac{1}{n} \sum_{i=1}^n w_i \left[V_\gamma \{ \exp(Y_i - X_i^T \beta) \} \{ \tau - I(Y_i < X_i^T \beta) \} \right], \tag{13}$$

where $\{w_1, \dots, w_n\}$ is a sequence of i.i.d nonnegative random variables with both mean and variance equal to 1. Let $\hat{\beta}_n^*(\gamma; \tau) = \arg \min_\beta W_n^*(\beta; \gamma, \tau) |_{\gamma}$, and then the distribution of $\sqrt{n}\{\hat{\beta}_n(\gamma; \tau) - \beta_0(\tau)\}$ can be approximated by the resampling distribution of $\sqrt{n}\{\hat{\beta}_n^*(\gamma; \tau) - \hat{\beta}_n(\gamma; \tau)\}$, where $\hat{\beta}_n(\gamma; \tau)$ is the minimizer of (4). The standard error of $\hat{\beta}_n(\gamma; \tau)$ can be approximately estimated by the empirical standard error of $\{\hat{\beta}_n^{*(m)}(\gamma; \tau) : m = 1, \dots, M\}$.

Remark 2. People can also define other criteria, such as selecting $\gamma^*(\tau)$ by minimizing the summation of the standardized $W_n(\beta, \gamma; \tau)$ and the standardized variance of $\hat{\beta}_n(\gamma; \tau)$, we denote this criterion as (III), then the selected $\gamma(\tau)$ shall be between 0 and $\gamma^*(\tau)$. In practice, it is hard to assess whether a criterion is better than another one. The criterion that fits practical needs is the best.

To illustrate the feasibility of criterion (II) and show the existence of optimal $\gamma(\tau)$, we next give two concrete examples with specified quantiles and distributions of ε .

Example 1. Assume that ε_i follows the standard log-normal distribution and is independent of X_i , i.e., $\varepsilon_i \sim N(0, 1)$, then $\gamma^*(\tau) = 1.12, 1.13, 1.14$ for $\tau = 0.25, 0.5, 0.75$, respectively.

First, note that

$$B(\gamma; \tau) = E[(\varepsilon^\gamma + \varepsilon^{-\gamma})/\gamma\{\tau - I(\varepsilon < 1)\}]^2 = (\tau - 1)^2 E\{(\varepsilon^\gamma + \varepsilon^{-\gamma})^2/\gamma^2 I(\varepsilon < 1)\} + \tau^2 E\{(\varepsilon^\gamma + \varepsilon^{-\gamma})^2/\gamma^2 I(\varepsilon \geq 1)\}.$$

By the definition,

$$E(\varepsilon^{2\gamma} I(\varepsilon < 1)) = \frac{\exp(2\gamma^2)}{\sqrt{2\pi}} \int_{-\infty}^0 \exp\{-x - 2\gamma^2/2\} dx = \exp(2\gamma^2)\Phi(-2\gamma),$$

$$E(\varepsilon^{-2\gamma} I(\varepsilon < 1)) = \frac{\exp(2\gamma^2)}{\sqrt{2\pi}} \int_{-\infty}^0 \exp\{-(x + 2\gamma)^2/2\} dx = \exp(2\gamma^2)\Phi(2\gamma),$$

hence, $B(\gamma; \tau) = \{\tau^2 + (1 - \tau)^2\}\{\exp(2\gamma^2) + 2\}/\gamma^2$. We can derive

$$E\{\varepsilon^\gamma(\tau - I(\varepsilon < 1))\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(\gamma x)(\tau - I(x < 0)) \exp(-x^2/2) dx = \frac{1}{\sqrt{2\pi}} \left\{ (\tau - 1) \int_{-\infty}^0 \exp(\gamma x) \exp(-x^2/2) dx + \tau \int_0^{+\infty} \exp(\gamma x) \exp(-x^2/2) dx \right\} = \exp(\gamma^2/2)\{(\tau - 1)\Phi(-\gamma) + \tau(1 - \Phi(-\gamma))\} = \exp(\gamma^2/2)\{\tau - \Phi(-\gamma)\}.$$

Thus, we have

$$\{J_\gamma(\tau) + f_\varepsilon(1)\}^{-2} B(\gamma; \tau) = \frac{\{\tau^2 + (1 - \tau)^2\}\{\exp(2\gamma^2) + 2\}}{\gamma^2 [\exp(\gamma^2/2)\{\tau - \Phi(-\gamma)\} + f_\varepsilon(1)]^2}.$$

There is no closed form solution for minimizing $\{J_\gamma(\tau) + f_\varepsilon(1)\}^{-2} B(\gamma; \tau)$, but through a numerical procedure, we find that $\{J_\gamma(\tau) + f_\varepsilon(1)\}^2/B(\gamma; \tau)$ reaches the maximum at 1.12, 1.13, and 1.14, see Figure 2, for $\tau = 0.25, 0.5, 0.75$, respectively. \square

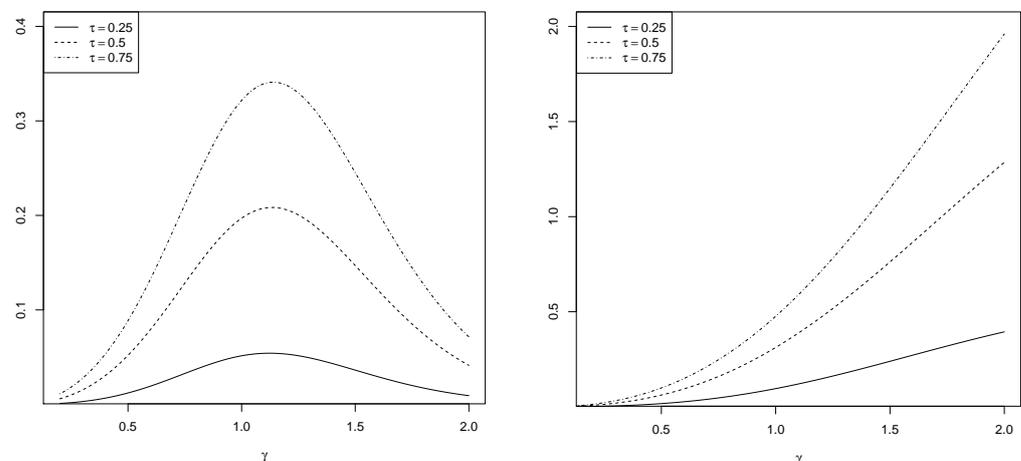


Figure 2. The left and right panels show the plot of $\{\gamma J_\gamma(0.5) + \gamma f_\varepsilon(1)\}^2/B(\gamma; \tau)$ versus γ in Examples 1 and 2, respectively.

