Quantile Regression with Left-Truncated and Right-Censored Data in a Reproducing Kernel Hilbert Space

Jinho Park

To cite this article: Jinho Park (2015) Quantile Regression with Left-Truncated and Right-Censored Data in a Reproducing Kernel Hilbert Space, Communications in Statistics - Theory and Methods, 44:7, 1523-1536, DOI: 10.1080/03610926.2013.777741

To link to this article: https://doi.org/10.1080/03610926.2013.777741

Published online: 26 Mar 2015.
Quantile Regression with Left-Truncated and Right-Censored Data in a Reproducing Kernel Hilbert Space

JINHO PARK

Department of Statistics, Inha University, Incheon, Korea

Li et al. (2007) developed an estimation method for quantile functions in a reproducing kernel Hilbert space for complete data, and Park and Kim (2011) proposed an estimation method using the \( \epsilon \)-insensitive loss. This article extends these estimation methods to left-truncated and right-censored data. As a measure of goodness of fit, the check loss and the \( \epsilon \)-insensitive loss were used to estimate the quantile function. The \( \epsilon \)-insensitive loss can shrink the estimated coefficients toward zero; hence, it can reduce the variability of the estimates. Simulation studies show that the estimated quantile functions based on the \( \epsilon \)-insensitive loss perform slightly better when \( \epsilon \) is adequately chosen.

Keywords  Quantile; Censoring; Truncation; Reproducing kernel Hilbert space

Mathematics Subject Classification  Primary 62G08; Secondary 62G05

1. Introduction

A popular approach to assessing the effect of covariates on a censored (and truncated) response is to fit a proportional hazard model (Cox, 1972). The Cox model assumes a semiparametric relationship between the covariates and the hazard function of survival time. There has also been considerable interest in modeling the conditional mean of survival time. Because the mean regression model directly fits survival times, it is simple to interpret the fitted model. The conditional mean regression models were studied by Miller (1976) and Buckley and James (1979) for censored data, and by Gross and Lai (1996) for censored and truncated data. The Cox model was extended to nonparametric or regularized (penalized) hazard models by Tibshirani (1997), Li and Li (2004), and Gui and Li (2005), and the mean regression model was also explored using nonparametric or regularized models. Park (2004) used a nonparametric regression model based on multivariate adaptive regression splines (Friedman, 1991), and Datta, et al. (2007) used the partial least squares and LASSO methodologies to estimate survival times.
As an alternative to the mean regression model, the conditional quantile regression model can be used for censored data as in Powell (1984) and Portnoy (2003). Compared with the mean regression model, the conditional quantile regression model allows non-homogeneous variability and more robust estimation. It also provides more information about the spread and skewness of the survival time.

Nonparametric methodologies have often been used to provide a flexible relationship between the response and covariates. Nonparametric estimations in a reproducing kernel Hilbert space (RKHS) have been used frequently to estimate the regression function. The estimation method can be applied to the quantile regression model. Li, et al. (2007) proposed a method for estimating the conditional quantile function in an RKHS, as explained below.

For a random sample \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) with the covariate \(X\) and the response \(Y\), the estimated \(\tau\)-th quantile function \(\hat{f}_\lambda\) in an RKHS with the kernel \(K(\cdot, \cdot)\) can be given by

\[
\hat{f}_\lambda = \arg\min_{f \in \mathcal{H}_K} \left\{ \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - f(X_i)) + \frac{\lambda}{2} ||f||^2_{\mathcal{H}_K} \right\},
\]

where \(\rho_\tau\) is the check loss given by \(\rho_\tau(x) = (\tau - I_{\{x<0\}})x\) with \(I_{\{x<0\}} = 1\) if \(x < 0\) and 0, otherwise || \(\cdot\)||\(\mathcal{H}_K\) is the norm induced by the kernel \(K(\cdot, \cdot)\) and \(\lambda\) is the smoothing parameter.

The above-estimating equation cannot be directly applied due to incomplete information when the survival data are censored and truncated. One method of estimating the model coefficients is using the E-M type iterative algorithms as presented in Buckley and James (1979) and Wang et al. (2008). Noniterative and weighted algorithms are also frequently used to analyze the censored survival data as reported in Gross and Lai (1996) and Park (2004).

This article proposes a weighted algorithm that can estimate the conditional quantile function for left-truncated and right-censored (LTRC) data in an RKHS. As a measure of the goodness of fit of quantile estimates, the check loss and a modified check loss with an insensitive zone are used. A criterion is introduced to select the smoothing parameter based on generalized approximate cross-validation (GACV). The remainder of this article is organized as follows. Section 2 discusses the development of an algorithm for estimating the quantile function. In Sec. 3, a method for selecting the optimal smoothing parameter using GACV is derived. In Sec. 4, some simulation results are presented, and the proposed method is used to analyze Stanford heart transplant data.

