Two-piece distribution based semi-parametric quantile regression for right censored data

Worku Biyadgie Ewnetu[1], Irène Gijbels[2] and Anneleen Verhasselt[1*]

[1]Center of Statistics, Data Science Institute, Hasselt University, Agoralaan D, Diepenbeek, 3590, Belgium.
[2]Department of Mathematics, KU Leuven, Celestijnenlaan 200 B, Leuven, 3001, Belgium.

*Corresponding author(s). E-mail(s): anneleen.verhasselt@uhasselt.be;
Contributing authors: workubiyadgie.ewnetu@uhasselt.be; irene.gijbels@kuleuven.be;

Abstract

Widely used methods such as Cox proportional hazards, accelerated failure time, and Bennet proportional odds models do not model the quantiles directly, but rather allow to assess the influence of the covariates only on the location of the distribution. Quantile regression allows to assess the effects of covariates, not only on a location parameter (such as a mean or median) but also on specific percentiles of the conditional distribution. In recent years, a large family of flexible two-piece asymmetric distributions where the location parameter coincides with a specific quantile of the distribution has been studied. In a conditional (regression) setting the use of such a family of two-piece asymmetric distributions has only been investigated in the complete data case in the literature. In this paper, we propose a semi-parametric procedure to estimate the conditional quantile curves of two-piece asymmetric distributions based on right censored survival data. We use a local likelihood estimation technique in a multi-parameter functional form, via which the effect of a covariate on the location, scale, and index of the conditional survival distribution can be assessed. The finite sample performance of the estimators is investigated via simulations, and the methodology is illustrated on real data examples.

Keywords: local likelihood, quantile regression, right random censoring, survival analysis, two-piece distributions

1 Introduction

In lifetime regression analysis we are mostly interested to investigate the association of the lifetime of patients or study objects with a set of covariates. In the spirit of this and related aims, several statistical models have been developed including the widely used Cox proportional hazards model [10], Bennett proportional odds model [4] and accelerated failure time (AFT) model [40]. These widely used methods do not model the quantiles directly, but rather allow to assess the influence of the covariates on the location of the distribution; in a sense that they do not allow to assess the effect of the covariates on the scale as well as shape of the conditional distribution. Conditional quantile functions on the other hand completely characterize the conditional distribution. There have been numerous studies to investigate the effect of covariates on a censored survival outcome via quantile regression models [5, 6, 9, 11, 30, 31, 39] among others. With a quantile regression approach one can study whether a covariate has a varying or constant effect across quantiles. Fitting a quantile regression model to lifetime data offers flexibility in exploring the effect of covariates on a specific percentile of the distribution as well as to estimate covariate effects on the scale and shape of the whole conditional survival distribution [27]. The concept of quantile regression (in both censored and uncensored data) has evolved over the years for a number of reasons: it is equivariant to monotonic transformations; quantiles are less sensitive to outlying observations; the inference takes into account the entire shape of the conditional distribution; and it is easy to draw a direct interpretation.

For fixed censoring schemes early papers on censored quantile regression include Powell [32, 33]. Under a random censoring scheme, Portnoy [31] analysed censored survival data using quantile regression through a weighted Kaplan-Meier approach. Similarly, Peng and Huang [30] studied quantile regression for randomly censored survival data by a Nelson-Aalen type estimation approach. However, both Portnoy [31] and Peng and Huang [30] methods are restricted by a strong assumption that the entire quantiles are linear in the covariates. Wang and Wang [39] casted doubt on the global linearity assumption, and relaxed the method using a locally linear quantile regression approach. This latter study dealt with a locally weighted censored quantile regression by estimating the censoring probability non-parametrically using a Kaplan-Meier approach. Fully parametric quantile regression for right censored data has been studied by Bottai and Zhang [5]. Their method also imposes a linearity assumption for the covariate effects and the error term follows an asymmetric Laplace distribution with a constant scale parameter. Recently, De Backer et al. [11] proposed minimum distance type estimation in linear censored quantile regression by circumventing the check-based modelling approach.

Together the aforesaid studies provide important insights into the flexibility of quantile regression to assess covariate effects on the survival times. However, most of the studies mentioned remain narrow in focus relying on a global linear assumption (e.g., Portnoy [31] and Peng and Huang [30]) as well as on a specific quantile linearity assumption (e.g., Wang and Wang [39]). In contrast to the literature (e.g., Rubio and Yu [35]), our method to conditional quantile is based on a multi-parameter regression framework where more than one distributional parameter can depend on the covariates. The linearity assumption on the covariate effects is relaxed and we add further distributional flexibility (see Section 2). Our approach is inspired by Gijbels et al. [21] in the area of complete data analysis. They proposed semi-parametric quantile regression using a rich family of asymmetric distributions, where the location parameter coincides with a particular quantile. Modelling survival data in a multi-parameter regression framework is not new in the literature; for example, Anderson [2] proposed a log-linear model where both the location and dispersion parameters depend on the covariates which can be seen as an extension of the classical AFT model. More recently, Burke and MacKenzie [7] studied a multi-parameter regression model for survival data by relaxing the classical proportional hazards assumption.

The paper is organized as follows. The definition of the family of two-piece asymmetric distributions and its basic properties in the conditional setting are described in Section 2. Section 3 briefly explores the semi-parametric estimation problem when the unknown parameters as well as quantile function are conditioned by both univariate and multivariate covariates. The asymptotic properties for the proposed estimators are provided in Section 4. Section 5 is devoted to an extensive simulation study conducted to investigate the finite sample performance of the proposed method. The use of the proposed methodology in real life data examples is illustrated in Section 6. Part of the simulation results and real data analysis are reported in the Supplementary material.

2 Two-piece asymmetric distributions

Let T be a lifetime random variable with support the positive real line $\mathbb{R}^+$. This can be duration of time until the event of interest occurs such as in biological organisms or failure time such as in mechanical systems. Consider an increasing, differentiable and invertible function $g\colon]0,+\infty[\to \mathbb{R}$ (with inverse g^{-1}). Let $T = g^{-1}(Z)$, where Z is a random variable with support in $\mathbb{R}$, and probability density function

$$\widetilde{f}_\alpha(z;\ \mu,\phi) = \frac{2\alpha(1-\alpha)}{\phi}\begin{cases} f_0\big\{(1-\alpha)\big(\dfrac{\mu-z}{\phi}\big)\big\} & \text{if } z<\mu \\ f_0\big\{\alpha\big(\dfrac{z-\mu}{\phi}\big)\big\} & \text{if } z\geq\mu, \end{cases} \tag{1}$$

where $f_0(\cdot)$ denotes a unimodal symmetric density with mean zero and variance one, $\mu \in \mathbb{R}$ and $\phi \in \mathbb{R}^+$ are, respectively, the location and scale parameters for $Z = g(T)$, and $\alpha \in (0, 1)$ is an index parameter controlling the allocation mass of the distribution to each side of μ. Gijbels et al. [20] named (1) a quantile based asymmetric (QBA) density since the αth quantile of Z equals the location parameter μ. The two-piece types of distributions go back to Fechner [16]. See also Wallis [38].

We denote the unconditional probability density and cumulative distribution function for the random variable T by $f_\alpha(t;\ \eta, \phi)$ and $F_\alpha(t;\ \eta, \phi)$, respectively. The probability density of $T = g^{-1}(Z)$ with parameters $\boldsymbol{\theta} = (\eta, \phi, \alpha)^T$ and link function $g(t)$ is obtained from (1), and given by

$$f_\alpha(t;\ \eta, \phi) = \frac{2\alpha(1-\alpha)g'(t)}{\phi} \begin{cases} f_0\big\{(1-\alpha)\big(\dfrac{g(\eta)-g(t)}{\phi}\big)\big\} & \text{if } t < \eta \\ f_0\big\{\alpha\big(\dfrac{g(t)-g(\eta)}{\phi}\big)\big\} & \text{if } t \geq \eta, \end{cases} \tag{2}$$

where $\eta = g^{-1}(\mu) \in \mathbb{R}^+$ (is the location parameter for T) and $g'(t)$ is the first derivative of $g(t)$. Hereafter, we call (2) a two-piece asymmetric (TPA in short) density and f_0 a reference density with its corresponding cumulative distribution F_0, survival S_0 and hazard h_0 functions.

Recently, the unconditional distributions defined in (2) have been studied by Ewnetu et al. [13], Rubio and Hong [34] for random right censored data. Ewnetu et al. [13] used a maximum likelihood estimation (MLE) and established the asymptotic properties of the estimators with an explicit expression of the Fisher information matrix.

By allowing the parameters of $\boldsymbol{\theta}$ to depend on the covariate X, the conditional TPA density of T given a covariate value $X = x$ is written as

$$\begin{aligned} &f_{\alpha(x)}(t;\ \boldsymbol{\theta}(x)) \\ &= \frac{2\alpha(x)\{1-\alpha(x)\}g'(t)}{\phi(x)} \begin{cases} f_0\big\{(1-\alpha(x))\big(\dfrac{g(\eta(x))-g(t)}{\phi(x)}\big)\big\} & \text{if } t < \eta(x) \\ f_0\big\{\alpha(x)\big(\dfrac{g(t)-g(\eta(x))}{\phi(x)}\big)\big\} & \text{if } t \geq \eta(x), \end{cases} \end{aligned}$$

where $\boldsymbol{\theta}(x) = \{\eta(x), \phi(x), \alpha(x)\}^T$. We assume a random covariate X with a probability density function $f_X(\cdot)$ and its support denoted by $\text{supp}(f_X)$. The conditional cumulative distribution of T given $X = x$ is given by

$$F_{\alpha(x)}(t;\ \boldsymbol{\theta}(x)) = \begin{cases} 2\alpha(x)F_0\big\{(1-\alpha(x))\big(\frac{g(t)-g(\eta(x))}{\phi(x)}\big)\big\} & \text{if } t < \eta(x) \\ 2\alpha(x) - 1 + 2(1-\alpha(x))F_0\big\{\alpha(x)\big(\frac{g(t)-g(\eta(x))}{\phi(x)}\big)\big\} & \text{if } t \geq \eta(x), \end{cases} \tag{3}$$

and the conditional quantile of order $\tau \in (0,1)$ is given by

$$\begin{aligned} Q_\tau(T \mid x) \equiv \Big\{F_{\alpha(x)}(t;\ \boldsymbol{\theta}(x))\Big\}^{-1} &= \inf\{t : F_{\alpha(x)}(t;\ \boldsymbol{\theta}(x)) \geq \tau\} \\ &= g^{-1}\Big(g\big(\eta(x)\big) + \phi(x) \cdot C_{\alpha(x)}(\tau)\Big), \quad (4) \end{aligned}$$

where $g^{-1}(\cdot) > 0$ and

$$\begin{aligned} C_{\alpha(x)}(\tau) &= \frac{1}{1-\alpha(x)} F_0^{-1}\left(\frac{\tau}{2\alpha(x)}\right)\mathbb{I}(\tau < \alpha(x)) \\ &+ \frac{1}{\alpha(x)} F_0^{-1}\left(\frac{1+\tau-2\alpha(x)}{2(1-\alpha(x))}\right)\mathbb{I}(\tau \geq \alpha(x)), \quad (5) \end{aligned}$$

whereas $F_0^{-1}(\cdot)$ is the quantile function of the reference distribution. In general, $C_{\alpha(x)}(\tau)$ is an increasing function of τ for a given covariate value $X = x$, resulting in a monotone increasing conditional quantile function $Q_\tau(T \mid x)$. In particular when $\tau = \alpha(x)$, $C_{\alpha(x)}(\tau) = 0$, and hence $Q_\tau(T \mid x) = \eta(x)$, indicating that the conditional quantile of order $\alpha(x)$ coincides with the location function given $X = x$. The conditional survival function is

$$S_{\alpha(x)}(t;\ \boldsymbol{\theta}(x)) = \begin{cases} 1 - 2\alpha(x) S_0\big\{(1-\alpha(x))\big(\frac{g(\eta(x))-g(t)}{\phi(x)}\big)\big\} & \text{if } t < \eta(x) \\ 2(1-\alpha(x)) S_0\big\{\alpha(x)\big(\frac{g(t)-g(\eta(x))}{\phi(x)}\big)\big\} & \text{if } t \geq \eta(x), \end{cases} \quad (6)$$

where $S_0(\cdot)$ is the survival probability function for the reference density.

To summarize, we assume $f_0(\cdot)$ to be known while the distributional parameters $\boldsymbol{\theta}$ are unknown smooth functions of the covariate X.

In the approach above, we start from a parametric family of distributions and then subsequently allow the parameters to be unknown functions of the covariate. Similar approaches have been extensively applied in the literature, for example in [19] and [3]. See also [26] for a recent review on what can be called distributional regression.

3 Semiparametric conditional quantile estimation

We are interested in the survival time T, however in real data applications, the variable of interest T is not always observed or it may be subject to different censoring schemes. In the presence of right random censoring one would only observe the couple $\big(\min\{T, C\}, \mathbb{I}(T \leq C)\big) = (Y, \Delta)$, where C is a random censoring variable with distribution G, Y is the observed response variable, and $\Delta = \mathbb{I}(T \leq C)$ is the censoring indicator which equals one if $T \leq C$ and zero else. We consider two crucial assumptions that are commonly used in standard survival models for censored data:

- *Conditionally independent censoring*: the survival time T and the censoring time C are conditionally independent given X.
- *Non-informative censoring*: the distribution of the censoring time C given $X = x$ does not give any information about the distribution of the survival time T.

In the unconditional context the assumptions of independence and non-informative censoring are classical. See for example [25].

Numerous studies have used likelihood methods to estimate the parameters in the distribution of T subject to the above two basic assumptions. Let $(x_i, y_i, \delta_i),\ \ i = 1, \dots, n$ be a realized i.i.d. sample from (X, Y, Δ). In the presence of right censoring, the conditional likelihood function is proportional to

$$\begin{aligned} \text{lik}_n(\boldsymbol{\theta}(\cdot)) &\propto \prod_{i=1}^{n_u} f_{\alpha(x_i)}(y_i;\ \boldsymbol{\theta}(x_i)) \prod_{i=1}^{n_c} S_{\alpha(x_i)}(y_i;\ \boldsymbol{\theta}(x_i)) \\ &= \prod_{i=1}^{n} \left\{ f_{\alpha(x_i)}(y_i;\ \boldsymbol{\theta}(x_i)) \right\}^{\delta_i} \left\{ S_{\alpha(x_i)}(y_i;\ \boldsymbol{\theta}(x_i)) \right\}^{1-\delta_i}, \end{aligned} \tag{7}$$

where n_u and n_c denote, respectively, the number of uncensored and censored observations. As we can see easily from (7), the contribution of an uncensored $(\delta_i = 1)$ observation in the conditional likelihood function is the conditional density while for a censored $(\delta_i = 0)$ observation it is the conditional survival probability.

If $\eta(x), \phi(x)$ and $\alpha(x)$ have a specified parametric functional form of the covariate, for example a polynomial with degree p_1 for $\eta(x) = (\eta_0 + \eta_1 x + \ldots + \eta_{p_1} x^{p_1})$, with degree p_2 for $\phi(x) = (\phi_0 + \phi_1 x + \ldots + \phi_{p_2} x^{p_2})$, and degree p_3 for $\alpha(x) = (\alpha_0 + \alpha_1 x + \ldots + \alpha_{p_3} x^{p_3})$, maximizing the likelihood function of (7) with respect to $\eta(x), \phi(x)$ and $\alpha(x)$ leads to a fully parametric estimation with estimator $\widehat{\eta}(x), \widehat{\phi}(x)$ and $\widehat{\alpha}(x)$, respectively. In most real applications, however, specifying a parametric form of these functions may be difficult and becomes too restrictive. In these circumstances a local weighted version of (7), the so-called *local kernel weighted conditional log-likelihood* may be used to estimate the unknown parameter functions. The main idea of using a local fitting approach is to estimate the unknown smooth functions non-parametrically within a local window around a value x in a grid. This technique provides estimates not only for the unknown smooth function but also for its derivatives.