Example 2. If $\epsilon_i \sim \text{Unif}(-1, 1)$, then $\gamma^*(\tau) \equiv \Gamma$ for any $\tau \in (0, 1)$. In particular, $\gamma^*(\tau) \equiv 2$ if we set $\Gamma = 2$.

First, by $E\{(\epsilon^\gamma + \epsilon^{-\gamma})^2 I(\epsilon < 1)\} = E\{(\epsilon^\gamma + \epsilon^{-\gamma})^2 I(\epsilon \geq 1)\} = 2 + \{\exp(2\gamma) - \exp(-2\gamma)\} / (4\gamma)$, we know that $B(\gamma; \tau) = \{\tau^2 + (1 - \tau)^2\} [\{\exp(2\gamma) - \exp(-2\gamma)\} / (4\gamma) + 2] / \gamma^2$. It further leads to

$$\begin{aligned} E\{\epsilon^\gamma(\tau - I(\epsilon < 1))\} &= \frac{1}{2} \int_{-1}^1 \exp(\gamma x) \{\tau - I(x < 0)\} dx \\ &= \frac{1}{2\gamma} \left\{ (\tau - 1) \int_{-1}^0 \exp(\gamma x) d(\gamma x) + \tau \int_0^1 \exp(\gamma x) d(\gamma x) \right\} \\ &= \{\tau \exp(\gamma) + (1 - \tau) \exp(-\gamma) - 1\} / (2\gamma), \end{aligned}$$

and then we have

$$\begin{aligned} &\{J_\gamma(\tau) + f_\epsilon(1)\}^{-2} B(\gamma; \tau) \\ &= \frac{\{\tau^2 + (1 - \tau)^2\} \{8\gamma + \exp(2\gamma) - \exp(-2\gamma)\}}{\gamma [\{\tau \exp(\gamma) + (1 - \tau) \exp(-\gamma) - 1\} + 2\gamma f_\epsilon(1)]^2} \end{aligned}$$

As is shown in Figure 2, $\{J_\gamma(\tau) + f_\epsilon(1)\}^{-2} / B(\gamma; \tau)$ reaches the maximum at Γ . \square

4. Simulation Study

For numerical implementation of the proposed method, we first obtained $\hat{\beta}_n(\gamma; \tau)$ by minimizing (4) for each fixed $\gamma \in [0, \Gamma]$ using the Nelder–Mead simplex algorithm, and then tuned and selected $\gamma(\tau)$ based on a criterion. We finally obtained the adapted value $\hat{\gamma}(\tau)$ and the corresponding coefficient $\hat{\beta}_n(\hat{\gamma}; \tau)$. To verify the theoretical properties of the proposed method, we conducted simulation studies under two simple scenarios with finite samples to illustrate the proposed method.

Scenario 1. We considered a simple univariate case,

$$Y_i = \alpha + \beta X_i + \epsilon_i, \quad i = 1, \dots, n,$$

where $\alpha = 0$, $\beta = 1$, X_i follows the standard normal distribution, and $\epsilon_i = \epsilon_i^* - \Phi^{-1}(\tau)$, with ϵ_i^* following the standard normal distribution and $\Phi^{-1}(\tau)$ indicating the τ -quantile of the standard normal distribution. We set the sample size $n = 200, 400$ and the number of replications as 500.

For Scenario 1, β_0 is of one dimension. So, we can plot the 3-dimensional surface of $W_n(\beta, \gamma; \tau = 0.5)$, see Figure 3, which shows that $W_n(\beta, \gamma; \tau)$ is local convex with respect to both γ and β . As illustrated previously in Example 1, if ϵ follows the standard normal distribution, the true value of $\gamma^*(\tau)$ in theory is around 1.08 by criterion (II). Table 1 summarizes the simulation results at different quantiles with various adaptive criteria. The results show that the estimated regression coefficients have small biases, and the biases demonstrate a clear trend of asymptotic consistency for all settings. According to the results in Table 1 and the box-plots in Figures 4 and 5, we also see that the adaptive parameter $\hat{\gamma}(0.5)$ selected by criteria (I) and (II) converge to 0.0, 1.13 and 0.8, respectively. The values of $\gamma(\tau)$ by criterion (I) are all equal to 0, indicating the estimates by criterion (I) are equivalent to that by the traditional quantile regression. Compared to the estimate by traditional quantile regression, the proposed estimate under the adaptive criterion (II) enhances the estimation efficiency of coefficients by 5~20%. Additionally, from the histogram in Figure 6 and the quantile–quantile (Q–Q) plot [21,22] in Figure 7, we see that the empirical distribution of the estimators follows a clear normal distribution.