### 2. Estimation of the Quantile Function

Suppose that \((X_1, Y_1), (X_2, Y_2), \ldots\) are independent and identically distributed random variables, where \(X_i\) is a (multivariate) covariate and \(Y_i\) is a survival time. Let \(C_i\) and \(T_i\) denote a right-censoring time and a left truncation time, respectively, and assume that \((C_i, T_i)\) are independent of \((X_i, Y_i)\). When the survival time \(Y_i\) is subject to right censoring, the censored survival time \(\hat{Y}_i = \min(Y_i, C_i)\) and the censoring indicator \(\delta_i = I\{Y_i \leq C_i\}\) are observed. If the survival times are subject to left truncation in addition to right censoring, an observation can only be made when \(\hat{Y}_i \geq T_i\). Let

\[(X_i, \hat{Y}_i, \delta_i, T_i), \quad i = 1, 2, \ldots, n \quad \text{with} \quad \hat{Y}_i \geq T_i\]

denote the observed data.
Let $\eta$ and $\overline{\eta}$ be the left and right boundaries of the interval, respectively, within which the survival time can be observed under left truncation and right censoring. Note that $\eta$ and $\overline{\eta}$ correspond to $\tau$ and $\overline{\tau}$ in Park (2004). Due to incomplete information that results from the censoring and truncation, the distribution of the survival time cannot be nonparametrically estimated outside the interval $[\eta, \overline{\eta}]$. First, suppose that $F(\eta) = 0$ and $F(\overline{\eta}) = 1$, where $F(\cdot)$ is the distribution function of the survival time. Then, the distribution function $F(\cdot)$ is estimable over the whole support of the survival time, and the product-limit estimator of the distribution function is given by

$$\hat{F}(y) = 1 - \prod_{y_{(i)} \leq y} \left\{ 1 - \frac{d(i)}{n(i)} \right\},$$

where $y_{(1)} < y_{(2)} < \ldots$ are the distinct uncensored survival times, $d(i)$ is the multiplicity of the uncensored survival times at $y_{(i)}$, $n(i)$ is the size of the risk set at $y_{(i)}$, i.e., $n(i)$ is the number of $j$ such that $T_j \leq y_{(i)} \leq \overline{Y}_j$ (Tsai, et al., 1987).

Now, suppose that $F(\eta) > 0$ or $F(\overline{\eta}) < 1$. Then, $F(\cdot)$ cannot be estimated on the outside of the interval $[\eta, \overline{\eta}]$. Thus, fix $a > \eta$ and $b < \overline{\eta}$. Gross and Lai (1996) showed that for any integrable function $g(\cdot)$, $E[g(X, \overline{Y})|a \leq Y \leq b]$ can be consistently estimated using

$$\frac{1}{\hat{F}_a(b)} \sum_{t=1}^{n} \delta_t I(a \leq \overline{Y}_t \leq b)g(X_t, \overline{Y}_t) \frac{\hat{S}_a(\overline{Y}_t)}{\#(\overline{Y}_t)}.$$  \hspace{1cm} (2)

In the above-estimating equation, $\hat{F}_a(b)$ is the product-limit estimator of the conditional distribution function $F_a(b) = P(Y \leq t|Y \geq a)$, which is given by

$$\hat{F}_a(b) = 1 - \prod_{a \leq y_{(i)} \leq t} \left\{ 1 - \frac{d(i)}{n(i)} \right\} \text{ for } t \geq a,$$

and $\hat{S}_a(t)$ is the estimator of the conditional survival function $S_a(b) = P(Y \geq t|Y \geq a)$ which is given by

$$\hat{S}_a(b) = \prod_{a \leq y_{(i)} < t} \left\{ 1 - \frac{d(i)}{n(i)} \right\} \text{ for } t \geq a,$$

where $\#(\overline{Y}_{(i)})$ is the size of the risk set at $\overline{Y}_{(i)}$. While $E[g(X, Y)]$ cannot be estimated due to incomplete information about the survival time $Y$, $E[g(X, Y)|a \leq Y \leq b]$ can be estimated within the interval $[a, b]$ where observations of the survival time $Y$ can be made.

### 2.1 Estimation with the Check Loss

For the nonparametric estimation of the $\tau$-th quantile in a hypothesis space $\mathcal{H}$, the theoretical quantile function $f_\tau(\cdot)$ can be defined as

$$f_\tau = \arg\min_{f \in \mathcal{H}} E[\rho_\tau(Y - f(X))].$$

However, as explained before, $E[\rho_\tau(Y - f(X))]$ is not estimable for censored and truncated data. Therefore, the optimal quantile function $f^*_\tau(\cdot)$ in $\mathcal{H}$ can be defined as

$$f^*_\tau = \arg\min_{f \in \mathcal{H}} E[\rho_\tau(Y - f(X))|a \leq Y \leq b].$$
Note that $f^*_\tau(\cdot)$ can be interpreted as the best approximation for the true quantile function in $\mathcal{H}$ with the presence of left truncation and right censoring. If observations of the survival time on the whole support of $Y$ can be made, $a$ and $b$ can be chosen so that they are near the left and right boundaries of the support of $Y$. Then, the condition $\{a \leq Y \leq b\}$ will be negligible, $f^*_\tau$ will be close to $f_\tau$, and the estimating Eq. (2) will be as follows:

$$
\sum_{i=1}^{n} \delta_i g(X_i, Y_i) \frac{\hat{S}_a(Y_i)}{#(Y_i)}.
$$

Note that the above-estimating equation is the same as that presented in Zhu (1992).