A local likelihood technique has been applied to a generalized linear model and the proportional hazards model of Cox [10] by Tibshirani and Hastie [37]. The motivation to use local likelihood and its applications in a variety of fields including survival analysis with censored data has been studied by Fan and Gijbels [14]. A number of studies have also applied local linear fitting for right censored data in non-parametric estimation problems, see for example, Fan and Gijbels [15], Kim et al. [24] and Yu and Jones [42]. Gannoun et al. [18] studied a local linear approach in non-parametric conditional quantile estimation for censored data. Most of the papers have been focusing on a local likelihood estimation technique with a single unknown smooth function of interest. Limited studies have been carried out on local polynomial estimation in multi-parameter likelihood models for the complete data case, see Aerts and Claeskens [1] and Gijbels et al. [21].

The local likelihood method is simple to formulate by localizing the global likelihood function defined in (7) using a kernel function $K(\cdot)$ (a symmetric density function) and a bandwidth parameter $h > 0$. First, we assume that $\eta(\cdot), \phi(\cdot)$ and $\alpha(\cdot)$ are unknown *smooth* functions of the covariate X and estimate them in a window around each x value. Once we have the estimators $\widehat{\eta}(x), \widehat{\phi}(x)$ and $\widehat{\alpha}(x)$, the conditional quantile of order τ, $Q_\tau(T \mid x)$ can easily be estimated at the second stage from (4) and the estimator is $\widehat{Q}_\tau(T \mid x) = g^{-1}\Big(g\big(\widehat{\eta}(x)\big) \ + \ \widehat{\phi}(x) \cdot C_{\widehat{\alpha}(x)}(\tau)\Big)$, where

$$
\begin{aligned}
C_{\widehat{\alpha}(x)}(\tau) = \frac{1}{1-\widehat{\alpha}(x)} F_0^{-1}\left(\frac{\tau}{2\widehat{\alpha}(x)}\right)\mathbb{I}(\tau < \widehat{\alpha}(x)) \\
+ \frac{1}{\widehat{\alpha}(x)} F_0^{-1}\left(\frac{1+\tau-2\widehat{\alpha}(x)}{2(1-\widehat{\alpha}(x))}\right)\mathbb{I}(\tau \geq \widehat{\alpha}(x)).
\end{aligned}
$$

Note that $\widehat{Q}_\tau(T \mid x) = \widehat{\eta}(x)$ when $\tau = \widehat{\alpha}(x)$, i.e., the estimator of the location function is the $\widehat{\alpha}(x)$th conditional quantile of the distribution. Note that the conditional quantile of $Z = g(T)$ is $\widehat{\mu}(x) + \widehat{\phi}(x) \cdot C_{\widehat{\alpha}(x)}(\tau)$, considering the monotone transformation $\widehat{\mu}(x) = g\big(\widehat{\eta}(x)\big)$. An advantage of the above approach is that the non-crossing property of $\widehat{Q}_\tau(T \mid x)$ inherently holds. Indeed, as can be seen from (4) combined with (5), for $0 < \tau_1 \leq \tau_2 < 1$, the estimator of the conditional quantile curves satisfy, $\widehat{Q}_{\tau_1}(T \mid x) \leq \widehat{Q}_{\tau_2}(T \mid x)$, in the support of the covariate X.

The potential of modelling survival time with the two-piece asymmetric distribution is to allow describing data by a large family of distributions which includes log-symmetric models as special cases (where it is assumed that $g(t) = \ln(t)$).

3.1 Single covariate

Consider for now a unidimensional continuous covariate X. For convenience of optimization, we reparametrize $\theta_1(\cdot) = g\{\eta(\cdot)\}$, $\theta_2(\cdot) = \ln\{\phi(\cdot)\}$ and $\theta_3(\cdot) = \text{logit}\{\alpha(\cdot)\} = \ln\left\{\frac{\alpha(\cdot)}{1-\alpha(\cdot)}\right\}$, which maps the parameter space of $\boldsymbol{\Theta}$, $\mathbb{R}^+ \times \mathbb{R}^+ \times (0,1)$ onto $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, and also guarantees the positivity of $\eta(\cdot)$ and $\phi(\cdot)$ while for $\alpha(\cdot)$ ensuring values in $(0,1)$. Let $\tilde{\boldsymbol{\theta}}(\cdot)$ denote the parameter vector $\big(\theta_1(\cdot),\ \theta_2(\cdot),\ \theta_3(\cdot)\big)^T \in \mathbb{R}^3$. Estimating $\tilde{\boldsymbol{\theta}}(\cdot)$ is then equivalent to estimating $\boldsymbol{\theta}(\cdot) = \big(\eta(\cdot), \phi(\cdot), \alpha(\cdot)\big)^T$. To formulate the local-likelihood function suppose that the unknown function $\theta_r(\cdot)$, $r \in \{1,2,3\}$ has a (p_r+1)th continuous derivative at the point x_0. For the data points x_i in the neighbourhood of x_0

we want to approximate $\theta_r(x_i)$ with a Taylor expansion of degree p_r ($\in \mathbb{N}$):

$$\begin{aligned}\theta_r(x_i) &\approx \theta_r(x_0) + \theta_r^{(1)}(x_0)\big(x_i - x_0\big) + \ldots + \frac{\theta_r^{(p_r)}(x_0)}{p_r!}\big(x_i - x_0\big)^{p_r} \\ &\equiv \sum_{j=0}^{p_r} \theta_{rj}(x_0)(x_i - x_0)^j \\ &\equiv \beta_{r0} \ + \ \beta_{r1}(x_i - x_0) + \ldots + \beta_{rp_r}(x_i - x_0)^{p_r} = \mathbf{x}_{i,p_r}^T \boldsymbol{\beta}_r \end{aligned} \tag{8}$$

where $\approx$ denotes the approximation by ignoring the higher orders in the Taylor expansion, $\mathbf{x}_{i,p_r} = \big(1, (x_i - x_0), (x_i - x_0)^2, \ldots, (x_i - x_0)^{p_r}\big)^T$, the notation $\beta_{rv} \equiv \theta_{rv}(x_0) = \dfrac{\theta_r^{(v)}(x_0)}{v!}$, for $v = 0, 1, \ldots, p_r$, and $\boldsymbol{\beta}_r = (\beta_{r0}, \ldots, \beta_{rp_r})^T$, with the superscript $^{(\nu)}$ denoting the νth derivative. From now on, for given x_0 for simplicity we may use β_{rv} instead of $\theta_{rv}(x_0)$. But it is important to keep in mind the dependence on x_0. Consider the ith observation and denote

$$\mathrm{L}_u(\boldsymbol{\beta};\ y_i, x_i, x_0) = \ln\{f_{\alpha(\mathbf{x}_{i,p_3}^T \boldsymbol{\beta}_3)}(y_i;\ \mathbf{x}_{i,p_1}^T \boldsymbol{\beta}_1, \mathbf{x}_{i,p_2}^T \boldsymbol{\beta}_2, \mathbf{x}_{i,p_3}^T \boldsymbol{\beta}_3)\},$$

$$\text{and } \mathrm{L}_c(\boldsymbol{\beta};\ y_i, x_i, x_0) = \ln\{S_{\alpha(\mathbf{x}_{i,p_3}^T \boldsymbol{\beta}_3)}(y_i;\ \mathbf{x}_{i,p_1}^T \boldsymbol{\beta}_1, \mathbf{x}_{i,p_2}^T \boldsymbol{\beta}_2, \mathbf{x}_{i,p_3}^T \boldsymbol{\beta}_3)\},$$

where $\boldsymbol{\beta} = \big\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3\big\}$. By incorporating the localizing weight $K_h(x - x_0)$ the *local kernel weighted conditional log-likelihood* function for the realized i.i.d. sample is given by

$$\mathcal{L}_n(\boldsymbol{\beta};\ x_0, h) = \sum_{i=1}^{n} \left(\delta_i \mathrm{L}_u(\boldsymbol{\beta};\ y_i, x_i, x_0) \ + \ (1 - \delta_i)\mathrm{L}_c(\boldsymbol{\beta};\ y_i, x_i, x_0)\right) K_h(x_i - x_0), \tag{9}$$

where $K_h(\cdot) = h^{-1}K(\cdot/h)$ is the kernel density function rescaled by the bandwidth parameter h. Henceforth, maximizing $\mathcal{L}_n(\boldsymbol{\beta};\ x_0, h)$ with respect to $\boldsymbol{\beta}$ gives the local kernel weighted maximum likelihood estimators $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\beta}}_3)$ at a local point x_0. That is,

$$\big\{\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\beta}}_3\big\} = \arg\max_{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3} \big\{\mathcal{L}_n(\boldsymbol{\beta};\ x_0, h)\big\}, \tag{10}$$

where $\widehat{\boldsymbol{\beta}}_r = (\widehat{\beta}_{r0}, \ldots, \widehat{\beta}_{rp_r})^T$ and the estimator $\widehat{\theta}_r^{(v)}(x_0)$ for $\theta_r^{(v)}(x_0)$ is $\widehat{\theta}_r^{(v)}(x_0) = v!\widehat{\theta}_{rv}(x_0) = v!\widehat{\beta}_{rv}$. The estimators for $\eta(\cdot), \phi(\cdot), \alpha(\cdot)$ at a local point x_0 are then given as

$$\widehat{\eta}(x_0) = g^{-1}(\widehat{\beta}_{10}), \qquad \widehat{\phi}(x_0) = \exp(\widehat{\beta}_{20}), \qquad \widehat{\alpha}(x_0) = \frac{\exp(\widehat{\beta}_{30})}{1 + \exp(\widehat{\beta}_{30})}. \tag{11}$$

Consequently, the estimator for the τth conditional quantile of the survival time at the local point x_0 is then given by

$$\widehat{Q}_\tau(T \mid x_0) = g^{-1}\Big(g\big(\widehat{\eta}(x_0)\big) \ + \ \widehat{\phi}(x_0) \cdot C_{\widehat{\alpha}(x_0)}(\tau)\Big). \tag{12}$$

The local kernel weighted conditional log-likelihood function defined in (9) is continuous and differentiable with respect to the components of $\boldsymbol{\beta}_r$ except when $y_i = g^{-1}(\theta_1(x_i))$, for $i = 1, \ldots, n$. However, one can use direct or derivative free optimization techniques to maximize it numerically. Note that the maximization problem defined in (10) is for a single x_0 to provide an estimate of $\theta_r(x_0)$. For the purpose of producing estimates of the smooth function $\theta_r(\cdot)$ on its support, the maximization needs to be done for a fine grid of x_0-values in the support of the covariate X.

3.2 Multiple covariates

In the previous section we mainly focused on curve smoothing via a local kernel weighted conditional log-likelihood for the censored data conditional on a single covariate. We shall now extend this type of smoothing technique to a d-dimensional covariate vector $\mathbf{X} = (X_1, \ldots, X_d)^T$. The goal is then fitting a d-dimensional surface to the observed data $(\mathbf{x}_i, y_i, \delta_i), i = 1, 2, \ldots, n$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$. Let $\boldsymbol{x}_0 = (x_{01}, \ldots, x_{0d})^T$ be a d-dimensional vector of local points. We then estimate $\big(\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \theta_3(\mathbf{x})\big)^T$ in an arbitrary point $\boldsymbol{x}_0$ by maximizing the local kernel weighted conditional log-likelihood.

We illustrate this idea with a local linear estimator, i.e., when $p_r = 1$, for $r \in \{1, 2, 3\}$. The Taylor approximation of order one of $\theta_r(\mathbf{x}_i)$ in a neighbourhood of $\boldsymbol{x}_0$ is given by

$$\begin{aligned} \theta_r(\mathbf{x}_i) &\approx \theta_{r0}(\boldsymbol{x}_0) + \sum_{j=1}^{d} \frac{\partial \theta_r(\boldsymbol{x})}{\partial x_j} \mid_{\mathbf{x}=\boldsymbol{x}_0} (x_{ij} - x_{0j}) \\ &\equiv \beta_{r0} + \sum_{j=1}^{d} \beta_{rj}(x_{ij} - x_{0j}) \equiv \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_r^{[d]}, \end{aligned}$$

where $\beta_{r0} = \theta_{r0}(\boldsymbol{x}_0)$, $\beta_{rj} = \theta_{rj}(\boldsymbol{x}_0) \equiv \dfrac{\partial \theta_r(\boldsymbol{x}_0)}{\partial x_j}$, for $j = 1, \ldots, d$; $\tilde{\mathbf{x}}_{i;\,d} = \big(1, (x_{i1} - x_{01}), \ldots, (x_{id} - x_{0d})\big)^T$ and $\boldsymbol{\beta}_r^{[d]} = (\beta_{r0}, \beta_{r1}, \ldots, \beta_{rd})^T$. In such surface fitting, we need to consider a d-dimensional multivariate kernel function $\tilde{K}(\cdot)$ with the properties that $\int \tilde{K}(\boldsymbol{u})d\boldsymbol{u} = 1$ and $\int \boldsymbol{u}\tilde{K}(\boldsymbol{u})d\boldsymbol{u} = 0$, and a positive definite matrix of bandwidth parameters $\boldsymbol{H}$ with determinant $\mid \boldsymbol{H} \mid$. In this case we define a rescaled version of $\tilde{K}$ by $\tilde{K}_{\boldsymbol{H}}(\boldsymbol{u}) = \mid \boldsymbol{H} \mid^{-1/2} \tilde{K}(\boldsymbol{H}^{-1/2}\boldsymbol{u})$.

The multivariable version of (9) for the observed i.i.d. sample is then

$$\mathcal{L}_n(\boldsymbol{\beta}^{[d]};\ \boldsymbol{x}_0, \boldsymbol{H})$$

$$= \sum_{i=1}^{n} \Big(\delta_i \mathrm{L}_u(\boldsymbol{\beta}^{[d]};\ y_i, \mathbf{x}_i, \boldsymbol{x}_0) + (1-\delta_i)\mathrm{L}_c(\boldsymbol{\beta}^{[d]};\ y_i, \mathbf{x}_i, \boldsymbol{x}_0) \Big) \tilde{K}_{\boldsymbol{H}}(\mathbf{x}_i - \boldsymbol{x}_0) \quad (13)$$

where $\boldsymbol{\beta}^{[d]} = \{\boldsymbol{\beta}_1^{[d]}, \boldsymbol{\beta}_2^{[d]}, \boldsymbol{\beta}_3^{[d]}\}$ with

$$\mathrm{L}_u(\boldsymbol{\beta}^{[d]};\ y_i, \mathbf{x}_i, \boldsymbol{x}_0) = \ln\{f_{\tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_3^{[d]}}(y_i;\ \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_1^{[d]}, \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_2^{[d]}, \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_3^{[d]})\}$$

$$\text{and } \mathrm{L}_c(\boldsymbol{\beta}^{[d]};\ y_i, \mathbf{x}_i, \boldsymbol{x}_0) = \ln\{S_{\tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_3^{[d]}}(y_i;\ \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_1^{[d]}, \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_2^{[d]}, \tilde{\mathbf{x}}_{i;\,d}^T \boldsymbol{\beta}_3^{[d]})\}.$$

Accordingly, maximizing (13) with respect to $\boldsymbol{\beta}^{[d]}$ at the local points $\boldsymbol{x}_0$ produces the local linear maximum likelihood estimators $\widehat{\boldsymbol{\beta}}^{[d]} = \{\widehat{\boldsymbol{\beta}}_1^{[d]}, \widehat{\boldsymbol{\beta}}_2^{[d]}, \widehat{\boldsymbol{\beta}}_3^{[d]}\}$, where $\widehat{\boldsymbol{\beta}}_r^{[d]} = (\widehat{\beta}_{r0}, \widehat{\beta}_{r1}, \ldots, \widehat{\beta}_{rd})$. Therefore, the d-variate local linear estimator of $\theta_r(\boldsymbol{x}_0)$ equals $\widehat{\theta}_{r0}(\boldsymbol{x}_0) = \widehat{\beta}_{r0}$, and for its first-order partial derivatives $\widehat{\dfrac{\partial\theta(\boldsymbol{x}_0)}{\partial x_j}} = \widehat{\theta}_{rj}(\boldsymbol{x}_0)$ for $j = 1, \ldots, d;\ r \in \{1,2,3\}$. Finally, we can derive the estimators for the parameter of interests and the conditional quantile function $Q_\tau(T \mid \mathbf{x})$ as in the single covariate case.