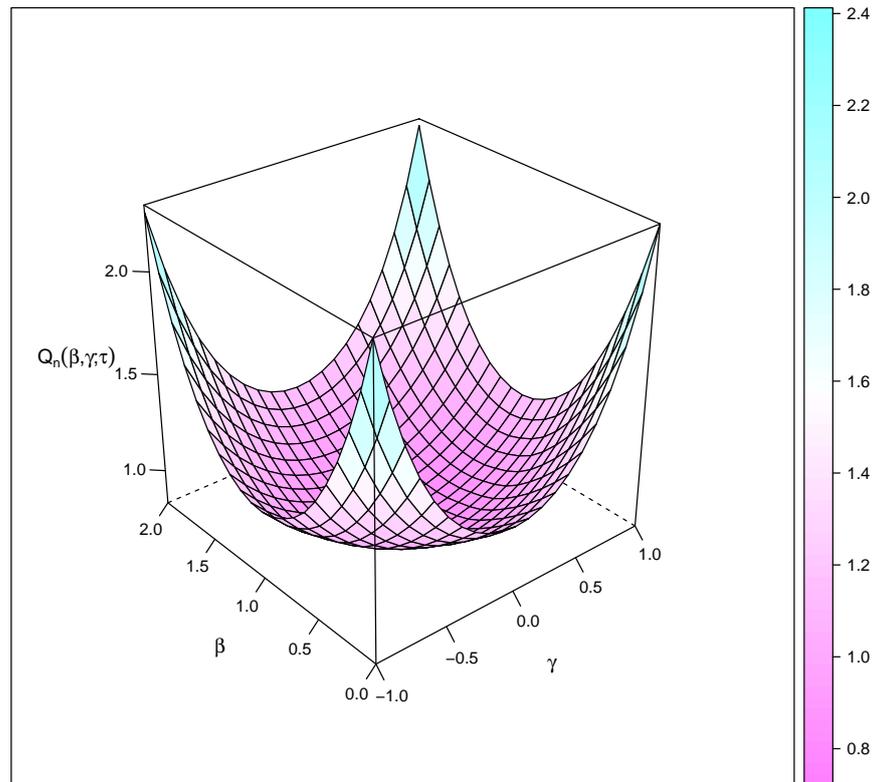


Figure 3. 3D surface of the objective function $W_n(\beta; \gamma, \tau = 0.5)$.

Table 1. Simulation results under Scenario 1.

Criterion	n	τ	$\beta_0 = 1$				γ	
			Est.	SD	SE	CP%	Est.	SD
I	200	0.25	0.995	0.067	0.069	94.8	0.000	0.000
		0.50	1.003	0.060	0.063	95.2	0.000	0.000
		0.75	1.001	0.073	0.070	93.0	0.000	0.000
	400	0.25	0.998	0.047	0.049	94.4	0.000	0.000
		0.50	0.999	0.044	0.044	93.0	0.000	0.000
		0.75	1.000	0.048	0.049	93.4	0.000	0.000
II	200	0.25	0.999	0.056	0.052	93.8	1.048	0.321
		0.50	1.002	0.057	0.050	93.6	1.070	0.339
		0.75	1.000	0.060	0.051	92.6	1.068	0.339
	400	0.25	1.002	0.041	0.039	94.0	1.126	0.246
		0.50	1.000	0.041	0.038	94.8	1.136	0.261
		0.75	0.999	0.042	0.039	93.5	1.137	0.242
III	200	0.25	0.996	0.051	0.054	96.2	0.781	0.234
		0.50	0.998	0.056	0.051	91.4	0.722	0.266
		0.75	0.999	0.061	0.054	90.8	0.760	0.248
	400	0.25	0.996	0.039	0.040	94.0	0.773	0.205
		0.50	1.001	0.038	0.037	94.8	0.713	0.216
		0.75	1.002	0.042	0.038	92.4	0.761	0.206

NOTE: Est., the empirical mean of the estimated coefficients; SD, the sample standard deviation of the estimated coefficients; SE, the empirical mean of standard errors based on the bootstrap method; CP, the coverage probability of the 95% confidence intervals.

Scenario 2. We considered the same model and settings as in Scenario 1, except that the error follows a uniform distribution, and $\epsilon_i = \epsilon_i^* - Q_{\epsilon^*}(\tau)$, where ϵ_i^* follows $\text{Uniform}(-1, 1)$, and $Q_{\epsilon^*}(\tau) = 2\tau - 1$.

The simulation results are presented in Table 2. It is shown that under criterion (I) that the estimated values of $\gamma^*(\tau)$ are all equal to 0 as well, and the corresponding estimated coefficients at different quantiles have small biases and reasonable coverage probabilities. Under the adaptive criterion (II), the estimated values of $\gamma^*(\tau)$ tend to converge to $\Gamma = 2$. In addition, the regression coefficients are all estimated accurately. Overall, the simulation results in Table 2 match well with the theoretical properties in Example 2 in Section 4. Specifically, the estimation efficiency of coefficients using the proposed method with criterion (II) increases by 60~100% over the traditional quantile regression.