For an estimation in an RKHS, it can be assumed that the estimated quantile function has the form $f(x) = \beta + h(x)$, where $h \in \mathcal{H}_K$ and $\mathcal{H}_K$ is an RKHS with the reproducing kernel $K(\cdot, \cdot)$. Using the estimating Eq. (2), an estimated quantile function can be defined as

$$
\hat{f}_\lambda = \arg \min_{\beta, h \in \mathcal{H}_K} \left\{ \sum_{i=1}^{n} W_i \rho_\tau(\hat{Y}_i - \beta - h(X_i)) + \frac{\lambda}{2} ||h||_{\mathcal{H}_K}^2 \right\},
$$

where

$$W_i = \frac{1}{\hat{F}_a(b)} \delta_i I(a \leq \hat{Y}_i \leq b) \frac{\hat{S}_a(\hat{Y}_i)}{#(\hat{Y}_i)}. \quad (3)$$

The above-estimating equation can be regarded as a weighted version of (1) or Eq. (4) of Li et al. (2007). Under an additional assumption that $P(T \leq C) = 1$, Shen (2014) developed a different estimation method for LTRC data. The weight given in Eq. (8) of Shen (2014) can also be used instead of the above $W_i$. The algorithm to estimate the quantile function $\hat{f}_\lambda$ will be explained in Sec. 2.2.

### 2.2 Estimation with the $\epsilon$-insensitive Check Loss

As an alternative to the check loss, a modified check loss can be used. An epsilon($\epsilon$)-insensitive check loss $\rho^*_\tau(\cdot)$ is defined as

$$
\rho^*_\tau(x) = \max(0, \rho_\tau(x) - \epsilon)
$$

$$= \begin{cases} 
\rho_\tau(x) - \epsilon & \text{if } \rho_\tau(x) \geq \epsilon, \\
0 & \text{if } \rho_\tau(x) < \epsilon.
\end{cases}
$$

The $\epsilon$-insensitive check loss sets the ordinary check loss to zero if it is less than $\epsilon$, and reduces it by $\epsilon$ if it is greater than $\epsilon$. When $\epsilon = 0$, the above loss is the same as the regular check loss. Park and Kim (2011) demonstrated that the $\epsilon$-insensitive check loss has the effect of shrinking the estimated coefficients for complete data. At the end of this section, the same shrinkage effect of the $\epsilon$-insensitive loss will be explained.

Using the $\epsilon$-insensitive check loss, the estimated $\tau$-th quantile function $\hat{f}_\lambda$ can be defined as

$$
\hat{f}_\lambda = \arg \min_{\beta, h \in \mathcal{H}_K} \left\{ \sum_{i=1}^{n} W_i \rho^*_\tau(\hat{Y}_i - \beta - h(X_i)) + \frac{\lambda}{2} ||h||_{\mathcal{H}_K}^2 \right\}, \quad (4)
$$
Quantile Regression with Censored Data in an RKHS

where \(W_i\) is the weight given in (3). By the representer theorem (Kimeldorf and Wahba, 1971), the estimated quantile function can be written as a linear combination of the kernels at observations \(X_j\), as follows:

\[
\hat{f}_\lambda(x) = \beta + \sum_{j=1}^{n} c_j K(X_j, x).
\]

When the optimizing objective (4) is rephrased using slack variables \(\xi_i\) and \(\xi_i^*\), the estimating equation can be written as follows:

\[
\begin{align*}
&\tau \sum_{i=1}^{n} W_i \xi_i + (1 - \tau) \sum_{i=1}^{n} W_i \xi_i^* \\
&\quad + \frac{\lambda}{2} \left\| \sum_{j=1}^{n} c_j K(X_j, \cdot) \right\|_{\mathcal{H}_K}^2 \\
&\text{subject to} \\
&\quad \hat{Y}_i - \left[ \beta + \sum_{j=1}^{n} c_j K(X_j, X_i) \right] \leq \frac{\xi_i}{\tau} + \xi_i, \\
&\quad \left[ \beta + \sum_{j=1}^{n} c_j K(X_j, X_i) \right] - \hat{Y}_i \leq \frac{\xi_i^*}{1 - \tau} + \xi_i^*, \quad i = 1, \ldots, n. \\
&\quad \xi_i \geq 0, \quad \xi_i^* \geq 0.
\end{align*}
\]