An obvious challenge in case of multiple covariates is the presence of the *"curse of dimensionality."* Nonetheless, several methods are reported in the literature to address this issue. The approaches include *additive modelling*, *partial linear modelling*, and *modelling with interactions*. See for example Fan and Gijbels [14] and references therein. For brevity we only elaborate on the approach of partial linear modelling.

3.3 Extension: partially linear modelling

It is also possible to specify some known parametric functional forms for some of the distributional parameters of interest. To be more specific, it could be possible to model the location parametrically while modelling the scale and index as unknown smooth functions. Let $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, be d_1- and d_2-dimensional covariate vectors with $d_1 + d_2 \le 2d$. Note that this allows to have common covariates in $\mathbf{X}_1$ and $\mathbf{X}_2$.

For the purpose of illustration we consider a multivariate covariate in the parametric modelling of the location parameter and a single covariate X_2 in non-parametric scale and index parameter functions. The non-parametric components $\theta_r(X_2)$ are approximated by Taylor expansion as defined by (8) for $r = 2, 3$. The location parameter is parametrically modelled as $\theta_1(\mathbf{X}_1) = \beta_{10} + \beta_{11}X_{11} + \beta_{12}X_{12} + \ldots + \beta_{1(d_1-1)}X_{1(d_1-1)} \equiv \mathbf{X}_1^T\boldsymbol{\beta}_1$, where $\mathbf{X}_1^T = (1, X_{11}, \ldots, X_{1(d_1-1)})$, $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \ldots, \beta_{1(d_1-1)})^T$ is a vector of regression coefficients that needs to be estimated.

Note that we can write an AFT type model for the transformed variable $Z = \ln(T)$

$$Z = \mathbf{X}_1^T\boldsymbol{\beta}_1 \ + \ \exp\{\theta_2(X_2)\}\varepsilon \quad (14)$$

where ε has a distribution defined in (3) with parameters $\mu(x) = 0, \phi(x) = 1, \alpha(x) = \alpha(X_2)$ and $g(t) = t$; a standard version of a two-piece asymmetric distribution. As there are parametric and non-parametric functions at the modelling stage, we call it *semi-parametric* or *partially linear* regression. It allows to investigate non-linear covariate effects while retaining the nice interpretability of the traditional linear regression on some covariates. It is useful to alleviate the loss of precision due to the curse of dimensionality that arises in fully non-parametric smoothing while at the same time reducing the bias due to functional misspecification in a fully parametric model.

Lee [28] studied partially linear quantile regression in complete data analysis. In the area of censored data, Chen and Khan [8] investigated semi-parametric estimation in a partially linear regression model. Further, a recent paper by Bravo [6] studied semi-parametric quantile regression with random censoring in which the parametric and non-parametric components are explicitly incorporated, and the estimation is done with a loss minimization approach. Christou and Akritas [9] explored single index quantile regression for censored data and proposed a non-iterative estimation algorithm for both parametric and non-parametric components. The latter two papers and their methods are discussed in Section 5.2.

Remark 1 *It is clear that the distribution* (3) *reduces to a log-symmetric distribution when taking* $g(t) = \ln(t)$ *and the scale parameter function* $\phi(x) \equiv \exp(\theta_2(x))$ *is fixed constant and the index parameter function* $\alpha(x) = 0.5$ *for all values of* x *and consequently model* (14) *reduces to a classical AFT model in which the random error term follows a log-symmetric distribution. In general, when both the scale and index parameters are constants in model* (14)*, the proposed method provides an exciting opportunity to advance the classical AFT model with a rich family of TPA distributions for the random error term.*

The unknown parametric and non-parametric components in model (14) are estimated by a profile likelihood approach. For the easy of illustration, the procedure is explained for a (univariate) local linear estimator of the scale and index parameter functions (i.e., $p_2 = p_3 = 1,\ d_2 = 1$) and a d_1-dimensional parametric linear model for the location function:

- S0: Get initial values for $\boldsymbol{\beta}_1$, $\theta_2(\cdot)$ and $\theta_3(\cdot)$ using parametric or fully non-parametric censored regression. Let $\widehat{\boldsymbol{\beta}}_1^{\{0\}}$, $\widehat{\theta}_2^{\{0\}}(\cdot)$ and $\widehat{\theta}_3^{\{0\}}(\cdot)$ be the estimates for $\boldsymbol{\beta}_1$, $\theta_2(\cdot)$ and $\theta_3(\cdot)$, respectively.
- S1: Fix $\boldsymbol{\beta}_1$ at the initial value $\widehat{\boldsymbol{\beta}}_1^{\{0\}}$ and estimate the non-parametric components. For given x_0, we estimate $\theta_2(x_0)$ and $\theta_3(x_0)$ and its first order partial derivatives at x_0 by a local linear smoothing approach. This can be done using

$$\left\{\widehat{\boldsymbol{\beta}}_2^{\{1\}}, \widehat{\boldsymbol{\beta}}_3^{\{1\}}\right\} = \underset{\boldsymbol{\beta}_2, \boldsymbol{\beta}_3}{\arg\max}\ \mathcal{L}_n^{\{1\}}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3;\ \widehat{\boldsymbol{\beta}}_1^{\{0\}}, x_0, h), \tag{15}$$

where $\widehat{\boldsymbol{\beta}}_2^{\{1\}} = (\widehat{\beta}_{20}^{\{1\}}, \widehat{\beta}_{21}^{\{1\}})^T$, $\widehat{\boldsymbol{\beta}}_3^{\{1\}} = (\widehat{\beta}_{30}^{\{1\}}, \widehat{\beta}_{31}^{\{1\}})^T$ and $\mathcal{L}_n^{\{1\}}(\boldsymbol{\beta}_2, \boldsymbol{\beta}_3;\ \widehat{\boldsymbol{\beta}}_1^{\{0\}}, x_0, h)$ is the local kernel weighted conditional log-likelihood function of $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$, similar to (9), replacing $\boldsymbol{\beta}_1$ by $\widehat{\boldsymbol{\beta}}_1^{\{0\}}$. The estimators for $\phi(x_0)$ and $\alpha(x_0)$ are then respectively, $\widehat{\phi}^{\{1\}}(x_0) = \exp(\widehat{\beta}_{20}^{\{1\}})$ and $\widehat{\alpha}^{\{1\}}(x_0) = \dfrac{\exp(\widehat{\beta}_{30}^{\{1\}})}{1+\exp(\widehat{\beta}_{30}^{\{1\}})}$, as $\widehat{\theta}_{r0}^{\{1\}}(x_0) \equiv \widehat{\beta}_{r0}^{\{1\}}, r = 2, 3$.

S2: Estimate the parametric component $\boldsymbol{\beta}_1$ by fixing the non-parametric components at their estimates obtained at S1. Maximize the profile likelihood function for $\boldsymbol{\beta}_1$ substituting $\widehat{\phi}^{\{1\}}(x_{2i})$ for $\phi(x_{2i})$ and $\widehat{\alpha}^{\{1\}}(x_{2i})$ for $\alpha(x_{2i})$. That is $\widehat{\boldsymbol{\beta}}_1 = \arg\max_{\boldsymbol{\beta}_1} \mathcal{L}_n^{\{2\}}\big(\boldsymbol{\beta}_1;\ \widehat{\phi}^{\{1\}}(\cdot), \widehat{\alpha}^{\{1\}}(\cdot)\big)$, where

$$\begin{aligned}
&\mathcal{L}_n^{\{2\}}\big(\boldsymbol{\beta}_1;\ \widehat{\phi}^{\{1\}}(\cdot), \widehat{\alpha}^{\{1\}}(\cdot)\big) \\
&= \sum_{i=1}^{n} \Big[\delta_i \ln\{f_{\widehat{\alpha}^{\{1\}}(x_{2i})}(y_i;\ g^{-1}(\mathbf{x}_{1i}^T\boldsymbol{\beta}_1), \widehat{\phi}^{\{1\}}(x_{2i}), \widehat{\alpha}^{\{1\}}(x_{2i}))\} \\
&\quad +(1-\delta_i)\ln\{S_{\widehat{\alpha}^{\{1\}}(x_{2i})}(y_i;\ g^{-1}(\mathbf{x}_{1i}^T\boldsymbol{\beta}_1), \widehat{\phi}^{\{1\}}(x_{2i}), \widehat{\alpha}^{\{1\}}(x_{2i}))\}\Big].
\end{aligned}$$

S3: Update the non-parametric components locally using (15) by substituting $\widehat{\boldsymbol{\beta}}_1$ for $\widehat{\boldsymbol{\beta}}_1^{\{0\}}$. Thus, the final estimators of the scale $\phi(x_0)$ and index $\alpha(x_0)$ are then given as $\widehat{\phi}(x_0) = \exp(\widehat{\beta}_{20})$ and $\widehat{\alpha}(x_0) = \dfrac{\exp(\widehat{\beta}_{30})}{1+\exp(\widehat{\beta}_{30})}$, respectively.

Remark 2 *In the above estimation algorithm we assumed that the unknown scale and index are functions of the same covariate X_2. In case when $\phi(\cdot)$ and $\alpha(\cdot)$ are functions of different sets of covariates the proposed estimation procedures needs to be modified.*

One possible option can be to estimate the two unknown functions separately at different stages. For example, one may split up S1 into two parts, say S1.1 and S1.2: estimate only $\alpha(\cdot)$ in S1.1 and then estimate $\phi(\cdot)$ in S1.2 after substituting $\widehat{\alpha}^{\{1\}}$ for $\alpha(\cdot)$ and $\widehat{\boldsymbol{\beta}}_1^{\{0\}}$ for $\boldsymbol{\beta}_1$ before starting S2. Another option can be to use a multivariate kernel function with a matrix of bandwidth parameters instead of univariate kernels in S1 and S3. In this option we only need to modify the weighting function in the proposed procedures (replacing $K_h(\cdot)$ with $\widetilde{K}_{\boldsymbol{H}}(\cdot)$).

4 Asymptotic properties

We briefly formulate the large sample properties (consistency and asymptotic normality) of the local kernel weighted MLE in the univariate case (from Section 3.1). Ewnetu et al. [13] investigated the asymptotic properties in the case of unconditional survival time data, while Gijbels et al. [21] investigated the asymptotic properties of the local kernel weighted MLE in the complete data case. We exploit the large sample theoretical properties combining the results of Ewnetu et al. [13] and Gijbels et al. [21]. The notations and assumptions are deferred to the Supplementary material (see Section S2).

The first proposition is stating that the expected score function at the true parameter is zero. Note that the score function differs from that in Gijbels et al. [21], as we have a likelihood for censored data.

Proposition 1 *Suppose that R1–R5 (in the Supplementary material) hold. Then the expectation of the score function at the true parameter is zero:* $\mathbb{E}_{Y,\Delta|X}\{\psi_r(Y,\Delta;\ \theta_1(x),\theta_2(x),\theta_3(x))\} = 0,\ r = 1,2,3,$ *where* $\psi_r(Y,\Delta;\ \cdot)$ *is the score function defined in S2.1 (in the Supplementary material).*

The following theorem provides the consistency of the estimator.

Theorem 2 *Suppose that the conditions of Proposition 1 on the conditional distributions of* T *and* C *hold. Assume that the smoothness and design conditions (S1)–(S2) (in the Supplementary material) and the conditions for kernel function and bandwidth (K1) and (K2) (in the Supplementary material) are satisfied. Let* x *be a point in the interior of the support of* f_X. *Then, as* $n \to +\infty$, *there exist solutions* $\widehat{\boldsymbol{\beta}} = \{\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\boldsymbol{\beta}}_3\}$ *of the local kernel weighted conditional log-likelihood equation* $\partial/\partial\beta_{rv}\{\mathcal{L}_n(\boldsymbol{\beta};\ x,h)\} = 0$, *such that* $\widehat{\beta}_{rv}$ *is weakly consistent for estimating* $\beta_{rv} = \theta_{rv}(x)$ *for all* $r \in \{1,2,3\};\ \ v = 0,1,2,\ldots,p_r$.

The next theorem states the asymptotic normality of the estimator. Note that the expression for Σ_x and $\boldsymbol{\Gamma}_x$ (in the Supplementary material) are similar as in Gijbels et al. [21], however the local Fisher information (see the Supplementary material) involved in their expression is now based on our local kernel weighted conditional log-likelihood for censored data.

Theorem 3 *Let all the conditions of Theorem 2 and (R6) (in the Supplementary material) be satisfied. Then for a point* x *in the interior of supp*(f_X), *as* $n \to \infty$,

$$\left(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Gamma}_x\boldsymbol{\Sigma}_x^{-1}\right)^{-1/2}\Big\{\sqrt{nh}\big[(\widehat{\boldsymbol{\beta}_1} - \boldsymbol{\theta}_1(x))\boldsymbol{H}_{p_1}, (\widehat{\boldsymbol{\beta}_2} - \boldsymbol{\theta}_2(x))\boldsymbol{H}_{p_2}, (\widehat{\boldsymbol{\beta}_3} - \boldsymbol{\theta}_3(x))\boldsymbol{H}_{p_3}\big]^T - \left(\boldsymbol{\Sigma}_x^{-1} - h\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Sigma}_x^{-1}\right)\mathbb{E}[\boldsymbol{W}^n(x)]\Big\} \xrightarrow{\mathcal{D}} \mathcal{N}_{p_1+p_2+p_3+3}(\boldsymbol{0}, \boldsymbol{I}_{p_1+p_2+p_3+3}),$$

where $\boldsymbol{0}$ *is a null vector of dimension* $(p_1+p_2+p_3+3)$ *and* $\boldsymbol{I}_{p_1+p_2+p_3+3}$, *is the identity matrix of dimension* $(p_1+p_2+p_3+3)$ *and the remaining notations are provided in Section S2 (in the Supplementary material).*

Corollary 1 *Assume the conditions of Theorem 3 hold. If the function* $g_r^{-1}(\cdot)$ *is differentiable and* $(g_r^{-1})'(\cdot)$ *is continuous at* $\vartheta_r(x)$, *where* $\vartheta_r(.)$ *is the parameter of interest. Then* $\widehat{\vartheta}_r(x)$ *has the same limiting distribution as* $\widehat{\theta}_r(x)$ *except that its asymptotic bias is divided by* $(g_r^{-1})'(\vartheta_r(x))$ *and asymptotic variance is divided by* $\{(g_r^{-1})'(\vartheta_r(x))\}^2$. *This is especially useful to provide the asymptotic distribution of the estimators of* $\vartheta_1(\cdot) = \eta(\cdot), \vartheta_2(\cdot) = \phi(\cdot)$, *and* $\vartheta_3(\cdot) = \alpha(\cdot)$.

5 Simulation study

We conduct an extensive simulation study with various scenarios to measure the finite sample performance of our estimator. The performance of the estimator depends on various aspects, including: the censoring rate, the sample size, the link function used to generate the survival time, and the nature of the unknown smooth parameter functions.

We split the simulation study into two parts. In the first part, we investigate the performance of the estimators for the unknown smooth functions in a univariate covariate setting. In the second part multiple covariates are incorporated and our estimator is compared with competitors when estimating the conditional quantile function. Even though the optimization is done with the transformed parameter functions, the final results are presented for the estimates of the originally defined parameter functions $\eta(x)$, $\phi(x)$ and $\alpha(x)$. All the computations are done using the statistical software R through the minimization package nloptr (version 1.2.2.2). The R codes are available at https://github.com/Ewnetu-github/tpalocal.git.

5.1 Part I: single covariate

We first investigate the performance of the local kernel weighted maximum likelihood estimator in the univariate case.

5.1.1 Data generating mechanism

We consider two scenarios to generate the data:

- Scenario I: The survival time $T \mid X = x \sim F_{\alpha(x)}(t;\ \boldsymbol{\theta}(x))$ (given in (3)) with a log-link function $g(t) = \ln(t)$ and a standard Laplace distribution for F_0. Details on this conditional distribution and parameter estimation in this setting are given in Section S1 of the Supplementary material.