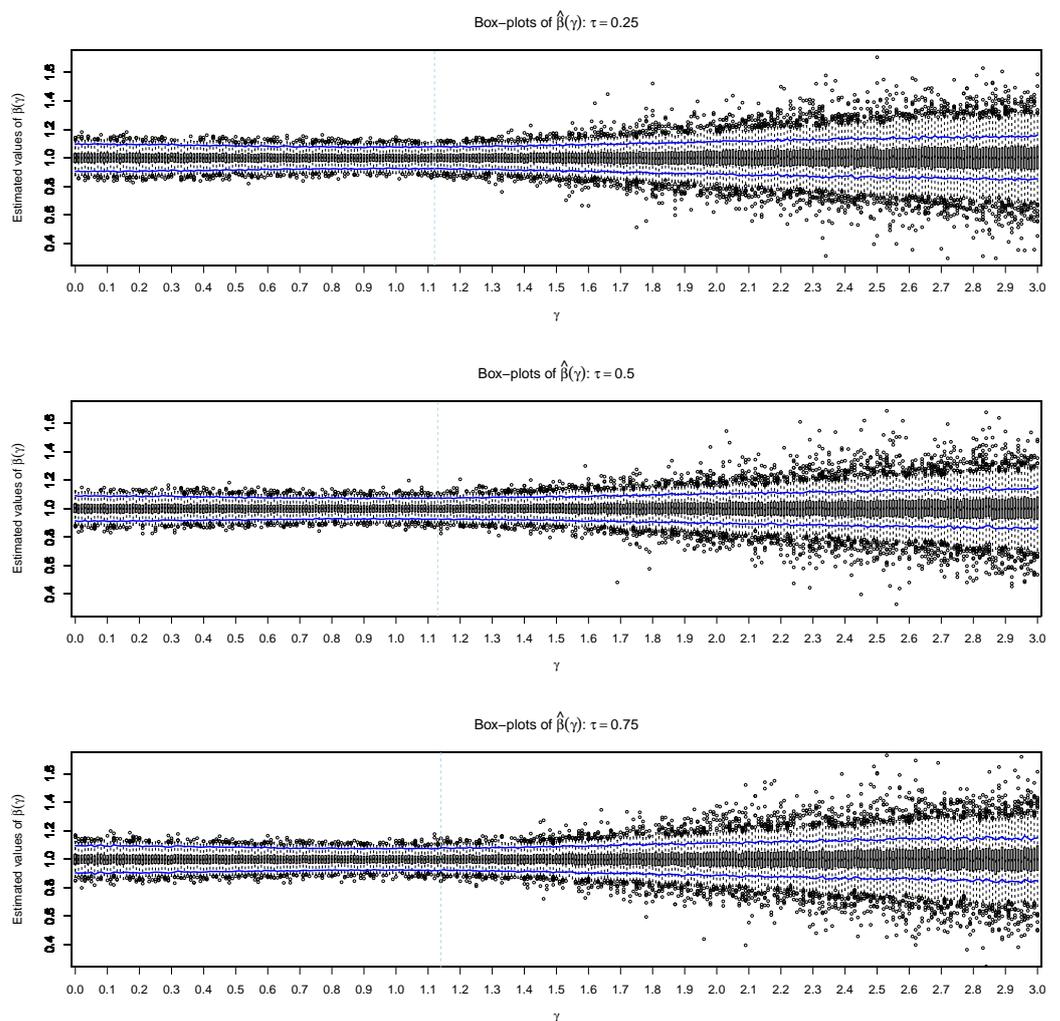


Figure 4. Box-plots of $\hat{\beta}_n(\gamma; \tau)$ at different values of γ and $\tau = 0.25, 0.50, 0.75$ under Scenario 1 with criterion (II). The solid lines with blue color indicate the point-wise upper and lower bounds of the 95% confidence interval for $\beta_0(\tau)$ at different γ , respectively.

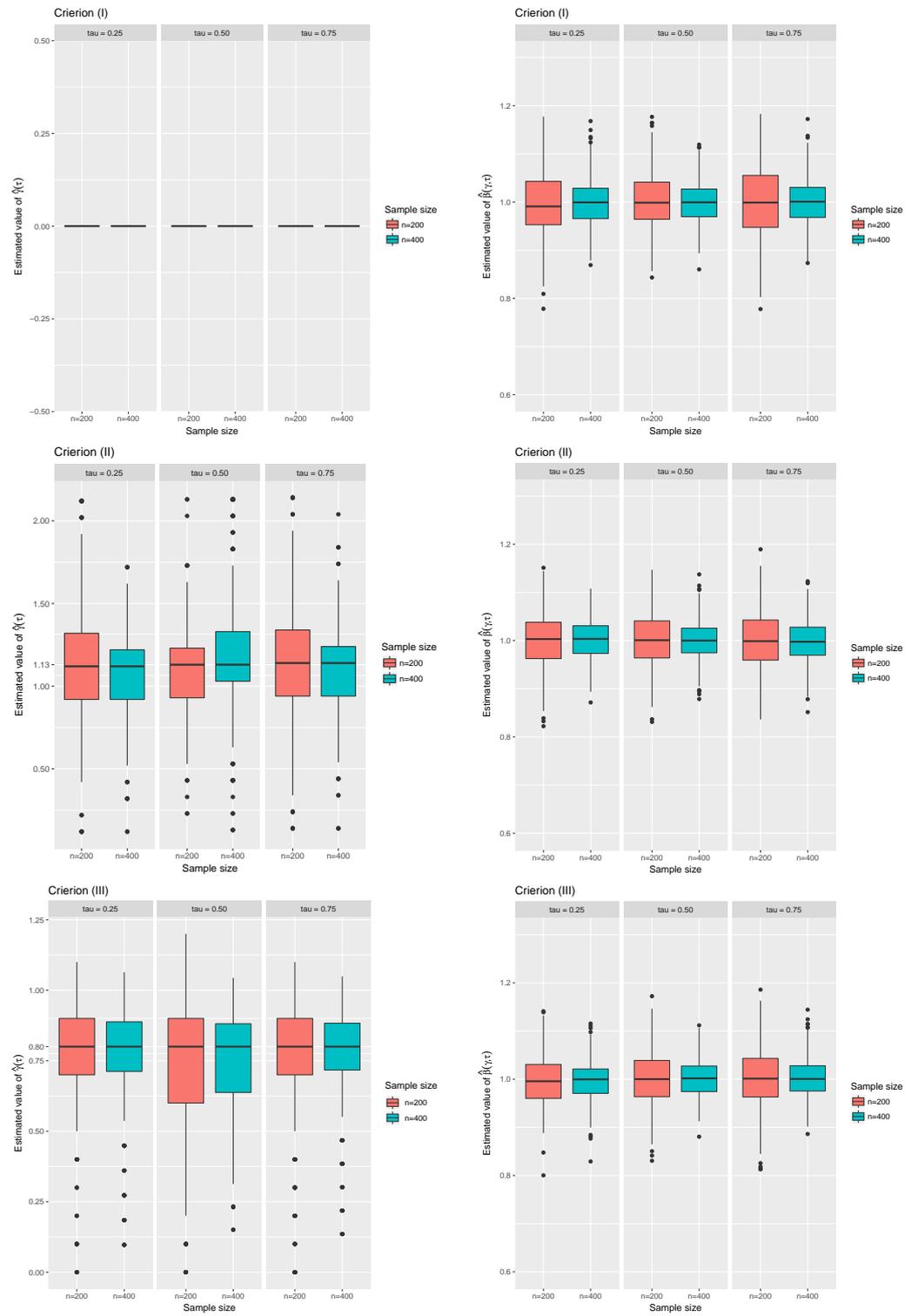


Figure 5. Boxplots of the values of $\hat{\beta}_n(\gamma; \tau)$ and $\hat{\gamma}(\tau)$ under Scenario 1 with three criteria, respectively.