Let \(K\) be the matrix with the \((i, j)\)-th element \(K(X_i, X_j)\). The optimization problem, under these constraints, can be solved using the Lagrangian function \(L(\cdot)\), which is defined as

\[
L(\beta, c, \xi, \xi^*, \alpha, \alpha^*, \zeta, \zeta^*) = \tau \sum_{i=1}^{n} W_i \xi_i + (1 - \tau) \sum_{i=1}^{n} W_i \xi_i^* + \frac{\lambda}{2} c^T K c \\
- \sum_{i=1}^{n} \alpha_i \left\{ \frac{\xi_i}{\tau} + \xi_i - \hat{Y}_i + \left[ \beta + \sum_{j=1}^{n} c_j K(X_j, X_i) \right] \right\} \\
- \sum_{i=1}^{n} \alpha_i^* \left\{ \frac{\xi_i^*}{1 - \tau} + \xi_i^* + \hat{Y}_i - \left[ \beta + \sum_{j=1}^{n} c_j K(X_j, X_i) \right] \right\} \\
- \sum_{i=1}^{n} \zeta_i \xi_i - \sum_{i=1}^{n} \xi_i^* \xi_i^*,
\]

where \(\alpha_i, \alpha_i^*, \zeta_i, \) and \(\xi_i^*\) are the non negative Lagrange multipliers; \(c, \xi, \xi^*, \alpha, \alpha^*, \zeta, \) and \(\zeta^*\) are vectors with corresponding \(n\) coordinates.

By setting the derivatives of the Lagrangian function to 0, the following constraints are obtained:

\[
\frac{\partial}{\partial \beta} L(\cdot) = 0 \iff -\sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \alpha_i^* = 0,
\]

\[
\frac{\partial}{\partial c} L(\cdot) = 0 \iff \lambda K c - (K \alpha - K \alpha^*) = 0,
\]

\[
\frac{\partial}{\partial \xi_i} L(\cdot) = 0 \iff \tau W_i - \alpha_i - \zeta_i = 0 \quad i = 1, 2, \ldots, n,
\]

\[
\frac{\partial}{\partial \xi_i^*} L(\cdot) = 0 \iff (1 - \tau) W_i - \alpha_i^* - \zeta_i^* = 0 \quad i = 1, 2, \ldots, n.
\]
The Karush-Kuhn-Tucker (KKT) conditions (Mangasarian, 1969) imply that

\[ \alpha_i \left[ \frac{\epsilon}{\tau} + \xi_i - \tilde{Y}_i + \beta + \sum_{j=1}^{n} c_j K(X_j, X_i) \right] = 0, \]

\[ \alpha_i^{*} \left[ \frac{\epsilon}{1-\tau} + \xi_i^{*} + \tilde{Y}_i - \beta - \sum_{j=1}^{n} c_j K(X_j, X_i) \right] = 0, \]

\[ \zeta_i \xi_i = 0, \ zeta_i^{*} \xi_i^{*} = 0. \]

From the above equations, the following results are obtained:

\[ \sum_{i=1}^{n} (\alpha_i - \alpha_i^{*}) = 0, \ c = \frac{1}{\lambda} (\alpha - \alpha^{*}), \]

\[ 0 \leq \alpha_i \leq \tau W_i, \ 0 \leq \alpha_i^{*} \leq (1-\tau)W_i \quad i = 1, 2, \ldots, n. \]

Using the above results, the Lagrangian function can be written as follows:

\[ L(\alpha, \alpha^{*}) \]

\[ = \sum_{i=1}^{n} \left( \tilde{Y}_i - \frac{\epsilon}{\tau} \right) \alpha_i - \sum_{i=1}^{n} \left( \tilde{Y}_i + \frac{\epsilon}{1-\tau} \right) \alpha_i^{*} - \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^{*}) K(X_i, X_j) (\alpha_j - \alpha_j^{*}). \]

Hence, the dual form is given by

\[ \max_{\alpha, \alpha^{*}} \left\{ \sum_{i=1}^{n} \left( \tilde{Y}_i - \frac{\epsilon}{\tau} \right) \alpha_i - \sum_{i=1}^{n} \left( \tilde{Y}_i + \frac{\epsilon}{1-\tau} \right) \alpha_i^{*} - \frac{1}{2\lambda} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^{*}) K(X_i, X_j) (\alpha_j - \alpha_j^{*}) \right\} \]

subject to \( \sum_{i=1}^{n} (\alpha_i - \alpha_i^{*}) = 0, \ 0 \leq \alpha_i \leq \tau W_i, \) and \( 0 \leq \alpha_i^{*} \leq (1-\tau)W_i \) for \( i = 1, 2, \ldots, n. \)

From the above dual problem, \( \alpha \) and \( \alpha^{*} \) can be found. The KKT conditions state that \( \zeta_i > 0 \) and \( \xi_i = 0 \) for the \( i \)-th sample with \( 0 < \alpha_i < \tau W_i. \) Hence,

\[ \frac{\epsilon}{\tau} - \tilde{Y}_i + \beta + \frac{1}{\lambda} \sum_{j=1}^{n} (\alpha_j - \alpha_j^{*}) K(X_j, X_i) = 0 \]

for the \( i \)-th sample with \( 0 < \alpha_i < \tau W_i. \) Similarly,

\[ \frac{\epsilon}{1-\tau} + \tilde{Y}_i - \beta - \frac{1}{\lambda} \sum_{j=1}^{n} (\alpha_j - \alpha_j^{*}) K(X_j, X_i) = 0 \]
for the $i$-th sample with $0 < \alpha_i^* < (1 - \tau)W_i$. Hence, $\hat{\beta}$ can be defined as the average of the following values:

$$-\frac{\epsilon}{\tau} + \hat{Y}_i - \frac{1}{\lambda} \sum_{j=1}^{n} (\alpha_j - \alpha_j^*)K(X_j, X_i) \quad \text{with} \quad 0 < \alpha_i < \tau W_i,$$

$$\frac{\epsilon}{1 - \tau} + \hat{Y}_i - \frac{1}{\lambda} \sum_{j=1}^{n} (\alpha_j - \alpha_j^*)K(X_j, X_i) \quad \text{with} \quad 0 < \alpha_i^* < (1 - \tau)W_i.$$

Using $\alpha$ and $\alpha^*$ from the solution of the dual problem and the above $\hat{\beta}$, the estimated quantile function $\hat{f}_\lambda$ can be written as follows:

$$\hat{f}_\lambda(x) = \hat{\beta} + \sum_{j=1}^{n} c_j K(X_j, x) = \hat{\beta} + \frac{1}{\lambda} \sum_{j=1}^{n} (\alpha_j - \alpha_j^*)K(X_j, x). \quad (5)$$

Note that the dual variables $\alpha_j$ and $\alpha_j^*$ cannot be nonzero at the same time, and that $0 \leq \alpha_j \leq \tau W_j$ and $0 \leq \alpha_j^* \leq (1 - \tau)W_j$. From the KKT conditions, it can be easily verified that:

(i) $\alpha_j = \tau W_j$, $0 \leq \alpha_j \leq \tau W_j$, and $\alpha_j = 0$ when $(X_j, \hat{Y}_j)$ is located above, on, and below the upper boundary of the $\epsilon$-insensitive loss zone, respectively; and,

(ii) $\alpha_j^* = 0$, $0 \leq \alpha_j^* \leq (1 - \tau)W_j$, and $\alpha_j^* = (1 - \tau)W_j$ when $(X_j, \hat{Y}_j)$ is located above, on, and below the lower boundary of the $\epsilon$-insensitive loss zone, respectively,

where the upper and lower boundaries of the $\epsilon$-insensitive loss zone are \{(x, y) : y = \hat{f}_\lambda(x) + \frac{\epsilon}{\tau}\} and \{(x, y) : y = \hat{f}_\lambda(x) - \frac{\epsilon}{1 - \tau}\}. Hence, the coefficient $c_j$ of the estimated quantile function in Eq. (5) has the following properties:

$$c_j = \frac{1}{\lambda} \tau W_j \quad \text{if} \quad \hat{Y}_j > \hat{f}_\lambda(X_j) + \frac{\epsilon}{\tau},$$

$$0 \leq c_j \leq \frac{1}{\lambda} \tau W_j \quad \text{if} \quad \hat{Y}_j = \hat{f}_\lambda(X_j) + \frac{\epsilon}{\tau},$$

$$c_j = 0 \quad \text{if} \quad \hat{f}_\lambda(X_j) - \frac{\epsilon}{1 - \tau} < \hat{Y}_j < \hat{f}_\lambda(X_j) + \frac{\epsilon}{\tau},$$

$$\frac{1}{\lambda} (1 - \tau)W_j \leq c_j \leq 0 \quad \text{if} \quad \hat{Y}_j = \hat{f}_\lambda(X_j) - \frac{\epsilon}{1 - \tau},$$

$$c_j = -\frac{1}{\lambda} (1 - \tau)W_j \quad \text{if} \quad \hat{Y}_j < \hat{f}_\lambda(X_j) - \frac{\epsilon}{1 - \tau}.$$

The region between the lower and upper boundaries of the $\epsilon$-insensitive loss zone widens as $\epsilon$ increases. Hence, more sample points are included in the $\epsilon$-insensitive loss zone for larger $\epsilon$, and the coefficients $c_j$ of those samples are equal to zero. Therefore, the $\epsilon$-insensitive loss has the effect of shrinking the coefficients $c_j$ toward zero. The shrinkage effect can reduce the variability of the quantile estimates. On the other hand, $\epsilon$-insensitive loss can induce bias because the minimizer of the $\epsilon$-insensitive loss may not be the true
τ-th quantile. Hence, ϵ has a typical bias-variance tradeoff effect. For an appropriate ϵ, the variance reduction can outweigh the increased bias. As demonstrated in Fig. 2 of Park and Kim (2011) for complete data, the performance of estimated quantiles improves up to a certain value of ϵ, and it worsens after that point. This pattern is also observed in simulation studies for LTRC data in Sec. 4.

For the optimal value of ϵ, the simulation studies in Sect. 4.2 show that the estimates usually perform best when ϵ has a value between one-tenth and one-fourth of the standard deviation of the response. Moreover, Park and Kim (2011) argued that ϵ needs to be close to 10% of the standard deviation of Y. In principle, we can use model selection criteria such as cross-validation to choose the optimal ϵ.