  The censoring time follows an exponential distribution $C \mid X = x \sim \text{Expo}(\theta_c x)$. We consider two values for $\theta_c > 0$, leading to 20% and 40% of censoring respectively.
- Scenario II: This is the same as Scenario I, except for the link function that is now a logit-link function $g(t) = \ln(e^t - 1)$.

We consider two sample sizes: $n = 100$ and 300. The covariate X has a truncated normal distribution with mean 0 and standard deviation 2. We consider two sets of unknown parameter functions (with $x \in (0, 2)$):

$$\begin{aligned}
\text{Model I:}\quad \theta_1(x) &= x + 2\exp(-16x^2)\\
\theta_2(x) &= \sin(2x) - 3\\
\theta_3(x) &= x^2 - \sin(x) - 2x\\
\text{Model II:}\quad \theta_1(x) &= x + \sin(2x) + 2\exp(-16x^2)\\
\theta_2(x) &= -x - 1.5\\
\theta_3(x) &= \sin(\pi x) + x^2.
\end{aligned}$$

The true location, scale and index parameter functions under these two models are presented in Figure 1. In total we have $2 \times 2 \times 2 \times 2 = 16$ data generating

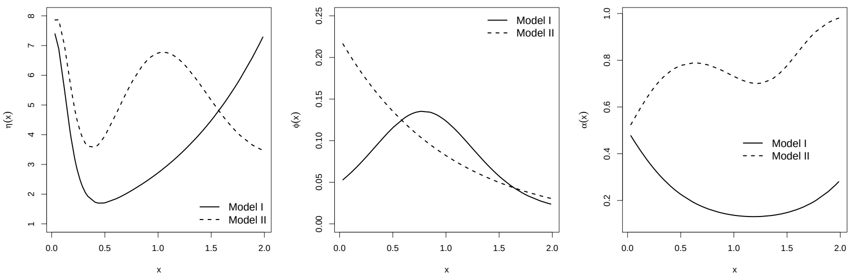


Fig. 1: True curves for the location (left), scale (middle) and index (right) parameter functions under the two simulation models.

mechanisms in this part of the simulation study. We consider 500 Monte Carlo simulations for each of the 16 settings.

5.1.2 Local kernel weighted estimator specifications

We use local linear smoothing for all unknown parameter functions ($p_1 = p_2 = p_3 = 1$) and an Epanechnikov kernel function given by $K(u) = (3/4)\big(1-u^2\big)\mathbb{I}(\mid u \mid\leq 1)$. In order to explore the sensitivity of the proposed method to the bandwidth choice, we consider an equispaced grid of bandwidth values ranging from 0.15 to 0.75 with interval lengths of 0.1. To start the optimization defined in (10) the initial values for β_{r0}'s ($r = 1, 2, 3$) are obtained from the fully parametric maximum likelihood estimates and for β_{r1} we use zero as initial value.

5.1.3 Performance measures

In order to measure the performance of the estimators under each scenario we use the approximate integrated squared error (AISE) for $\widehat{\theta}_r(x)$'s and approximate quantile residual loss (AQRL) for $\widehat{Q}_\tau(T \mid x)$. The AISE is based on a grid of m points (x_j's; we take $m = n$ in the analysis) in the domain of X and defined by AISE $= \frac{R_X}{m}\sum_{j=1}^{m}\big(\widehat{\theta}_r(x_j) - \theta_r(x_j)\big)^2$, $r = 1, 2, 3$, where R_X is the range of the observed X-values and $\widehat{\theta}_r(x_j)$ is the estimator (based on (11)) of the true $\theta_r(x_j)$. The AQRL is defined as

$$\text{AQRL} = \frac{R_X}{m}\sum_{j=1}^{m}\rho_\tau\big(\widehat{Q}_\tau(T \mid x_j) - Q_\tau(T \mid x_j)\big), \tag{16}$$

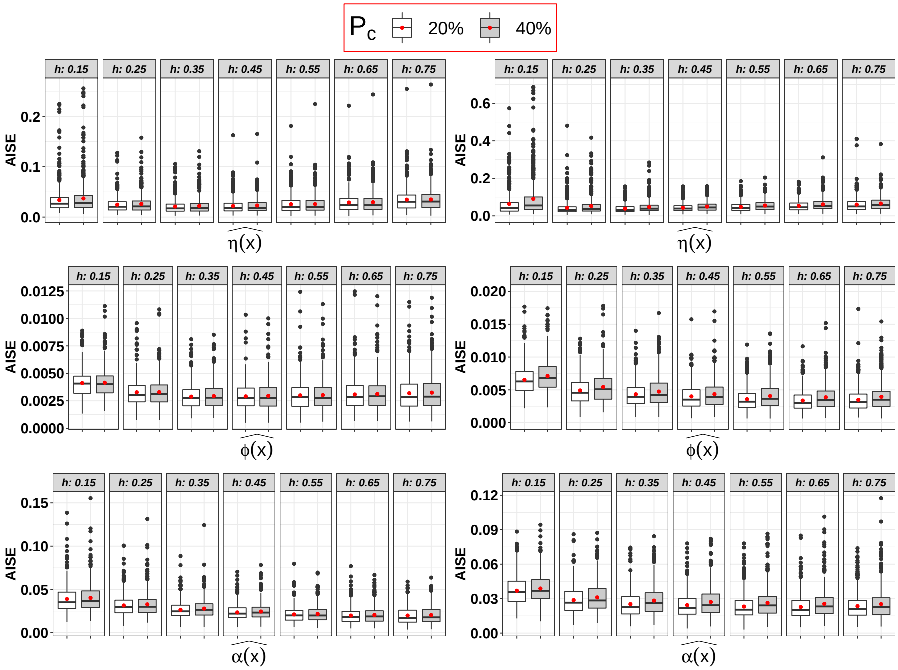


Fig. 2: **Scenario II.** AISE values across seven bandwidth values for the three estimates over 500 simulated samples of size $n = 100$ with 20% and 40% censoring proportions; Model I (left column) and Model II (right column).

where $\rho_\tau(u) = u\big(\tau - \mathbb{I}(u < 0)\big)$ is the quantile loss function. The conditional quantile is estimated for three different values of $\tau = 0.25, 0.5,\ 0.75$: $\widehat{Q}_\tau(T \mid x_j)$ (based on (12)).

5.1.4 Results

The results for Scenario II for $n = 100$ are presented here, while the figures for Scenario II for $n = 300$, and the results for Scenario I can be found in the Supplementary material. Boxplots of the AISE of the estimators over the 500 simulated samples of size $n = 100$ from Scenario II are depicted in Figure 2. The red dot in each boxplot represents the average of AISE over these 500 simulated samples. Tables 1 and 2 present, for respectively Models I and II, the median AISE values for Scenario II for both sample sizes.

It can be noted from Figure 2, Figure S1 in the Supplementary material, and Tables 1 and 2 that the AISE decreases as the sample size increases while it tends to increase with the censoring proportion for all estimates, as expected. Nonetheless, the impact of censoring on the performance of the estimators is very little, in particular for larger sample size. In terms of the bandwidth, we observe that the performances of the three estimators are not so sensitive to the bandwidth values, and that the sensitivity decreases as sample size increases.

Table 1: **Scenario II.** Median AISE values for the estimators $\widehat{\eta}(x)$, $\widehat{\phi}(x)$ and $\widehat{\alpha}(x)$ for Model I, for sample sizes $n = 100$ and $n = 300$.

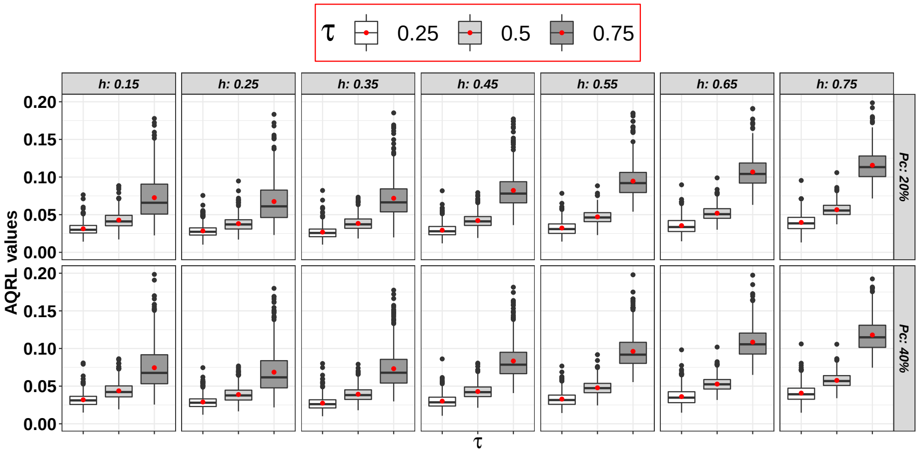
sample size	estimator	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
			bandwidth value h						
$n = 100$	$\widehat{\eta}$	20%	0.027	0.021	0.017	0.018	0.020	0.024	0.031
		40%	0.028	0.021	0.018	0.018	0.020	0.024	0.031
	$\widehat{\phi}$	20%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
		40%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
	$\widehat{\alpha}$	20%	0.035	0.030	0.025	0.022	0.020	0.018	0.017
		40%	0.037	0.030	0.027	0.023	0.020	0.019	0.018
$n = 300$	$\widehat{\eta}$	20%	0.012	0.009	0.009	0.011	0.013	0.015	0.020
		40%	0.012	0.010	0.009	0.011	0.013	0.015	0.020
	$\widehat{\phi}$	20%	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		40%	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	$\widehat{\alpha}$	20%	0.020	0.017	0.016	0.016	0.013	0.012	0.013
		40%	0.020	0.018	0.017	0.016	0.013	0.012	0.013

Table 2: **Scenario II.** Median AISE values for the estimators $\widehat{\eta}(x)$, $\widehat{\phi}(x)$ and $\widehat{\alpha}(x)$ for Model II, for sample sizes $n = 100$ and $n = 300$.

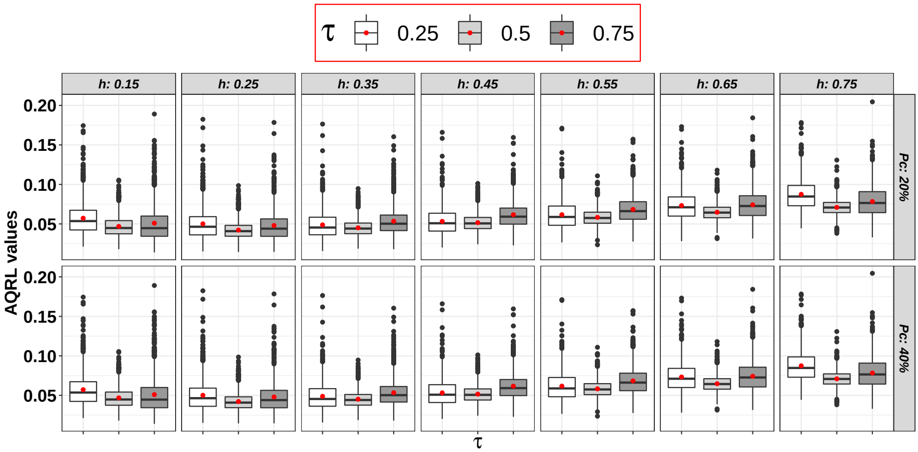
sample size	estimator	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
			bandwidth value h						
$n = 100$	$\widehat{\eta}$	20%	0.041	0.032	0.032	0.039	0.042	0.046	0.050
		40%	0.056	0.038	0.039	0.045	0.050	0.054	0.057
	$\widehat{\phi}$	20%	0.006	0.005	0.004	0.004	0.003	0.003	0.003
		40%	0.007	0.005	0.004	0.004	0.004	0.004	0.004
	$\widehat{\alpha}$	20%	0.036	0.027	0.023	0.022	0.020	0.020	0.021
		40%	0.037	0.029	0.026	0.025	0.024	0.023	0.023
$n = 300$	$\widehat{\eta}$	20%	0.018	0.016	0.019	0.024	0.026	0.029	0.034
		40%	0.021	0.018	0.022	0.025	0.028	0.031	0.038
	$\widehat{\phi}$	20%	0.003	0.003	0.003	0.003	0.002	0.002	0.003
		40%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
	$\widehat{\alpha}$	20%	0.018	0.015	0.014	0.014	0.014	0.016	0.018
		40%	0.020	0.016	0.016	0.016	0.016	0.017	0.018

Comparing the two simulation models, Model I reveals a better performance of $\widehat{\eta}(x)$ and $\widehat{\phi}(x)$ compared to that of Model II whilst showing approximately the same performance for $\widehat{\alpha}(x)$. This is owing to the fact that the location function has a more complex shape (a peak and valley) in Model II while only a valley in Model I.

The AQRL values under Scenario II are presented in Figure 3 for $n = 100$, and for Models I and II. Table 3 lists the median AQRL values for Model



(a) Model I



(b) Model II

Fig. 3: **Scenario II.** AQRL values across seven bandwidth values for three quantiles over 500 simulated samples of size $n = 100$ with 20% and 40% censoring proportions; Model I (a) and Model II (b).

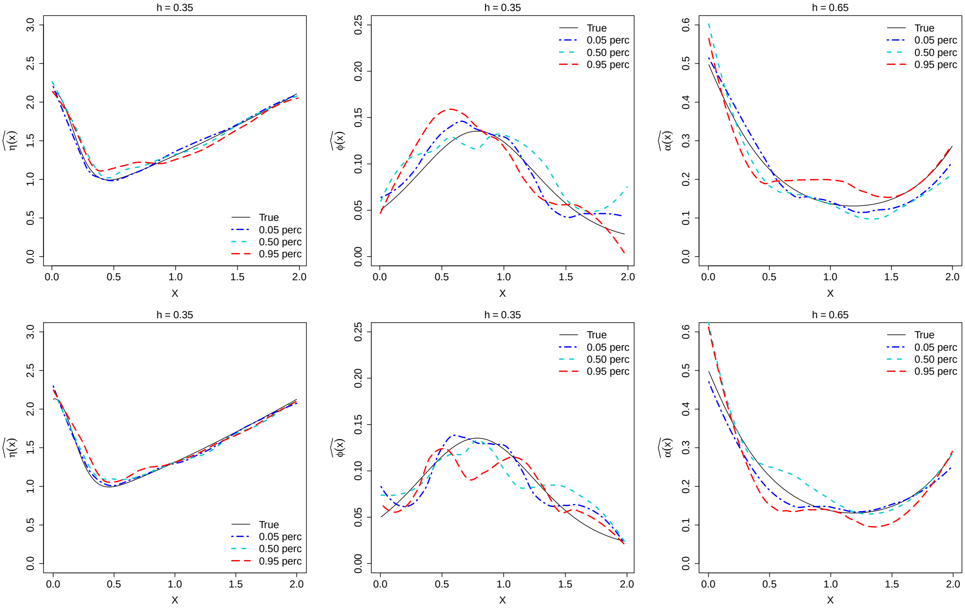
I (a similar table for Model II can be found in the Supplementary material; see Table S1). As can be seen from Figure 3a and Table 3, for Model I, the lower order quantile ($\tau = 0.25$) estimates show a better performance than the higher order quantiles ($\tau = 0.5, 0.75$). This is due to the fact that higher order quantiles of the conditional distribution of the observed survival time are affected by the right censoring. In Figure 3b, showing boxplots for Model II, the lower order quantile estimate shows inferior performance compared to the upper order quantiles. This might be due to the larger variability in the

Table 3: **Scenario II.** Median AQRL values for the quantile estimator $\widehat{Q}_\tau(T \mid x)$, over a grid of x-values. Results for Model I, for three values of τ, two censoring proportions P_c, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