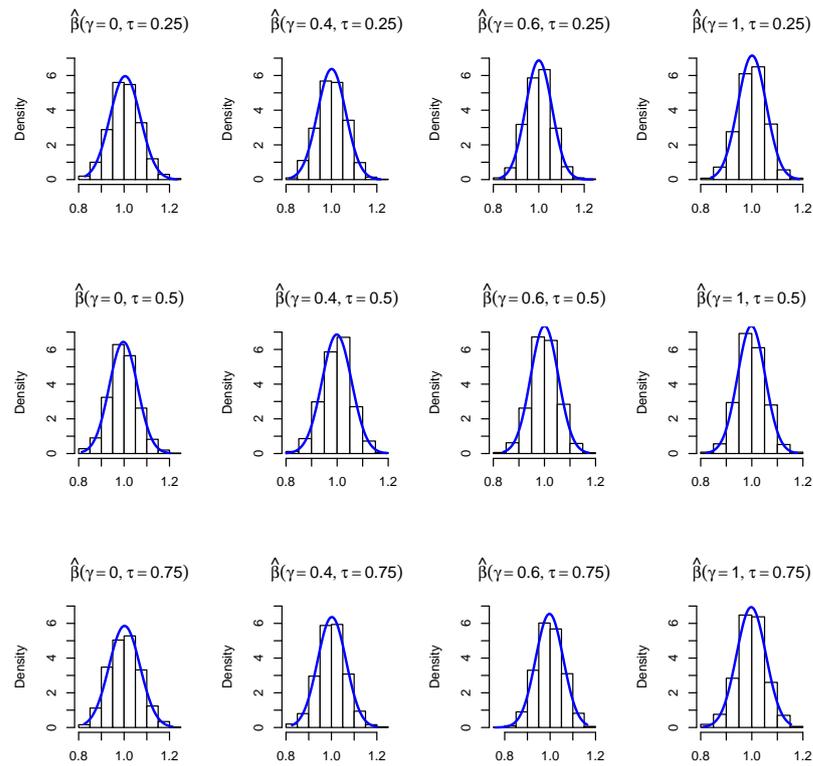


Figure 6. Histograms of $\hat{\beta}_n(\gamma; \tau)$'s at different values of γ and $\tau = 0.25, 0.50, 0.75$ with $n = 400$ under Scenario 1.

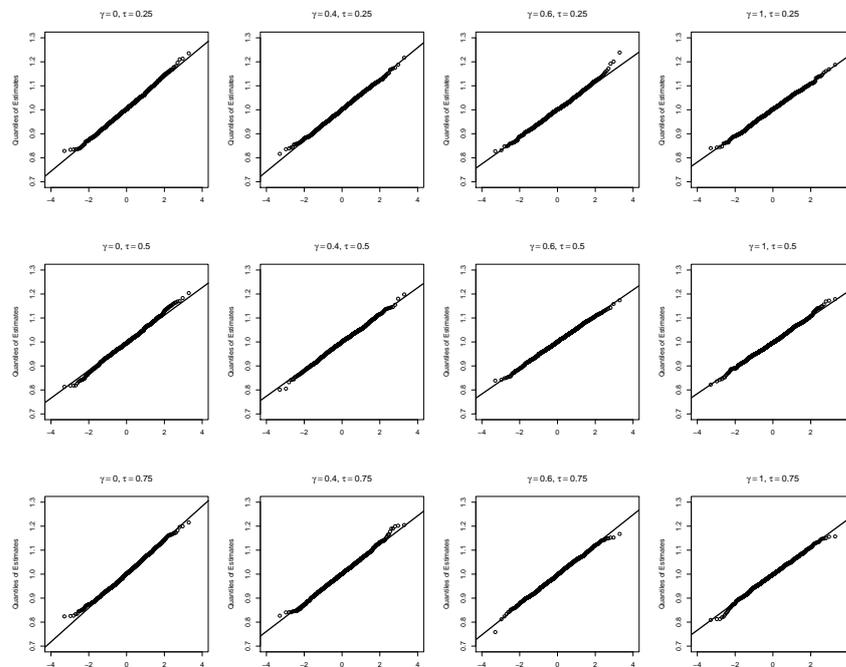


Figure 7. Q–Q plots of $\hat{\beta}_n(\gamma; \tau)$'s at different values of γ and $\tau = 0.25, 0.50, 0.75$ with $n = 400$ under Scenario 1.

Table 2. Simulation results under Scenario 2.

Criteria	n	τ	β ₀ = 1				γ	
			Est.	SD	SE	CP%	Est.	SD
I	200	0.25	1.005	0.062	0.063	95.0	0.000	0.000
		0.50	1.005	0.072	0.073	95.0	0.000	0.000
		0.75	0.997	0.066	0.063	93.4	0.000	0.000
	400	0.25	1.000	0.043	0.043	94.8	0.000	0.000
		0.50	0.999	0.052	0.052	94.4	0.000	0.000
		0.75	0.999	0.043	0.045	95.2	0.000	0.000
II	200	0.25	0.998	0.038	0.039	94.0	1.927	0.250
		0.50	0.999	0.037	0.038	95.2	1.996	0.081
		0.75	1.001	0.038	0.040	92.8	1.924	0.257
	400	0.25	1.001	0.027	0.026	95.8	1.940	0.221
		0.50	0.999	0.026	0.026	95.0	1.998	0.036
		0.75	1.000	0.027	0.029	94.0	1.942	0.201

NOTE: Est., the empirical mean of the estimated coefficients; SD, the sample standard deviation of the estimated coefficients; SE, the empirical mean of the estimated standard errors based on the bootstrap method; CP, the coverage probability of the 95% confidence intervals.

5. Application

We applied the proposed method to a prostate cancer study [23], where the prostate cancer data contain the medical records of 97 male patients who received a radical prostatectomy. The description of data is summarized in Table 3.