Another advantage of the ϵ-insensitive loss is sparsity. Note that the estimated quantile function can be written as a linear combination of kernels at the sample points:

\[
\hat{f}_\lambda(x) = \beta + \sum_{j=1}^{n} c_j K(X_j, x).
\]

The number of nonzero coefficients \(c_j\) decreases as ϵ increases. Sparsity plays an important role in statistics and machine learning (Friedman et al., 2004).

3. Selection of the Smoothing Parameter

The estimated quantile function \(\hat{f}_\lambda(\cdot)\) depends on the smoothing parameter \(\lambda\). The estimated function \(\hat{f}_\lambda(\cdot)\) will overfit the data if \(\lambda\) is too small and \(\hat{f}_\lambda(\cdot)\) will underfit the data if \(\lambda\) is too large. Hence, the optimal smoothing parameter that best balances the goodness of fit to the data and complexity of the model must be selected.

There are several methods for selecting the smoothing parameter. The popular criteria are the Bayesian information criterion (BIC) and GACV. Yuan (2006) demonstrated that GACV is comparable with the other commonly used measures such as BIC and the approximate cross-validation (ACV). In this study, the GACV is used to select the smoothing parameter.

For complete data \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n),\) the ACV is given by

\[
ACV(\lambda) = \sum_{i=1}^{n} \rho_\tau(Y_i - \hat{f}_\lambda(X_i)) / \left(1 - \frac{\partial \hat{f}_\lambda(X_i)}{\partial Y_i}\right),
\]

and the GACV is defined as the average (or sum) of \((1 - \frac{\partial \hat{f}_\lambda(X_i)}{\partial Y_i})\) in the denominator of the ACV, as follows:

\[
GACV(\lambda) = \sum_{i=1}^{n} \rho_\tau(Y_i - \hat{f}_\lambda(X_i)) / n - \sum_{i=1}^{n} \frac{\partial \hat{f}_\lambda(X_i)}{\partial Y_i}
\]

(see Li, et al., 2007).

For the censored and truncated data \((X_i, \tilde{Y}_i, \delta_i, T_i)\), the following weighted check loss is used:

\[
\sum_{i=1}^{n} W_i \rho_\tau(\tilde{Y}_i - \hat{f}_\lambda(X_i))
\]
with $W_i = \frac{1}{F_{\tau}(b)} \delta_i I(a \leq \tilde{Y}_i \leq b) \frac{\hat{S}_i(\tilde{Y}_i)}{\hat{F}_i(\tilde{Y}_i)}$. Since $\sum_{i=1}^n W_i \rho(\tilde{Y}_i - f(X_i))$ is a consistent estimator of $E[\rho(\tilde{Y} - f(X)) | a \leq Y \leq b]$ for any function $f(\cdot)$, it is reasonable to modify the GACV (6) by

$$GACV(\lambda) = \frac{\sum_{i=1}^n W_i \rho(\tilde{Y}_i - \hat{f}_\lambda(X_i))}{n - \sum_{i=1}^n \partial \hat{f}_\lambda(X_i) / \partial \tilde{Y}_i}.$$  

Park and Kim (2011) showed that $\sum_{i=1}^n \partial \hat{f}_\lambda(X_i) / \partial Y_i$ is the same as the number of sample points on the boundary of the $\epsilon$-insensitive loss zone. It can be easily verified that the arguments in Park and Kim (2011) also hold for the censored and truncated data. Hence, the divergence measure is given by

$$\sum_{i=1}^n \frac{\partial \hat{f}_\lambda(X_i)}{\partial \tilde{Y}_i} = \# \{ (X_i, \tilde{Y}_i) : \tilde{Y}_i = \hat{f}_\lambda(X_i) + \frac{\epsilon}{\tau}, \delta_i = 1 \text{ or } \tilde{Y}_i = \hat{f}_\lambda(X_i) - \frac{\epsilon}{1-\tau}, \delta_i = 1 \}.$$  

In this article, the GACV (7) with the divergence parameter given in Eq. (8) is used to select the smoothing parameter.

4. Real Data Analysis and Simulation Studies

4.1 Stanford Heart Transplant Data

The Stanford heart transplant data consist of the survival times (days), ages, and mismatch scores of patients after receiving heart transplants. The data are only right censored, so it can be assumed that $T = 0$. In order to determine the weights in the estimation equation in (4), the data are stratified into four groups based on the ages of patients: $\leq 30$, $30–39$, $40–49$, and $\geq 50$ due to the inappropriateness of simple random censorship (Leurgans, 1987; Gross and Lai, 1996).

A key interest in the Stanford heart transplant data is the effect of age on survival time. In order to estimate the quantile functions, the radial basis kernel with the scale parameter $\sigma = 12$ was used. The smoothing parameter was chosen using a grid search between $2^{-15}$ and $2^0$, and $\epsilon$ was set to 0.1.