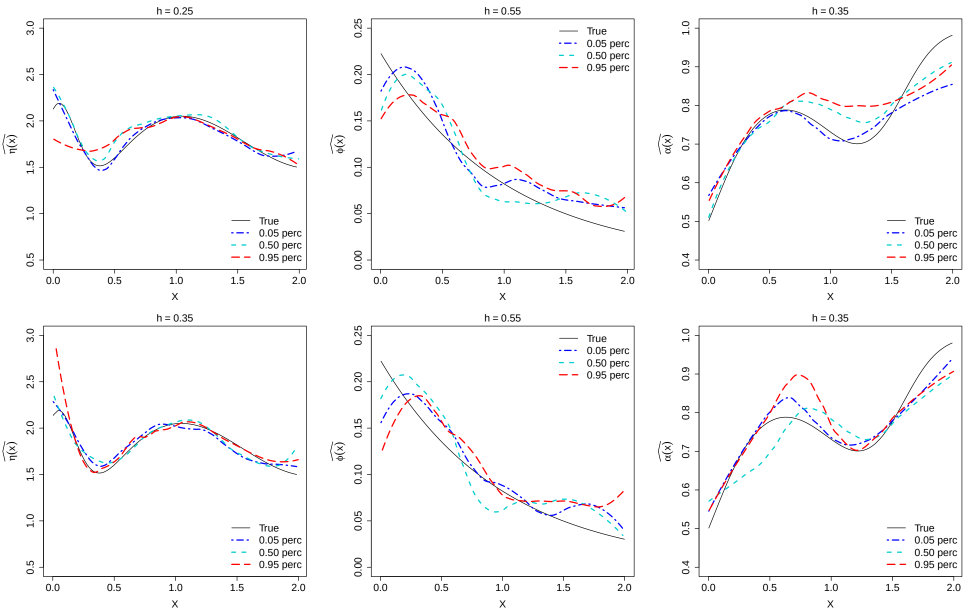
			bandwidth value h						
sample size	τ value	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	0.25	20%	0.030	0.027	0.026	0.028	0.031	0.034	0.039
		40%	0.031	0.028	0.026	0.029	0.032	0.035	0.039
	0.5	20%	0.041	0.037	0.037	0.041	0.046	0.051	0.056
		40%	0.042	0.038	0.038	0.042	0.047	0.052	0.057
	0.75	20%	0.066	0.061	0.066	0.078	0.092	0.104	0.113
		40%	0.068	0.062	0.068	0.079	0.092	0.106	0.115
$n = 300$	0.25	20%	0.021	0.020	0.020	0.025	0.028	0.031	0.034
		40%	0.022	0.020	0.021	0.025	0.029	0.031	0.035
	0.5	20%	0.026	0.026	0.030	0.036	0.042	0.048	0.053
		40%	0.027	0.026	0.030	0.036	0.043	0.049	0.053
	0.75	20%	0.045	0.046	0.056	0.071	0.089	0.104	0.115
		40%	0.045	0.047	0.057	0.072	0.090	0.106	0.116

data in the lower quantiles and the large amount of censored observations in the lower quantiles (see the scatter plot in Figure 5b).

In order to illustrate the quality of the estimators, three representative curves of the estimates together with the true curves are reported. These three representative curves are selected from a list of 500 curves corresponding to the 0.05th percentile (best estimated), 0.50th percentile (median) and 0.95th percentile (worst estimated) of the AISE (AQRL for the quantiles) values. The bandwidth yielding the smallest average AISE (AQRL) value is selected among the list of all candidates for each of the estimates. Figure 4 depicts the true and estimated parameter functions and Figure 5 presents a scatter plot with the estimated quantile curves under Scenario II. The data in the scatter plot are from the simulated sample corresponding to the 0.5th percentile of AQRL values. The quality of the estimators observed from these figures is in line with what we observed from the boxplots. The best, median, and worst estimated curves are closer to the true curve for the 0.25th quantile compared to the upper quantile, for Model I. The corresponding figures under Scenario I are presented in Figures S8a and S8b for Model I and Model II, respectively. The estimates under Scenario II are better than under Scenario I. This is due to the link function used ($g(t) = \ln(e^t - 1)$, resp. $g(t) = \ln(t)$). Indeed, the link function has an impact on the asymptotic variance of the estimator $\widehat{\eta}(x) = g^{-1}(\widehat{\theta}_1(x))$, see Corollary 1.

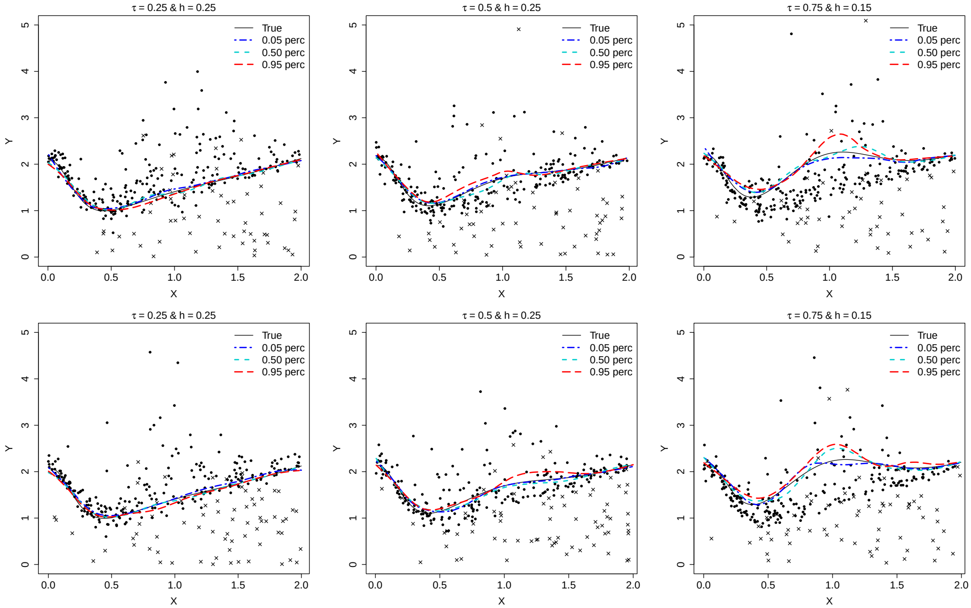


(a) Model I

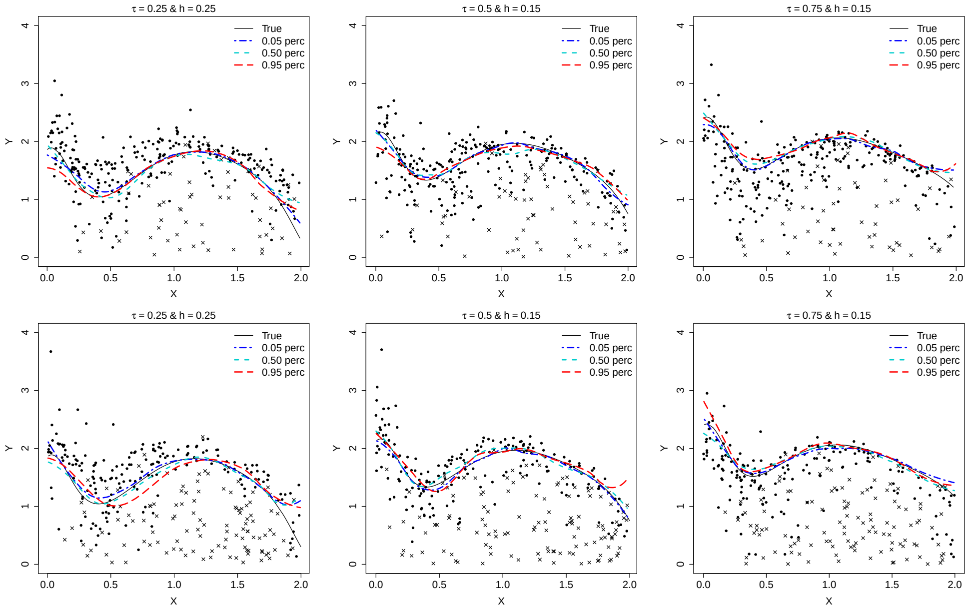


(b) Model II

Fig. 4: **Scenario II.** True and estimated curves of $\widehat{\eta}(x), \widehat{\phi}(x)$ and $\widehat{\phi}(x)$ for $n = 300$ and 20% censoring (first row) and 40% censoring (second row). Model I (a) and Model II (b).



(a) Model I



(b) Model II

Fig. 5: **Scenario II.** True and estimated curves of $\widehat{Q}_\tau(x)$; $n = 300$ with 20% censoring (first row) and 40% censoring (second row). The scatter plot represents the sample data selected according to the 0.50th AQRL percentile (• represent the uncensored and × the censored cases).

5.2 Part II: Multiple covariates

In this part, we compare our proposed estimator (TPA) with competitors in a partially linear model setting. We consider model (14) with TPA Laplace error distribution with a fixed index parameter. For comparison purpose we include results on the following methods available in the literature: non-parametric quantile regression (NP) of Gannoun et al. [18], semi-parametric quantile regression (SPQR) of Bravo [6] and single-index quantile regression (SIQR) studied by Christou and Akritas [9].

We are interested in the performance of the estimate of: *(i)* the parametric component, more specifically $\boldsymbol{\beta}_1$ in (14); *(ii)* the non-parametric component $\phi(x) = \exp(\theta_2(x))$; *(iii)* the conditional τth quantile function of $Z = \ln(T)$ given $\mathbf{X} = \mathbf{x}$ ($Q_\tau(Z \mid \mathbf{x})$).

5.2.1 Data generating mechanism

We consider sample size $n = 300$, 500 Monte Carlo simulations and 20% censoring proportion in three different models.

- **Model I - location and scale have no covariates in common:**

$$Z = \mathbf{X}_1^T\boldsymbol{\beta}_1 \ + \ \exp\big(\sin(2\pi X_2) \ - \ 1.5\big)\varepsilon, \tag{17}$$

  where $\mathbf{X}_1 = (X_{11}, X_{12})^T$ are independently generated from a standard uniform distribution Unif$(0,1)$ and $X_2 \sim$ Unif$(0,2)$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T = (1,2)^T$ and $\theta_2(x_2) \equiv \ln(\phi(x_2)) = \sin(2\pi x_2) - 1.5$. The random error term ε has a TPA Laplace distribution with parameters $\mu = 0, \phi = 1, \alpha = 0.25$ and hence $Q_\tau(\varepsilon) = \widetilde{F}_\alpha^{-1}(\tau;\ \mu = 0, \phi = 1, \alpha = 0.25)$. The τth quantile of the TPA Laplace distribution with $\mu = 0,\ \phi = 1$ and index parameter α is given by

$$\widetilde{F}_\alpha^{-1}(\tau) = \begin{cases} \dfrac{1}{1-\alpha}\ln\left(\dfrac{\tau}{\alpha}\right) & \text{if } \tau < \alpha \\ -\dfrac{1}{\alpha}\ln\left(\dfrac{1-\tau}{1-\alpha}\right) & \text{if } \tau > \alpha. \end{cases}$$

  The conditional τth quantile function for Z given $\mathbf{X} = (X_{11}, X_{12}, X_2)^T$ is thus given by $Q_\tau(Z \mid \mathbf{X}) = \mathbf{X}_1^T\boldsymbol{\beta}_1 \ + \ \exp\big(\sin(2\pi X_2) \ - \ 1.5\big)\widetilde{F}_\alpha^{-1}(\tau)$.

  We can rewrite the model in (17) as

$$Z = Q_\tau(Z \mid \mathbf{X}) \ + \ \exp\big\{\sin(2\pi X_2) - 1.5\big\}\varepsilon_\tau^*,$$

  where $\varepsilon_\tau^* = \big(\varepsilon - \widetilde{F}_\alpha^{-1}(\tau)\big)$ with its τth quantile equals zero. This guarantees that the assumption of SPQR of [6] (that the error has zero quantile with order τ) is satisfied. The censoring variable $C \sim N(\theta_c, 1)$, with θ_c such that there is 20% censoring.

- **Model II - same covariates are involved in location and scale:**

$$Z = \exp\big(\mathbf{X}^T\boldsymbol{\beta}_1\big) \ + \ \big(\sin(2\pi\mathbf{X}^T\boldsymbol{\beta}_1) + 2\big)\varepsilon, \tag{18}$$

where $\mathbf{X} = \mathbf{X}_1$ from Model I. The parameter vector $\boldsymbol{\beta}_1$, the error ε and the censoring distribution are the same as in Model I. This model is directly taken from Christou and Akritas [9] except that the random error term follows a TPA distribution here.

- **Model III - Model II with conditional censoring distribution:** the *conditional* censoring distribution is $C \mid X \sim \text{Expo}(\theta_c X)$ and $X = X_{12} \sim \text{Unif}(0, 1)$.

5.2.2 Performance measures

We investigate the performance of the estimators on five different quantile levels: $\tau = 0.1$, 0.25, 0.50, 0.75 and 0.9. The finite sample performance of the estimator of the regression coefficient vector $\widehat{\boldsymbol{\beta}}_1$ is measured using the approximated *bias* and standard error (*SE*). The performance of conditional quantile estimator is assessed by the weighted AQRL (WAQRL) given by

$$\text{WAQRL} = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{1 - \widehat{G}(Y_i \mid \cdot)} \rho_\tau \Big(\widehat{Q}_\tau \{ Z_i \mid \mathbf{x}_i \} - Q_\tau (Z_i \mid \mathbf{x}_i) \Big).$$

The motivation to use a weight in this performance measure is that all competitors use $1 - \widehat{G}(Y_i \mid \cdot)$ in their objective function to account for censoring (where the Kaplan-Meier estimator is used). In Section S3 of the Supplementary material, we illustrate in a numerical example that using the weighted or the unweighted version of the criterion does not alter conclusions.

5.2.3 Local kernel weighted estimator specifications

The local kernel weighted estimate of the τth order conditional quantile function in model (17) is $\widehat{Q}_\tau(Z \mid \mathbf{X} = \mathbf{x}) = \mathbf{x}_1^T \widehat{\boldsymbol{\beta}}_1 + \exp\big\{\widehat{\beta}_{20}(x_2)\big\} \widetilde{F}_\alpha^{-1}(\tau)$, where $\widehat{\boldsymbol{\beta}}_1 = (\widehat{\beta}_{10}, \ \widehat{\beta}_{11})^T$ and $\widehat{\beta}_{20}(x_2) = \widehat{\theta}_{20}(x_2)$ is the local kernel weighted estimator of the non-parametric component $\theta_2(x_2)$. As mentioned earlier in Section 2, the αth quantile of any TPA family of distributions equals to the location parameter itself so that $\widetilde{F}_\alpha(\tau = \alpha) = 0$. Therefore, when $\alpha = \tau$ the conditional quantile estimates equals the parametric component $\mathbf{x}_1^T \widehat{\boldsymbol{\beta}}_1$.

As the non-parametric component in model (18) involves two covariates, we use a multivariate local linear kernel weighted estimator with a product Epanechnikov kernel function, that is $K(x_1, x_2) = K_{h_1}(x_1) \cdot K_{h_2}(x_2)$ with equal bandwidth parameter values $h_1 = h_2 = h$. Our local kernel weighted estimator is based on the algorithm given in Section 3.3. At the initial stage of the algorithm, we fit a generalized linear model for the location and scale parameters of a power exponential distribution assuming $Z \sim \text{PE}\big(\mu, \sigma, \nu\big)$ [36]. For instance, for Model I we model: the location $\mu(\boldsymbol{X}_1) = \beta_{11} X_{11} + \beta_{12} X_{12}$, the scale $\log(\sigma(X_2)) = \beta_{20} + \beta_{21} X_2$ and the shape ν is fixed at one.