Table 3. Summary of prostate cancer data.

Components	Total	Components	Total
cancer volume (cc)		PSA (ng./mL.)	
mean (SD)	7.0 (7.9)	mean (SD)	23.74 (40.8)
median	4.25	median	13.35
range	45.39	range	265.2
prostate weight (gm.)		capsular penetration (cm.)	
mean (SD)	45.5 (45.6)	mean (SD)	2.36 (3.72)
median	37.45	median	0.45
range	438.5	range	18.0
Age (year)		Gleason score –no.(%)	
mean (SD)	63.9 (7.45)	= 6	35 (36)
median	65.00	= 7	56 (58)
range	38.00	= 8	1 (1)
		= 9	5 (5)
BPHA (cm ²)		Percentage of Gleason	
mean (SD)	2.64 (2.94)	scores 4 or 5 –no.(%)	
median	1.35	[0, 20]	61 (63)
range	9.99	(20, 60]	24 (25)
SVI -no. (%)		(60, 100]	12 (12)
+	21 (22)		
–	76 (78)		

NOTE: ‘SVI’, seminal vesicle invasion; ‘BPHA’, benign prostatic hyperplasia amount.

The response variable of interest is the level of prostate antigen (PSA), and there are eight predictor variables, including the log of cancer volume, the log of prostate weight, age, the log of the amount of benign prostatic hyperplasia, seminal vesicle invasion, the log of capsular penetration, Gleason score and the percentage of Gleason scores of 4 or 5. The research goal is to study the covariate effects of the predictor variables at the level of prostate antigen.

We considered a linear quantile model for the association between prostate antigen and the predictor variables. For convenience, we first standardized the predictors as well as the dependent variables, and then included the intercept in the model. We took 200 bootstrap samples for the variance estimation of coefficients, and selected the adaptive parameter using criterion (II). We obtained the estimate $\hat{\gamma}(\tau) = 2.0$ for all grid points $\tau \in \{0.1, 0.2, \dots, 0.8, 0.9\}$. Then, we further acquired the estimated regression coefficients with 95% Confidence Intervals (CIs) at the 0.25, 0.50 and 0.75 quantiles, which are summarized in Table 4, where the results are also compared to those of the traditional quantile regression. As is shown in the table, there is not much difference between the estimated coefficients by the traditional quantile regression and the proposed method with criterion (II). However, referring to the values of the ‘CI ratio’ (‘CI Ratio’ is defined as the ratio of the length of 95% CI by traditional QR over that by the proposed method using criterion (II)), we can see overall that the lengths of 95% CIs by the proposed method are significantly smaller than that by a traditional quantile regression. Figure 8 presents the estimated curves of coefficients with 95% confidence bands. The figure shows that the proposed method with criterion (II) provides smoother estimates and more stable and narrower confidence bands compared to those by a regular quantile regression, especially at the region of tail quantiles. From the results in both Table 4 and Figure 8, we concluded that the cancer volume, the prostate weight, and the seminal vesicle invasion are significantly associated with the level of prostate antigen, which is as expected since a high level of prostate antigen is generally regarded as strong evidence of prostate cancer. While the effects of the amount of benign prostatic hyperplasia, capsular penetration, Gleason score and the percentage of Gleason scores of 4 or 5 at the level of prostate antigen are statistically insignificant.

Table 4. The estimated covariate effects with 95% CIs for the prostate cancer study.

Coeff. $\hat{\beta}(\tau)$	Traditional QR (Criterion (I))		Criterion (II)		CI Ratio
	Est.	95% CI	Est.	95% CI	
$\tau = 0.25$					
Intercept	−0.354	(−0.513, −0.167)	−0.217	(−0.333, −0.102)	1.498
Volume	0.696	(0.486, 0.864)	0.611	(0.450, 0.772)	1.174
Weight	0.282	(0.135, 0.473)	0.238	(0.101, 0.374)	1.238
Age	−0.033	(−0.203, 0.150)	−0.147	(−0.272, −0.023)	1.418
Hyperplasia	0.130	(0.007, 0.435)	0.102	(−0.039, 0.244)	1.512
Seminal	0.275	(0.087, 0.408)	0.248	(0.113, 0.383)	1.189
Capsular	−0.259	(−0.504, −0.090)	−0.150	(−0.324, 0.024)	1.190
Gleason	−0.022	(−0.189, 0.152)	0.039	(−0.137, 0.215)	0.969
Percentage	0.246	(−0.091, 0.480)	0.128	(−0.048, 0.305)	1.618
$\tau = 0.5$					
Intercept	−0.057	(−0.185, 0.071)	0.009	(−0.090, 0.109)	1.286
Volume	0.544	(0.363, 0.725)	0.601	(0.459, 0.742)	1.279
Weight	0.239	(0.059, 0.419)	0.220	(0.091, 0.349)	1.395
Age	−0.173	(−0.307, −0.039)	−0.116	(−0.222, −0.010)	1.264
Hyperplasia	0.201	(0.042, 0.360)	0.112	(−0.004, 0.228)	1.371
Seminal	0.287	(0.087, 0.487)	0.240	(0.095, 0.386)	1.375
Capsular	−0.159	(−0.343, 0.026)	−0.111	(−0.274, 0.052)	1.132
Gleason	0.127	(−0.050, 0.304)	0.071	(−0.087, 0.230)	1.117
Percentage	0.099	(−0.050, 0.248)	0.084	(−0.081, 0.248)	0.906
$\tau = 0.75$					
Intercept	0.378	(0.312, 0.514)	0.253	(0.143, 0.363)	0.918
Volume	0.570	(0.348, 0.752)	0.592	(0.452, 0.733)	1.438
Weight	0.136	(−0.067, 0.342)	0.199	(0.058, 0.339)	1.456
Age	−0.135	(−0.305, 0.035)	−0.121	(−0.241, −0.002)	1.423
Hyperplasia	0.155	(−0.021, 0.333)	0.087	(−0.048, 0.222)	1.311
Seminal	0.334	(0.198, 0.439)	0.261	(0.186, 0.405)	1.100
Capsular	−0.089	(−0.383, 0.065)	−0.070	(−0.227, 0.088)	1.422
Gleason	−0.061	(−0.210, 0.143)	−0.019	(−0.161, 0.123)	1.243
Percentage	0.170	(−0.016, 0.226)	0.125	(−0.059, 0.308)	0.659