Standard parametric models usually assume a quadratic relationship between the age and the logarithm of the survival time. Figure 1 presents the regression line of Buckley and James (1979), the median regression line of Zhou (1992), and the median regression function estimated in an RKHS with the $\epsilon$-insensitive loss. In the figure, the symbols ‘+’ and ‘o’ represent censored and uncensored data, respectively. The quadratic regression lines indicate that the survival time increases up to an age near the early thirties, and it decreases after that point. From the estimated median regression function in an RKHS, it can be seen that the median survival time increased rapidly up to an age slightly over 20, and that it does not change significantly between the ages of 20 and 40. Note that this pattern is not revealed readily in the quadratic models.

Figure 2 shows the estimated 25% and 75% quantile functions with the median regression functions. The 25% and 75% quantile functions exhibit patterns similar to those of the median regression function, except that the decreasing points occur later. An interesting point is that the survival time has an asymmetric distribution. The 75% quantiles are closer to the median compared with the 25% quantiles, especially at ages between 20 and 40.
Figure 1. The dashed and dash-dotted lines represent the mean regression function suggested by Buckley and James (1979) and the median regression function by Zhou (1992), respectively. The solid line represents the median regression function estimated in an RKHS.

4.2 Numerical Simulation

In order to evaluate the performance of the suggested estimation method, random samples were generated from the model:

\[ X_i \sim \text{Uniform}(-2, 2), \]
\[ Y_i = 2 + \frac{\sin(\pi X_i)}{\pi X_i} + 0.25 \exp(-0.25(X_i + 1)\epsilon_i). \]

Figure 2. Estimated quantile functions. The solid line represents the median regression function, the lower dashed line represents the 25% quantile function, and the upper dash line represents the 75% quantile function.
Figure 3. Median regression functions for the simulated random sample. The solid line represents the true median regression function, and the dashed line represents the estimated regression function.

\[ C_i = \frac{1}{2} + \frac{1}{4} \zeta_i, \quad T_i = \frac{1}{4} + \frac{1}{4} \nu_i \]

where \( \epsilon_i \sim N(0, 1) \), \( \zeta_i \sim \chi^2_{10} \) and \( \nu_i \sim \chi^2_5 \). This simulation model has non-homogeneous variability. The conditional variance of the survival time decreases as the covariate increases from \(-2\) to \(2\).

Figure 3 shows one simulated data-set. From the generated samples, the data with \( \tilde{Y}_i = \min(Y_i, C_i) < T_i \) were left-truncated and were not used for analysis. Among the 200 generated samples, about 25% were left-truncated, and 25% of the remaining samples were right-censored. In Fig. 3, the censored and uncensored data are denoted by symbols “+” and “o”, respectively. The radial basis kernel with \( \sigma = 0.8 \) was used, \( \epsilon \) was set to 0, and the smoothing parameter was obtained using a grid search method. The dashed line represents the estimated median regression function, and the solid line is the true median regression function. It can be seen that the estimated line is very close to the true line, except in the area where the covariate is less than \(-1\). This may be attributed to the fact that the variability is very high in this area as compared with the other areas.

The simulation studies were repeated in order to investigate the performance of the estimated quantile functions for various \( \tau \) and \( \epsilon \). The performance was measured in terms of the mean squared error (MSE), mean absolute deviation (MAD), and prediction error (PE). These are defined as follows:

\[
\text{MSE} = \frac{1}{1000} \sum_{i=1}^{1000} \left( \widehat{f}_\lambda(X^*_i) - f(X^*_i) \right)^2,
\]

\[
\text{MAD} = \frac{1}{1000} \sum_{i=1}^{1000} \left| \widehat{f}_\lambda(X^*_i) - f(X^*_i) \right|,
\]
Table 1
Mean squared error (MSE), mean absolute deviation (MAD), and prediction error (PE) of the estimated quantile functions