The bandwidth parameter h is selected adaptively using 5-fold cross-validation by minimizing the WAQRL values. First, we split up the data into

5 non overlapping groups with approximately equal size. We estimate the conditional quantile $Q_\tau(Z \mid \mathbf{x})$ in each group $j = 1, 2, ..., 5$ using all observations but excluding those in the jth group. Then the performance of the conditional quantile estimator is measured

$$\mathrm{WAQRL}_j(h) = \frac{1}{n} \sum_{i \in S^{(-j)}}^{n} \frac{\delta_i}{1 - \widehat{G}^{(-j)}(Y_i \mid \cdot)} \rho_\tau \Big(\widehat{Q}_\tau^{(-j)} \{Z_i \mid \mathbf{x}_i\} - Q_\tau(Z_i \mid \mathbf{x}_i) \Big),$$

where $S^{(-j)}$ is the set of observations without the jth group and the superscript $(^{-j})$ illustrates the estimation is done with all observations except for the jth part and h is the candidate bandwidth parameter. We repeat this procedure for all $j = 1, 2, ..., 5$ and ten equally spaced grid values for h from 0.05 to 0.75 for Model I and from 0.15 to 0.75 for Models II and III. Finally, the bandwidth value yielding the smallest AQRL value is selected among the ten candidates.

5.2.4 Competitors specifications

Christou and Akritas [9] provided us with the R code for calculation of their estimator (including bandwidth selection) in the SIQR model. We implemented ourself the estimator of Bravo [6] in the SPQR model, based on their majorize-minimize algorithm. Estimation in the NP model is done via the rq function in R package quantreg (version 5.75), by forming a new weight defined by $\frac{\delta_i}{1 - \widehat{G}(Y_i \mid \cdot)} \tilde{K}_h(\mathbf{x}_i - \boldsymbol{x}_0)$, where $\widehat{G}(Y_i \mid \cdot)$ is the estimate of $G(Y_i \mid \cdot)$. Herein $\tilde{K}_h$ is a multivariate kernel (a product kernel here) with the same bandwidth h in each dimension. We use a Kaplan-Meier estimator for the (unconditional) censoring distribution in Models I and II. In the situation of a conditional censoring distribution (Model III), we estimate $G(\cdot \mid x)$ using the local Kaplan-Meier estimator as defined by expression (2.7) in [9].

5.2.5 Results

Due to an identifiability issue in the SIQR model, it is commonly assumed to impose that $\|\boldsymbol{\beta}_1\| = 1$ or that β_{11}=1 (see e.g. [9]). As a result, we have fixed $\beta_{11} = 1$ and only estimate β_{12} for the parametric component. Furthermore, there are no common covariates in parametric and non-parametric components in Model I, therefore both components for the NP estimator are estimated separately via non-parametric censored regression.

Table 4 presents the approximated bias and SE for $\widehat{\beta}_{12}$, and the average WAQRL over the 500 simulation runs, for five different quantiles. From this table, we observe that the bias as well as the SE values for local kernel weighted estimator are approximately equal across the different quantile levels. This is due to the fact that $\boldsymbol{\beta}_1$ is not specific to τ under the TPA method. TPA and SPQR provide quite similar SE for $\widehat{\beta}_{12}$, while the latter produces some bias in all simulation models. On the other hand, the variability of $\widehat{\beta}_{12}$ increases with the order of quantile for the SIQR method. In general, we observe that

Table 4: Models I, II and III: Approximated bias and standard errors (SE) for $\widehat{\beta}_{12}$, and average WAQRL values (av.WAQRL).

Criteria	Methods	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
Model I						
Bias	*TPA*	0.008	0.017	0.008	0.007	0.007
	SPQR	−0.054	−0.050	−0.044	−0.020	0.007
	SIQR	0.255	0.122	0.235	0.342	0.782
SE	*TPA*	0.087	0.088	0.086	0.087	0.087
	SPQR	0.092	0.092	0.084	0.085	0.089
	SIQR	1.061	0.652	0.838	1.584	4.313
av.WAQRL	*TPA*	0.125	0.018	0.158	0.434	0.800
	SPQR	0.310	0.170	0.163	0.465	1.147
	SIQR	0.108	0.052	0.156	0.403	0.752
	NP	0.655	0.475	0.261	0.266	0.538
Model II						
Bias	*TPA*	0.001	0.010	0.016	0.027	0.031
	SPQR	−0.077	−0.024	0.019	0.076	0.116
	SIQR	0.413	0.105	0.154	0.424	1.883
SE	*TPA*	0.121	0.112	0.116	0.127	0.135
	SPQR	0.137	0.115	0.114	0.133	0.152
	SIQR	1.699	0.572	1.062	1.803	5.413
av.WAQRL	*TPA*	0.436	0.172	0.472	1.201	2.252
	SPQR	1.124	0.162	1.054	4.394	10.620
	SIQR	0.617	0.356	0.589	1.351	2.485
	NP	0.615	0.427	0.699	1.563	2.644
Model III						
Bias	*TPA*	0.009	0.014	0.027	0.040	0.041
	SPQR	−0.124	−0.056	−0.001	0.052	0.091
	SIQR	0.581	0.138	0.140	0.489	0.916
SE	*TPA*	0.128	0.118	0.121	0.135	0.142
	SPQR	0.163	0.127	0.118	0.139	0.159
	SIQR	2.524	0.721	0.850	2.363	3.238
av.WAQRL	*TPA*	0.459	0.459	0.489	1.195	2.234
	SPQR	0.975	0.975	1.125	4.524	10.805
	SIQR	0.751	0.751	0.716	1.389	2.728
	NP	0.702	0.702	0.728	1.611	2.807

the SIQR method induces important bias and SE for the estimates of the parametric component in all simulation settings.

When comparing the methods based on the performance of the estimator of the conditional quantiles, it is seen from Table 4 that TPA and SIQR have a comparable performance under Model I, while the former outperforms the latter in the other two simulation models. In contrast to the performance of the parametric component estimates, the SPQR method is revealing the worst performance (large av.WAQRL value) in all simulation settings except when $\tau = 0.25$. An explanation for this exception is that $\widetilde{F}_{\alpha=0.25}(\tau = 0.25) = 0$, leads to $\widehat{Q}_\tau(T \mid \mathbf{x}) = \mathbf{x}^T \widehat{\boldsymbol{\beta}}_1$, and hence the SPQR performs similarly at this particular quantile.

Overall, the proposed TPA method outperforms the three competitors, especially w.r.t. the estimation of the conditional quantile. Furthermore, SPQR outperforms SIQR in estimating the parametric component, however, the reverse holds when estimating the quantile function. Therefore, SPQR and SIQR are less accurate to estimate the non-parametric and parametric components, respectively. Due to the presence of conditional censoring the AQRL value is slightly higher for all methods under Model III.

6 Real data analysis

We apply our methodology on two data examples: the Small Cell Lung Cancer (SCLC) data used by Ying et al. [41] and the Mayo Clinic Primary Biliary Cirrhosis (PBC) data that has been analysed frequently (e.g. in Fleming and Harrington [17], Jin et al. [23], Rubio and Hong [34]). In order to construct confidence intervals for the parameters of interest, one can use the asymptotic normality result of our estimator. The asymptotic variance, however depends on the unknown distribution of the covariate and the censoring variable which makes it less useful in practice as it therefore depends on the performance of estimators of these distributions. Therefore, we construct confidence intervals and calculate standard errors based on bootstrap resampling. We generate $B = 1000$ bootstrap samples in the single covariate analysis and $B = 500$ samples in the multiple covariate setting. Three different members of the TPA family are used in the modelling: Laplace, logistic and normal. We compare the methods based on their predictive ability of the quantile function via AQRL (where in (16) $Q_\tau(T \mid x_j)$ is replaced by Y_j and the grid points are the observations).

6.1 Example 1: SCLC data

In this clinical study, there were 121 patients with small cell lung cancer randomly assigned to two treatment groups (arm A and arm B), with 62 patients assigned to arm A and 59 patients to arm B. We refer the readers to Ying et al. [41] for a more detailed description of the clinical design that was conducted to evaluate the two treatment regimes. The data is available in the R package emplik. The survival times are measured in days administering the two treatments together with patient's entry age. At the time of analysis, each death time was either observed or administratively censored. Thus, it is reasonable to assume that the censoring variable does not depend on the covariates. We denote the base 10 logarithm of the patient's death time by Z, of whom 19% were censored observations. We denote the treatment indicator and the patient's entry age by X_1 ($X_1 = 1$ if a patient is in group B and 0 otherwise) and X_2, respectively.

As a preliminary analysis we fit a non-parametric median censored regression for the survival time on patients age (univariate analysis). The histogram of the residuals obtained from this preliminary analysis is displayed in Figure 6, together with the scatter plot of the data. From this figure, it can be noted that the distribution of the residuals is not far from symmetry. Therefore, we expect the index parameter to be close to 0.5 and we estimate the treatment effect on the survival time, adjusting for the patient's age in the scale parameter of the distribution using the following semi-parametric quantile regression model

$$\begin{aligned} Z &= \beta_{10} \ + \ \beta_{11} X_1 \ + \ \exp\left\{\theta_2(X_2)\right\} \cdot \varepsilon, \\ Q_\tau(Z \mid \mathbf{X}) &= \beta_{10} \ + \ \beta_{11} X_1 \ + \ \exp\left\{\theta_2(X_2)\right\} \cdot \widetilde{F}_\alpha^{-1}(\tau), \end{aligned} \tag{19}$$

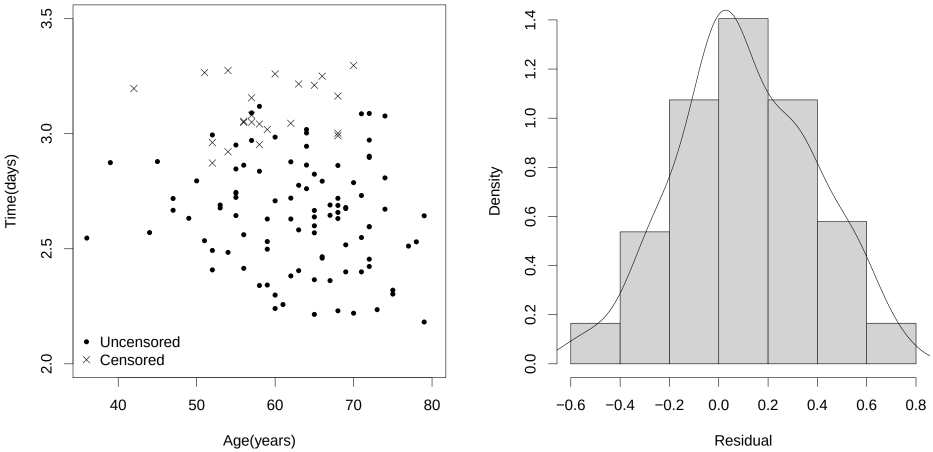


Fig. 6: SCLC data: scatter plot (left) the histogram of residuals (right) obtained by fitting a nonparametric median censored regression.

where $\widetilde{F}_{\alpha}^{-1}(\tau)$ denotes the quantile function for the TPA distributions with parameters $\mu = \ln(\eta) = 0,\ \phi = 1$ and index α. The unknown function $\theta_2(\cdot)$ is estimated local linearly.

Based on the local kernel weighted maximum likelihood estimator, we find that $\widehat{\alpha} = 0.45,\ 0.48,\ 0.51$ for the Laplace, logistic and normal TPA distributions, respectively. The Epanechnikov kernel is used as a weighting function for the non-parametric estimation. The bandwidth parameter h is selected via 5-fold cross-validation by minimizing the AQRL for the conditional τth quantile function of Z.

Table 6 provides the estimated regression coefficients with their bootstrap standard errors and 95% bootstrap confidence intervals for the parametric components. Estimates based on the TPA normal model have smaller bootstrap standard errors and narrower bootstrap confidence intervals compared to the other two distributions. This is in line with the AQRL values presented in Table 5 (first 3 columns), in particular for the first two quantiles. The result for the 0.5th quantile is consistent with the median regression model studied by Ying et al. [41]. For example, in Ying et al. [41] the predicted median survival time for a patient with age 62-year-old in treatment Arm A was 603 days. This predicted value coincides with our estimate of the 0.5th conditional quantile function. For the TPA Laplace distribution $\widehat{Q}_{\tau=0.5}(Z \mid X_1 = 0, X_2 = 62) = \widehat{\beta}_{10} + \exp(\widehat{\beta}_{20}) \cdot \widetilde{F}_{\widehat{\alpha}}^{-1}(\tau = 0.5) = 2.78$, changing to the original time scale it becomes $10^{2.78} \approxeq 603$ days. For treatment Arm B, the predicted median survival time is $\widehat{Q}_{\tau=0.5}(Z \mid X_1 = 1, X_2 = 62) = \widehat{\beta}_{10} + \widehat{\beta}_{11} + \exp(\widehat{\beta}_{20}) \cdot \widetilde{F}_{\widehat{\alpha}}^{-1}(\tau = 0.5) = 2.6 \approx 398$ days, indicating that patients treated in Arm A are more beneficial than those in Arm B.

Table 5: SCLC: approximate quantile residual loss (AQRL) for three different quantiles estimate

	Unconditional α			Conditional $\alpha(\cdot)$		
Models	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$
TPA Laplace	0.080	0.092	0.088	0.090	0.114	0.092
TPA logistic	0.074	0.093	0.093	0.094	0.118	0.092
TPA normal	0.075	0.093	0.090	0.100	0.123	0.093

Table 6: SCLC: estimates, bootstrap standard error (in parenthesis) and confidence intervals for the parametric regression coefficients.

	$\widehat{\beta}_{10}$		$\widehat{\beta}_{11}$	
Model	Estimate (se)	95%CI	Estimate (se)	95%CI
TPA Laplace	2.742 (0.045)	[2.690, 2.874]	−0.171 (0.059)	[−0.317, − 0.085]
TPA logistic	2.710 (0.030)	[2.667, 2.784]	−0.151 (0.044)	[−0.228, − 0.052]
TPA normal	2.723 (0.029)	[2.664, 2.773]	−0.136 (0.043)	[−0.213, − 0.050]

We now consider model (19), where also the index parameter is a function of X_2, and therefore estimating both the scale and index functions non-parametrically. The estimated index function $\widehat{\alpha}(\cdot)$ is depicted in Figure S11 (in the Supplementary material) for the two treatment arms separately. From that figure we note that the estimated index function is below 0.5 for Arm A, whereas for Arm B, the index function increases from about 0.4 to 0.8 with age. This implies that there is a difference in the skewness of the distribution for different age categories and treatment arms as well. For Arm B in the lower age groups, the distribution of patient's survival time is right skewed while in the larger age categories it is left skewed. Further, the differences between the estimates for the three TPA models, are quite different in the case of Arm A (especially in the lower age categories), whereas for Arm B the three estimated curves are almost parallel. Table 5 summarizes the AQRL (unconditional α and conditional $\alpha(\cdot)$) in the three fitted models. The AQRL values obtained with a conditional $\alpha(\cdot)$ are larger than that of the unconditional α in all fitted models. Besides that the modelling with a conditional index parameter reveals different skewing behaviour for the two treatment arms, it is not showing an improved performance in terms of estimating the conditional quantiles.

In Figure 7 the estimated quantile curves for the TPA Laplace distribution versus patient's age are presented. From that figure we conclude that the survival time in treatment Arm A is better than in Arm B across the entire age range and the three different percentiles of the survival time.

These data have also been analysed with an AFT model with least absolute deviation estimation [22] and a heteroscedastic AFT model Zhou et al. [43]. Despite the fact that the effect of entry age was not significant in previous studies, a closer inspection of this plot also shows a non-linear relationship with patient's age in the 0.25th and 0.75th quantiles but constant effects on the 0.5th quantile (median).

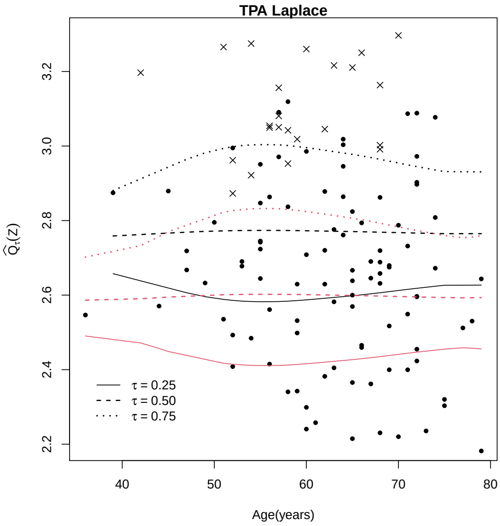


Fig. 7: SCLC: estimated quantile curves fitted by TPA Laplace model. The solid line for the 0.25th quantile, dashed for the 0.5th quantile, and dotted line for 0.75th quantile; grey line for the treatment Arm A and red line for treatment Arm B (• represent the uncensored and × the censored cases).