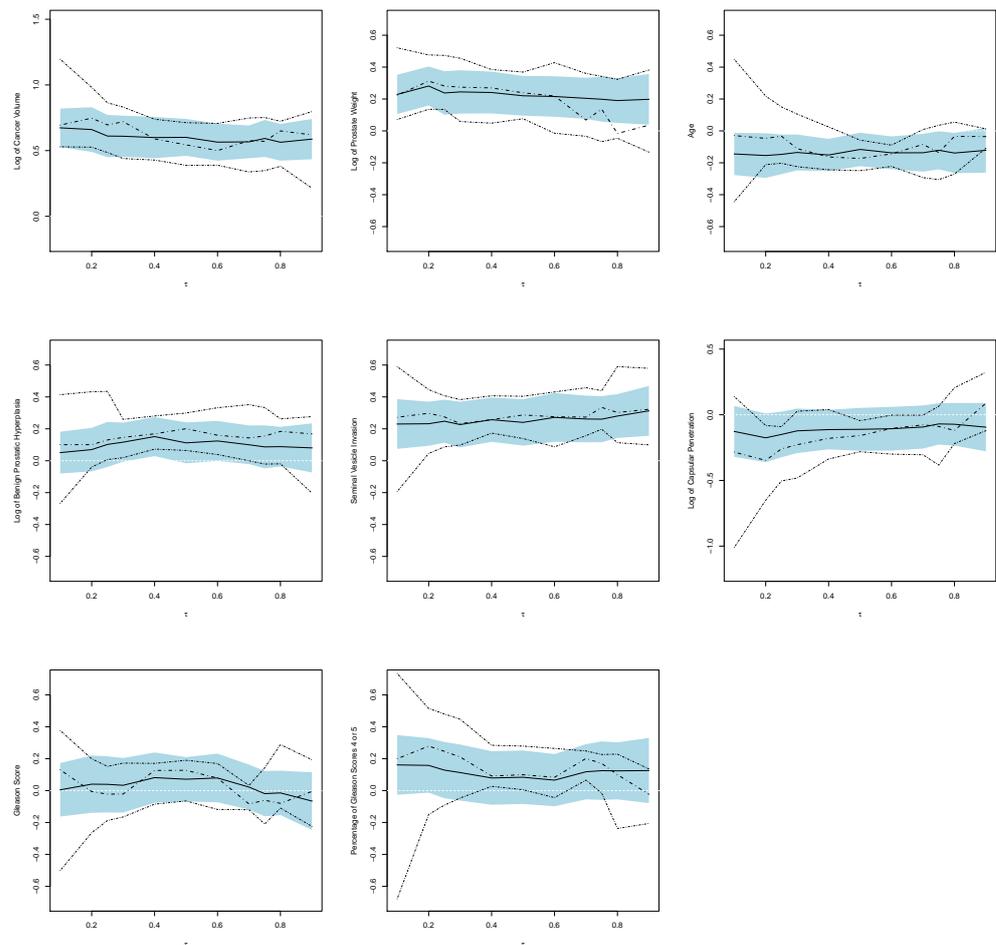


Figure 8. The estimated curves of covariate effects along the quantiles for the prostate cancer study. The solid lines present the estimated coefficients by proposed method with criterion (II), and the dash-dot lines are by the traditional quantile regression. The shaded areas give the 95% confidence bands of the estimated curves by proposed method with criterion (II). The 95% confidence bands of the curves by regular quantile regression are shown between the lower and upper dash-dot lines.

6. Discussion

We propose a general class of loss functions to conduct a quantile regression, which naturally unifies the absolute and the relative error criteria. The consistency and asymptotic normality of the resulting estimators are established. Numerical studies demonstrate the good performance of the proposed method using finite samples. Although our proposal is based on a quantile regression, similar ideas can also be extended to other statistical framework such as the M-estimator. Further research includes extending the proposed procedure to censored data, see, e.g., [7,24,25], or longitudinal data [6,26], and even to the case with a diverging number of parameters or ultra-high dimensionality situations [27].

Author Contributions: Conceptualization, T.H. and B.L.; methodology, T.H.; simulation study, B.L.; mathematical proof, B.L.; formal analysis, B.L.; investigation, T.H. and B.L.; resources, T.H.; data curation, B.L.; writing—original draft preparation, T.H.; writing—review and editing, B.L.; visualization, B.L.; project administration, B.L.; funding acquisition, B.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (No. 11901013), Beijing Natural Science Foundation (No. 1204031), and the Fundamental Research Funds for the Central Universities. Tao Hu's work was partly supported by the Beijing Talent Foundation

Outstanding Young Individual Project and the Academy for Multidisciplinary Studies of Capital Normal University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The prostate cancer data used in this article is publicly available in the R package bayesQR at <http://127.0.0.1:23924/library/bayesQR/html/Prostate.html> (accessed on 31 March 2021).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Lemma A1. Let $Q(x, \gamma; \epsilon, \tau) = \gamma^{-1}[\exp\{\gamma(\epsilon - x)\} - \exp\{-\gamma(\epsilon - x)\}]\{\tau - I(\epsilon < x)\}$, where $x \in \mathbb{R}$, $\tau \in (0, 1)$. Then, (i) for any fixed $\gamma > 0$, $Q(x, \gamma; \epsilon, \tau)$ is a strictly convex function with respect to $x \in \mathbb{R}$; (ii) for any fixed x and ϵ , $Q(x, \gamma; \epsilon, \tau)$ is monotone increasing with respect to γ on $(0, \Gamma)$. In particular, $Q(x, \gamma; \epsilon, \tau)$ reaches the minimum when $\gamma = 0$.