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>MSE</th>
<th>MAD</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>$0.005004(0.0001577)$</td>
<td>$0.05224(0.0007559)$</td>
<td>$0.08533(0.0001579)$</td>
</tr>
<tr>
<td>0.01</td>
<td>$0.004557(0.0001504)$</td>
<td>$0.05001(0.0006620)$</td>
<td>$0.08483(0.0001492)$</td>
</tr>
<tr>
<td>0.02</td>
<td>$0.004499(0.0001441)$</td>
<td>$0.04926(0.0007025)$</td>
<td>$0.08485(0.0001514)$</td>
</tr>
<tr>
<td>0.05</td>
<td>$0.003938(0.0001190)$</td>
<td>$0.04680(0.0006345)$</td>
<td>$0.08456(0.0001439)$</td>
</tr>
<tr>
<td>0.10</td>
<td>$0.004283(0.0001283)$</td>
<td>$0.04924(0.0007383)$</td>
<td>$0.08507(0.0001577)$</td>
</tr>
<tr>
<td>0.00</td>
<td>$0.005540(0.0001626)$</td>
<td>$0.05610(0.0007874)$</td>
<td>$0.07491(0.0001510)$</td>
</tr>
<tr>
<td>0.01</td>
<td>$0.005381(0.0001715)$</td>
<td>$0.05392(0.0007308)$</td>
<td>$0.07454(0.0001474)$</td>
</tr>
<tr>
<td>0.02</td>
<td>$0.005024(0.0001757)$</td>
<td>$0.05259(0.0007499)$</td>
<td>$0.07453(0.0001497)$</td>
</tr>
<tr>
<td>0.05</td>
<td>$0.004637(0.0001380)$</td>
<td>$0.05091(0.0007487)$</td>
<td>$0.07416(0.0001382)$</td>
</tr>
<tr>
<td>0.10</td>
<td>$0.006630(0.0001759)$</td>
<td>$0.06332(0.0008895)$</td>
<td>$0.07594(0.0001604)$</td>
</tr>
<tr>
<td>0.00</td>
<td>$0.008860(0.0002824)$</td>
<td>$0.07001(0.0009751)$</td>
<td>$0.03930(0.0001290)$</td>
</tr>
<tr>
<td>0.01</td>
<td>$0.008808(0.0003205)$</td>
<td>$0.06882(0.0011113)$</td>
<td>$0.03912(0.0001296)$</td>
</tr>
<tr>
<td>0.02</td>
<td>$0.007990(0.0002531)$</td>
<td>$0.06627(0.0010096)$</td>
<td>$0.03879(0.0001186)$</td>
</tr>
<tr>
<td>0.05</td>
<td>$0.012421(0.0003705)$</td>
<td>$0.08781(0.0013233)$</td>
<td>$0.04026(0.0001401)$</td>
</tr>
<tr>
<td>0.10</td>
<td>$0.028773(0.0007507)$</td>
<td>$0.13670(0.0018674)$</td>
<td>$0.04518(0.0002142)$</td>
</tr>
</tbody>
</table>

$$PE = \frac{1}{1000} \sum_{i=1}^{1000} \rho_{\tau}(Y_i^* - \hat{f}_{\lambda}(X_i^*)),$$

where $X_i^*$ and $Y_i^*$ are 1000 newly generated samples from the model. Note that the original check loss ($\rho_{\tau}$) was used to evaluate the performance, even though the $\epsilon$-insensitive check loss was used to estimate the quantile functions. The process was repeated 400 times in order to compute the averages of MSE, MAD, and PE, and their standard errors. The performance of the estimated quantile function was investigated for $\tau = 0.5$, 0.7, and 0.9. Various $\epsilon$ values were used in order to evaluate the effect of the $\epsilon$-insensitive check loss ($\epsilon = 0.00, 0.01, 0.02, 0.05,$ and 0.1).

Table 1 presents the results of the Monte Carlo simulations. The table shows that the performance of the estimated quantile improves as $\epsilon$ increases up to 0.02 or 0.05, and it worsens after that point. The table indicates that the estimated quantiles have the best performance when $\epsilon$ is set to 0.05 for $\tau = 0.5, 0.7$ and $\epsilon$ is set to 0.02 for $\tau = 0.9$. As demonstrated by Park and Kim (2011), the estimated quantile function performed better when $\epsilon$ is close to one-tenth of the standard deviation of $Y$. In our simulation model, the conditional variance of $Y$ given $X$ is $\frac{1}{4} \exp(-\frac{1}{4}(X + 1))$, and $X$ is uniformly distributed from $-2$ to 2. The standard deviation of $Y$ ranges from $\frac{1}{4}e^{-0.75}$ to $\frac{1}{4}e^{0.25}$, and the average standard deviation of $Y$ can be calculated by

$$\frac{1}{4} \int_{-2}^{2} \frac{1}{4} \exp \left( -\frac{1}{4}(x + 1) \right) \, dx = \frac{1}{4}[e^{0.25} - e^{-0.75}],$$
which is approximately 0.203. Hence, in our simulation studies for LTRC data, the estimated quantiles perform best when $\epsilon$ has a value between one-tenth and one-fourth of the standard deviation of $Y$.

5. Concluding Remarks

In this article, a method for estimating quantile functions was proposed for LTRC data in an RKHS. The proposed estimation method is based on a non-iterative and weighted algorithm. An important issue in regularized estimation is the selection of the smoothing parameter. A modified version of GACV was employed as the criterion for smoothing parameter selection.

As a measure of the goodness of fit of quantiles, the check loss and $\epsilon$-insensitive check loss were used, and they were compared using simulation studies. The studies showed that the $\epsilon$-insensitive loss produces better estimates when $\epsilon$ has a value between one-tenth and one-fourth of the standard deviation of the response. This is attributable to the fact that the $\epsilon$-insensitive loss shrinks the estimated model coefficients toward zero. This shrinkage effect leads to less variability in the prediction, and it can improve the overall accuracy of the prediction as demonstrated in the simulation studies.

Recently, algorithms to fit the entire solution path (Hastie et al., 2004) for every value of the smoothing parameter $\lambda$ have been developed for a regularized estimation method. Although the grid search method was used to select the optimal smoothing parameter in this study, the solution path analysis needs to be investigated further to observe the effect of the smoothing parameter.

Funding

This work was supported by the INHA UNIVERSITY Research Grant.

References