6.2 Example 2: PBC data

The PBC data can be found in Appendix D of Fleming and Harrington [17]. The PBC data set includes information about the survival time (in days) together with the status of the patient at the end of the study (0/1/2 for censored, transplant, dead) and prognostic factors for 418 patients. The first 312 cases in the data set participated in the randomized trial and the remaining 106 cases did not participate in the clinical trial, but their basic measurements have been recorded and followed for survival. Transplanted cases are considered as censored, and hence there were about 61.5% of censored observations.

6.2.1 Single covariate analysis

We first consider a single covariate analysis for all unknown functions $(\eta(\cdot), \phi(\cdot), \alpha(\cdot))$ and estimate them local linearly. Fleming and Harrington [17] showed that serum bilirubin (in mg/dl) is the strongest predictor of survival probability of patients. Therefore, we perform a univariate analysis for the 312 cases that participated in the randomized trial. Apart from the conditional quantile function, we also estimate the conditional survival function given two values of serum bilirubin ($X = 0.5$, and $X = 2.35$). The point and interval estimates (Table 7) for the conditional survival function at times $t = 1$ and $t = 5$ are calculated. Li and Datta [29] studied the influence of serum bilirubin on the survival probabilities using a non-parametric regression approach for these specific values of bilirubin and at $t = 1$ and $t = 5$.

Table 7: PBC (univariate): estimated conditional survival function together with a bootstrap 90% confidence interval for two time points (in years) and at two serum bilirubin values (in mg/dl).

Model	$t = 1$				$t = 5$			
	$X = 0.5$		$X = 2.35$		$X = 0.5$		$X = 2.35$	
TPA Laplace	0.997	[0.992, 0.999]	0.941	[0.942, 0.976]	0.946	[0.915, 0.968]	0.668	[0.604, 0.746]
TPA logistic	0.998	[0.989, 0.999]	0.968	[0.947, 0.982]	0.944	[0.907, 0.969]	0.663	[0.602, 0.715]
TPA normal	0.998	[0.996, 0.999]	0.960	[0.938, 0.981]	0.941	[0.904, 0.967]	0.637	[0.592, 0.694]

In the TPA Laplace model the 90% bootstrap confidence interval for the 5-year survival probability of a patient with serum bilirubin of 0.5 is at least 0.915 and can it be as high as 0.968. In other words, this interval indicates that the proportion of patients in the population whose lifetimes would exceed 5-year given a serum bilirubin 0.5 can be found in between 0.915 and 0.968. This result also coincides with that of Li and Datta [29].

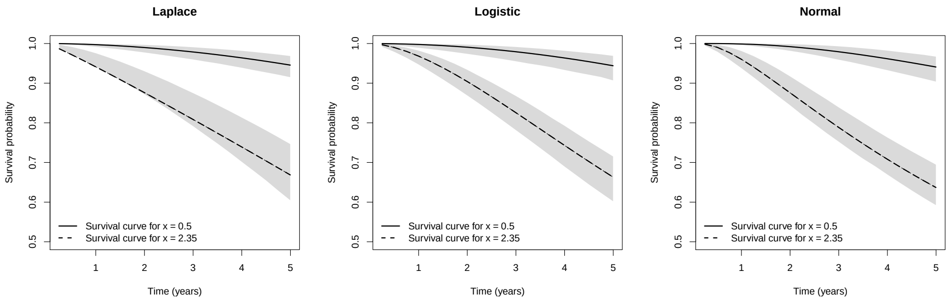


Fig. 8: PBC (univariate): 90% pointwise bootstrap confidence intervals for the conditional survival function at two values of serum bilirubin ($X = 0.5$ and $X = 2.35$) fitted by three TPA distributions: Laplace (left), logistic (middle) and normal (right).

Figure 8 visualizes the 90% pointwise bootstrap confidence intervals for the conditional survival function in the first 5 years for the two bilirubin levels. The survival probability of a patient with serum bilirubin value of 0.5 is noticeably larger than that with serum bilirubin value 2.35 across all time points. The difference between the two curves increases with the survival time.

Table 8: PBC (univariate): average AQRL values over 1000 bootstrap samples for eight different quantiles.

	τ							
Models	0.025	0.05	0.075	0.10	0.25	0.50	0.75	0.90
TPA Laplace	0.269	0.532	0.788	1.028	1.840	1.451	0.723	0.289
TPA logistic	0.260	0.498	0.732	0.964	1.864	1.449	0.723	0.289
TPA normal	0.272	0.508	0.747	0.984	1.858	1.446	0.723	0.289

In order to assess the performance of the fitted models we compute the AQRL values across eight different quantiles, summarized in Table 8. As most of the uncensored observations are found in the lower quantiles of the distribution we measure the performance with three very small values of $\tau = 0.025, 0.05, 0.075$, and further $\tau = 0.1, 0.25, 0.50, 0.75, 0.90$. All the three models result in a comparable performance for all quantile orders. Figure 9 depicts the 90% pointwise bootstrap confidence intervals for the conditional quantile function for some quantile levels. It can be noted that the estimated curve drops very quickly as the quantile level increases. This indicates that the impact of changes in the serum bilirubin on the upper quantiles is stronger than for the lower quantiles of the conditional distribution. This concurs with the estimated survival curves depicted in Figure 8, where the differences between the two curves increase as time goes by.

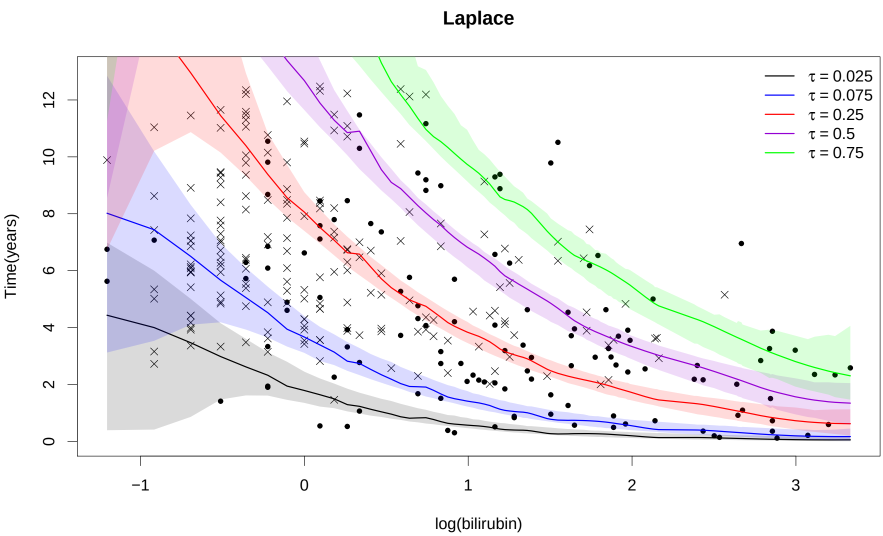


Fig. 9: PBC (univariate): 90% pointwise bootstrap confidence intervals for the conditional quantile function fitted by TPA Laplace model (• represent the uncensored and × the censored cases).

6.2.2 Multiple covariate analysis

The PBC data is also studied by Jin et al. [23], where rank based inference for the semi-parametric AFT model is proposed, and by Rubio and Hong [34], who fitted an AFT model taking two-piece errors with two constant scale parameters. Both authors analysed the data with five covariates: age (in years), logarithm of the serum albumin (log(albumin) in mg/dl), logarithm of the serum bilirubin (log(bilirubin) in mg/dl), edema, and logarithm of the prothrombin time (log(protime) in seconds).

We consider the same set of covariates in this multivariable part, specifically with the aim of illustrating the proposed method in Section 3.3. We consider all covariates in the parametric part while only the continuous covariates (that is excluding edema) in the non-parametric components. All the five covariates are thus involved in modelling the location of the distribution while only the four continuous covariates are used to estimate the scale and index parameters in a multivariate local linear kernel weighted smoothing approach. As all covariates have different units of measurement, we use the standardized version of the covariates in this part of the analysis.

Let $\mathbf{X}_1$ denote the 6-dimensional vector containing one and the set of all the five covariates included in the parametric component and $\mathbf{X}_2$ the 4-dimensional vector denoting those covariates included in the two non-parametric components. We consider the following semi-parametric quantile regression model for T using (4) with $g(t) = \ln(t)$ and $g^{-1} = \exp(t)$:

$$Q_\tau(T \mid \mathbf{X}) = \exp\left(\mathbf{X}_1^T \boldsymbol{\beta}_1 \ + \ \phi(\mathbf{X}_2)\cdot C_{\alpha(\mathbf{X}_2)}(\tau)\right), \tag{20}$$

where $\boldsymbol{\beta}_1 = (\beta_{10}, \beta_{11}, \ldots, \beta_{15})^T$. The scale $\phi(\cdot)$ and index $\alpha(\cdot)$ are unknown smooth functions and need to be estimated using a multivariate local linear kernel weighted approach. We use a product kernel function with an equal bandwidth parameter for all components. We compare our multivariate local kernel weighted maximum likelihood estimator with an MLE in an AFT model with log-normal error, based on their AQRL for eight different quantiles, $\tau=$ 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 0.9.

Figure 10 depicts the estimates for the non-parametric components against each standardized covariate together with a LOESS smoothed curve for the TPA normal model. This provides us with information on the relation between each covariate and the scale and index parameters. From this figure, we conclude that log(protime) has a the strongest effect on both the scale and index parameters, in contrast to the other covariates. Therefore, we decide to change model (20), to a model where only log(protime) is used to model the unknown scale and index parameters. As such the non-parametric components are modelled with a single covariate. No modification are made on the parametric component:

$$Q_\tau(T \mid \mathbf{X}) = \exp\left(\mathbf{X}_1^T \boldsymbol{\beta}_1 \ + \ \phi(\log(\text{protime}))\cdot C_{\alpha(\log(\text{protime}))}(\tau)\right). \tag{21}$$

Boxplots of the AQRL values for our multivariate local linear kernel weighed MLE and the MLE in the AFT model, based on 500 bootstrap samples, for models (20) and (21) are summarized in Figures 11 and 12, respectively. The estimators based on Model (20) result in larger AQRL values compared to Model (21). TPA Laplace and logistic distributions result in a better performance in the upper three quantiles. From Figure 12 it is clear that

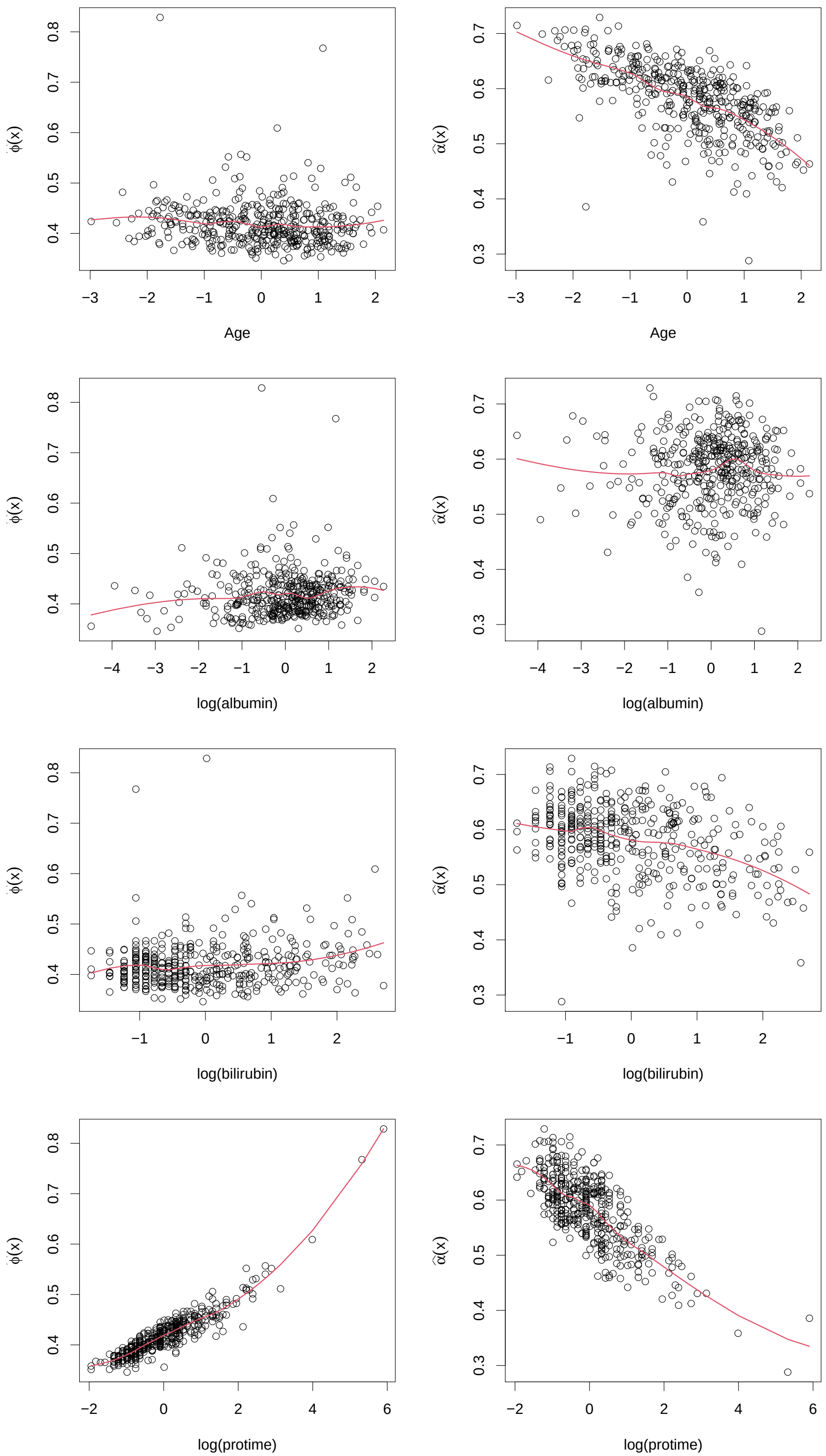


Fig. 10: PBC data (multivariate): scatter plots of local estimates from TPA normal model against each continuous covariate together with simple loess smoothed curve (red line).

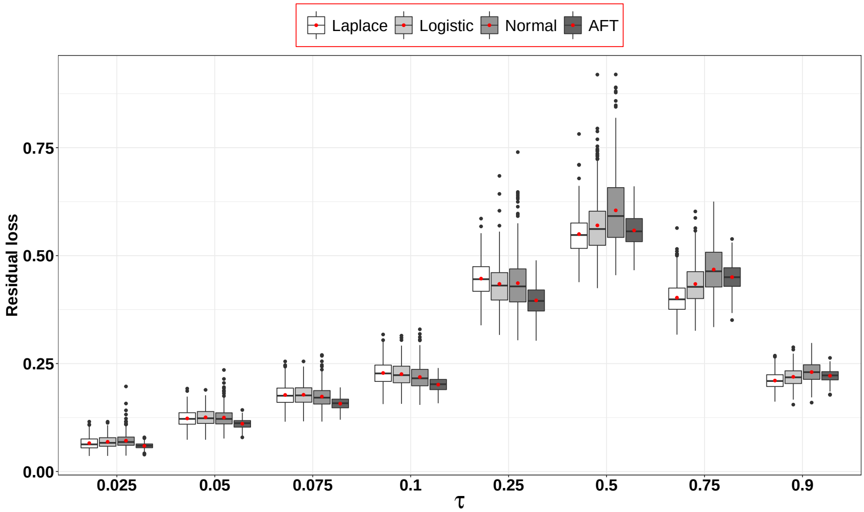


Fig. 11: PBC (multivariable): AQRL from 500 bootstrap samples for model (20).