Proof of Lemma A1. The first part of Lemma 1 follows directly from the fact that

$$\frac{d^2Q(x, \gamma; \epsilon, \tau)}{dx^2} = \gamma[\exp\{\gamma(\epsilon - x)\} - \exp\{-\gamma(\epsilon - x)\}]\{\tau - I(\epsilon < x)\} > 0.$$

The second part of Lemma A1 follows from the first derivative of $Q(x, \gamma; \epsilon, \tau)$ with respect to γ . Let $u = \exp(\epsilon - x)$. By simple derivation, we obtain

$$\begin{aligned} \frac{dQ(x, \gamma; \epsilon, \tau)}{d\gamma} &= \frac{\tau - I(u < 1)}{\gamma^2} \{u^\gamma(\gamma \ln(u) - 1) + u^{-\gamma}(\gamma \ln(u) + 1)\} \\ &= \frac{\tau - I(u < 1)}{\gamma^2} \left\{ \frac{(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1)}{u^\gamma} \right\} \\ &= \begin{cases} \frac{\tau - 1}{\gamma^2} \frac{(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1)}{u^\gamma}, & \text{if } 0 < u < 1, \\ \frac{\tau}{\gamma^2} \frac{(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1)}{u^\gamma}, & \text{if } u \geq 1. \end{cases} \end{aligned}$$

If $0 < u < 1$, then $\gamma \ln(u) - 1 < -1$, $u^{2\gamma} < 1$, and $\gamma \ln(u) + 1 < 1$. Thus, $(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1) < 0$ and then $\frac{\tau - 1}{\gamma^2} \frac{(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1)}{u^\gamma} > 0$. Using a similar argument, we can also show $\frac{\tau}{\gamma^2} \frac{(\gamma \ln(u) - 1)u^{2\gamma} + (\gamma \ln(u) + 1)}{u^\gamma} > 0$. Hence, $Q(x, \gamma; \epsilon, \tau)$ is monotone increasing with respect to γ . In particular, by L'Hôpital's rule

$$\lim_{\gamma \rightarrow 0} Q(x, \gamma; \epsilon, \tau) = 2(\epsilon - x)\{\tau - I(\epsilon < x)\} = Q(x, \gamma; \epsilon, \tau)|_{\gamma=0}.$$

This completes the proof of Lemma 1. \square

References

1. Koenker, R. *Quantile Regression*; Cambridge University Press: New York, NY, USA, 2005.
2. Koenker, R.; Bassett, G.S. Regression quantiles. *Econometrica* **1978**, *46*, 33–50.
3. Beyerlein, A. Quantile regression—Opportunities and challenges from a user's perspective. *Am. J. Epidemiol.* **2014**, *180*, 330–331.
4. Koenker, R. Quantile regression: 40 years on. *Annu. Rev. Econ.* **2017**, *9*, 155–176.
5. Fenske, N.; Fahrmeir, L.; Hothorn, T.; Rzehak, P.; Höhle, M. Boosting structured additive quantile regression for longitudinal childhood obesity data. *Int. J. Biostat.* **2013**, *9*, 1–18.
6. Koenker, R. Quantile regression for longitudinal data. *J. Multivar. Anal.* **2004**, *91*, 74–89.
7. Peng, L.; Huang, Y. Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **2008**, *103*, 637–649.
8. Angrist, J.; Chernozhukov, V.; Fernández-Val, I. Quantile Regression under misspecification, with an application to the U.S. wage structure. *Econometrica* **2006**, *74*, 539–563.

9. Waldmann, E. Quantile regression: A short story on how and why. *Stat. Model.* **2018**, *18*, 203–218.
10. Narula, S.; Wellington, J. Prediction, linear regression and the minimum sum of relative errors. *Technometrics* **1977**, *19*, 185–190.
11. Khoshgoftaar, T.M.; Bhattacharyya, B.B.; Richardson, G.D. Predicting software errors, during development, using nonlinear regression models: A comparative study. *IEEE Trans. Reliab.* **1992**, *41*, 390–395.
12. Chen, K.; Guo, S.; Lin, Y.; Ying, Z. Least absolute relative error estimation. *J. Am. Stat. Assoc.* **2010**, *105*, 1104–1112.
13. Yang, Y.; Ye, F. General relative error criterion and M-estimation. *Front. Math. China* **2013**, *8*, 695–715.
14. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **1964**, *26*, 211–252.
15. Sakia, R.M. The Box-Cox transformation technique: A review. *Statistician* **1992**, *41*, 169–178.
16. Gneiting, T. Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **2011**, *106*, 746–762.
17. Yin, G.; Cai, J. Quantile regression models with multivariate failure time data. *Biometrics* **2005**, *61*, 151–161.
18. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199.
19. Knight, K. Limiting distribution for L_1 regression estimators under general conditions. *Ann. Stat.* **1998**, *26*, 755–770.
20. Su, L.; Xu, P. Common threshold in quantile regressions with an application to pricing for reputation. *Econom. Rev.* **2017**, *38*, 1–37.
21. Almeida, A.; Loy, A.; Hofmann, H. ggplot2 Compatible Quantile-Quantile Plots in R. *R J.* **2019**, *10*, 248–261.
22. Wilk, M.B.; Gnanadesikan, R. Probability plotting methods for the analysis of data. *Biometrika* **1968**, *55*, 1–17.
23. Stamey, T.A.; Kabalin, J.N.; McNeal, J.E.; Johnstone, I.M.; Freiha, F.; Redwine, E.A.; Yang, N. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. radical prostatectomy treated patients. *J. Urol.* **1989**, *141*, 1076–1083.
24. Powell, J.L. Censored regression quantiles. *J. Econom.* **1986**, *32*, 143–155.
25. Wang, H.; Wang, L. Locally weighted censored quantile regression. *J. Am. Stat. Assoc.* **2009**, *104*, 1117–1128.
26. Farcomeni, A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput.* **2012**, *22*, 141–152.
27. He, X.; Wang, L.; Hong, H.G. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann. Stat.* **2013**, *41*, 342–369.