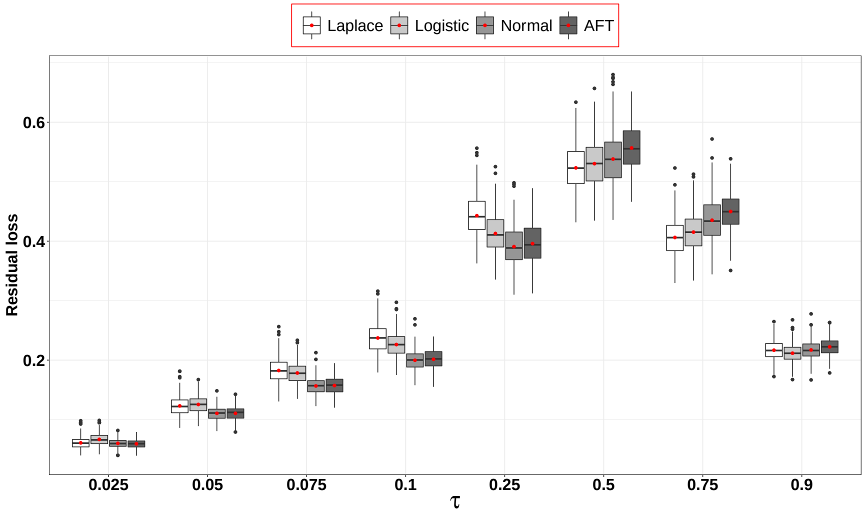


Fig. 12: PBC (multivariable): AQRL value from 500 bootstrap samples for model (21).

TPA normal and AFT distribution lead to a similar performance in the lower three quantiles. However, the former shows better predictive performance than the latter in the upper five quantiles. From the two partially linear models, the one with only log(protime) in the scale and index function (21) provides the best results according to the AQRL.

The estimated regression coefficients with their bootstrapped standard errors in model (21) are summarized in Table 9. The estimated coefficients for the five covariates in the TPA normal and the AFT models are fairly similar. These are furthermore similar to those reported in Ding [12].

Table 9: PBC (multivariable): estimates and associated bootstrap standard errors (in parenthesis) for model (21).

Estimates	TPA Laplace		TPA logistic		TPA normal		AFT	
intercept	7.560	(2.007)	7.332	(1.366)	7.622	(1.388)	7.836	(1.551)
age	−0.030	(0.006)	−0.027	(0.005)	−0.026	(0.005)	−0.026	(0.005)
log(albumin)	1.348	(0.573)	1.494	(0.452)	1.501	(0.500)	1.533	(0.499)
log(bilirubin)	−0.568	(0.071)	−0.569	(0.054)	−0.581	(0.056)	−0.593	(0.055)
edema	−0.669	(0.353)	−0.795	(0.235)	−0.884	(0.241)	−0.898	(0.225)
log(protime)	−2.177	(0.830)	−2.209	(0.518)	−2.342	(0.498)	−2.428	(0.569)

7 Further discussion and conclusion

We study conditional quantile curve estimates in the context of a large class of two-piece asymmetric distributions based on right censored data. The main features of the large family of two-piece asymmetric distributions are: the αth quantile of the survival distribution coincides with its location parameter; and the family includes any symmetric density with mean zero and variance one. A link function ensures the flexibility to use any unimodal density as a basis, including densities with support the whole real line. In the regression setting the three parameters (location, scale and index) of a member of the family may depend on covariates in a non-parametric way.

The unknown but smooth parameter functions are estimated by maximizing a weighted local polynomial likelihood function, and estimation of the τth quantile curve is straightforwardly obtained by substituting the estimated parameter functions. In case some of the functions (location, scale and index) are modelled parametrically and the others non-parametrically, a two stages estimation procedure is proposed, via a profile likelihood approach.

We establish the large sample properties of the estimators for an entire class of conditional TPA distributions. The relevance and practical use of the proposed method is demonstrated on real data applications. Although we focus on right random censorship in this paper, the methodology can be extended to other types of censoring, such as left and interval censoring.

Among the interesting research issues to tackle in further research are the development of methods for data-driven bandwidth selection. Hereby one can think of a global bandwidth h but also of a variable bandwidth, since at boundary regions it might be better to use a somewhat larger bandwidth.

In this paper we worked under the standard assumptions made in survival analysis, among which that of non-informative censoring, and the conditional independence between survival and censoring time. It would be particularly interesting to consider settings in which the parameter functions might be influenced by the censoring mechanism.

Supplementary information. We provide the following additional material in the Supplementary material:

- Section S1 includes the conditional TPA Laplace distribution with its parameter estimation.
- Essential notations and conditions for the asymptotic properties are provided in Section S2.
- Section S3 summarizes the simulation result for Scenario I, presents the other figures and tables for Scenario II, and provides a small study concerning the weighted and unweighted AQRL.
- Some additional results for the SCLC data example are reported in Section S4.

Acknowledgments. The authors are grateful to an Associate Editor and two reviewers for their comments which led to an improvement of the manuscript. We thank the authors of Christou and Akritas [9] to provide us with the R code to calculate their estimator in the SIQR model. The second author gratefully acknowledges support from Research Grant FWO G0D6619N of the Flemish Science Foundation, and from the C16/20/002 project of the Research Fund KU Leuven. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.

Declarations

The authors do not have a conflict of interest.

References

[1] Aerts, M., & Claeskens, G. (1997). Local polynomial estimation in multi-parameter likelihood models. *Journal of the American Statistical Association*, **92**(440), 1536–1545.

[2] Anderson, K. M. (1991). A nonproportional hazards Weibull accelerated failure time regression model. *Biometrics*, **47**(1), 281–288.

[3] Antoniadis, A., Gijbels, I., Lambert-Lacroix, S., & Poggi, J.-M. (2016). Joint estimation and variable selection for mean and dispersion in proper dispersion models. *Electronic Journal of Statistics*, **10**(1), 1630–1676.

[4] Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine*, **2**(2), 273–277.

[5] Bottai, M., & Zhang, J. (2010). Laplace regression with censored data. *Biometrical Journal*, **52**(4), 487–503.

[6] Bravo, F. (2018). Semiparametric quantile regression with random censoring. *Annals of the Institute of Statistical Mathematics*, **72**(1), 265–295.

[7] Burke, K., & MacKenzie, G. (2017). Multi-parameter regression survival modeling: An alternative to proportional hazards. *Biometrics*, **73**(2), 678–686.

[8] Chen, S., & Khan, S. (2001). Semiparametric estimation of a partially linear censored regression model. *Econometric Theory*,**17**(3), 567–590.

[9] Christou, E., & Akritas, M. G. (2019). Single index quantile regression for censored data. *Statistical Methods & Applications*, **28**(4), 655–678.

[10] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**(2), 187–202.

[11] De Backer, M., El Ghouch, A., & Van Keilegom, I. (2020). Linear censored quantile regression: a novel minimum-distance approach. *Scandinavian Journal of Statistics*, **47**(4), 1275–1306.

[12] Ding, Y. (2010). *Some New Insights About the Accelerated Failure Time Model.* Doctoral dissertation, University of Michigan.

[13] Ewnetu, WB., Gijbels, I., & Verhasselt, A., (2023). Flexible two-piece distributions for right censored survival data. *Lifetime Data Analysis*, **29**, 34–65.

[14] Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, CRC Press.

[15] Fan, J., & Gijbels, I. (1994). Censored regression: local linear approximations and their applications. *Journal of the American Statistical Association*, **89**(426), 560–570.

[16] Fechner, G. (1897). *Kollektivmasslehre.* Engelmann, Leipzig.

[17] Fleming, T.R. & Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

[18] Gannoun, A., Saracco, J., Yuan, A., & Bonney, G. E. (2005). Nonparametric quantile regression with censored data. *Scandinavian Journal of Statistics,* **32**(4), 527–550.

[19] Gijbels, I., Prosdocimi, I. & Claeskens, G. (2010). Nonparametric estimation of mean and dispersion functions in extended generalized linear models. *Test*, **19**(3), 580–608.

[20] Gijbels, I., Karim, R., & Verhasselt, A. (2019). On quantile-based asymmetric family of distributions: Properties and inference. *International Statistical Review*, **87**(3), 471–504.

[21] Gijbels, I., Karim, R. & Verhasselt, A. (2021). Semiparametric quantile regression using quantile-based asymmetric family of densities. *Computational Statistics and Data Analysis*, 157, 107129.

[22] Huang, J., Ma, S., & Xie, H. (2007). Least absolute deviations estimation for the accelerated failure time model. *Statistica Sinica*, **17**(4), 1533–1548.

[23] Jin, Z., Lin, D. Y., Wei, L. J., & Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, **90**(2), 341–353.

[24] Kim, H. T., & Truong, Y. K. (1998). Nonparametric regression estimates with censored data: Local linear smoothers and their applications. *Biometrics*, **54**(4), 1434–1444.

[25] Kleinbaum, D. G. & Klein, M. (2012). *Survival Analysis. A Self Learning Text.* Third Edition. Springer New York.

[26] Kneib, T., Silbersdorff, A. & Säfken, B. (2023). Rage against the mean – a review of distributional regression approaches. *Econometrics and Statistics*, **26**, 99–123.

[27] Koenker, R. (2015). Quantile Regression. *International Encyclopedia of the Social & Behavioral Sciences*, **19**(2), 712–718.

[28] Lee, S. (2003). Efficient semi-parametric estimation of a partially linear quantile regression model. *Econometric Theory*, **19**(1), 1–31.

[29] Li, G., & Datta, S. (2001). A bootstrap approach to non-parametric regression for right censored data. *Annals of the Institute of Statistical Mathematics*, **53**(4), 708–729.

[30] Peng, L., & Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, **103**(482), 637–649.

[31] Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**(464), 1001–1012.

[32] Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, **25**(3), 303–325.

[33] Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, **32**(1), 143–155.

[34] Rubio, F. J., & Hong, Y. (2016). Survival and lifetime data analysis with a flexible class of distributions. *Journal of Applied Statistics*, **43**(10), 1794–1813.

[35] Rubio, F. J., & Yu, K. (2017). Flexible objective Bayesian linear regression with applications in survival analysis. *Journal of Applied Statistics*, **44**(5), 798–810.

[36] Stasinopoulos, M. D., Rigby, R. A., Heller, G. Z., Voudouris, V., & De Bastiani, F. (2017). *Flexible Regression and Smoothing: using GAMLSS in R.* CRC Press.

[37] Tibshirani, R., & Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association,* **82**(398), 559–567.

[38] Wallis, K. F. (2014). The two-piece normal, binormal, or double Gaussian distribution: Its origin and rediscoveries. *Statistical Science*, **29**(1), 106–112.

[39] Wang, H. J., & Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**(487), 1117–1128.

[40] Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, **11**(14-15), 1871–1879.

[41] Ying, Z., Jung, S. H., & Wei, L. J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, **90**(429), 178–184.

[42] Yu, K., & Jones, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**(441), 228–237.

[43] Zhou, M., Kim, M. O., & Bathke, A. C. (2012). Empirical likelihood analysis for the heteroscedastic accelerated failure time model. *Statistica Sinica*, **22**(1), 295–316.

SUPPLEMENTARY MATERIAL

for

Two-piece distribution based semi-parametric quantile regression for right censored data

Worku B. Ewnetu, Irène Gijbels and Anneleen Verhasselt

In this supplement, we provide the following additional material:

- Section S1 includes the conditional TPA Laplace distribution with its parameter estimation.
- Essential notations and conditions for the asymptotic properties are provided in Section S2.
- Section S3 summarizes the simulation results for **Scenario I**, provides additional information on the simulations for **Scenario II,** and presents a small study concerning the weighted and unweighted AQRL.
- Some additional results for the SCLC data example are reported in Section S4.

S1 Conditional TPA Laplace distribution

Consider a standard Laplace density $f_0(s) = \frac{1}{2}e^{-|s|}$ together with the corresponding survival function $S_0(s) = \frac{1}{2} - \frac{1}{2}\text{sgn}(s)\big(1 - e^{-|s|}\big)$, and the quantile function $F_0^{-1}(s) = -\text{sgn}(\tau - \frac{1}{2}) \ln\big(1 - 2 \mid \tau - \frac{1}{2} \mid\big)$. Then the conditional TPA Laplace density for the survival time T given a single covariate $X = x$ can be obtained from (3) and written as

$$f_{T|X}(t;\, \boldsymbol{\theta}(x)) = \frac{\alpha(x)(1-\alpha(x))g'(t)}{\phi(x)} \exp\bigg(- \frac{1}{\phi(x)}\rho_{\alpha(x)}\big\{g(t) - g(\eta(x))\big\}\bigg)$$

and the corresponding conditional survival function is

$$\begin{aligned} S_{T|X}(t;\, \boldsymbol{\theta}(x)) = &\big(\mathbb{I}\{t \geq \eta(x)\} - \alpha(x)\big) \exp\bigg(- \frac{1}{\phi(x)}\rho_{\alpha(x)}\big\{g(t) - g(\eta(x))\big\}\bigg) \\ &+ \mathbb{I}\{t < \eta(x)\}, \end{aligned}$$

where $\rho_{\alpha(x)}(u)$ is the check (loss) function defined as $\rho_{\alpha(x)}(u) = u\big(\alpha(x) - \mathbb{I}\{u < 0\}\big)$ for $0 < \alpha(x) < 1$ and $\mathbb{I}(\cdot)$ is the indicator function. This loss function assigns weight $1 - \alpha(x)$ and $\alpha(x)$ to $t < \eta(x)$ and $t \geq \eta(x)$, respectively.

S1.1 Single covariate

Considering these two functions together with the expression of (8) and the data (x_i, y_i, δ_i), the contribution of each uncensored datum in the local likelihood function is then

$$\begin{aligned} \text{L}_u(\boldsymbol{\beta};\, y_i, x_i, x_0) = &\ln\big(\alpha(x_i)\{1 - \alpha(x_i)\}\big) + \ln\big\{g'(y_i)\big\} \\ &- \mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2 - \rho_{\alpha(x_i)}\Bigg(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Bigg) \end{aligned}$$

and for the censored case will be given as

$$\begin{aligned} \text{L}_c(\boldsymbol{\beta};\, y_i, x_i, x_0) = &\,\omega_i \ln\Bigg(1 - \alpha(x_i)\exp\Big\{ - \rho_{\alpha(x_i)}\Big(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Big)\Big\}\Bigg) \\ &+ (1 - \omega_i)\Bigg\{\ln\big(1 - \alpha(x_i)\big) - \rho_{\alpha(x_i)}\Bigg(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Bigg)\Bigg\}, \end{aligned}$$

where $\alpha(x_i) = \dfrac{\exp\big(\mathbf{x}_{i,p_3}^T\boldsymbol{\beta}_3\big)}{1 + \exp(\mathbf{x}_{i,p_3}^T\boldsymbol{\beta}_3)}$ and $\omega_i = \mathbb{I}\big\{y_i < g^{-1}(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1)\big\}$. Combining these two equations with the localizing weighted function, the local kernel weighted conditional log-likelihood sample is

$$
\begin{aligned}
&\mathcal{L}_n(\boldsymbol{\beta};\ x_0,h)\\
&=\sum_{i=1}^{n}\delta_i\Bigg(\ \ln\big(\alpha(x_i)\{1-\alpha(x_i)\}\big)\ +\ \ln\big\{g'(y_i)\big\}-\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\\
&\qquad\qquad -\rho_{\alpha(x_i)}\Big(\frac{g(y_i)-g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Big)\Bigg)K_h(x_i-x)\\
&\quad+\sum_{i=1}^{n}(1-\delta_i)\Bigg(\omega_i\ln\Big[1-\alpha(x_i)\exp\Big\{-\rho_{\alpha(x_i)}\Big(\frac{g(y_i)-g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Big)\Big\}\Big]\\
&\quad+(1-\omega_i)\Bigg\{\ln\big(1-\alpha(x_i)\big)-\rho_{\alpha(x_i)}\Big(\frac{g(y_i)-g^{-1}\big(\mathbf{x}_{i,p_1}^T\boldsymbol{\beta}_1\big)}{\exp\big(\mathbf{x}_{i,p_2}^T\boldsymbol{\beta}_2\big)}\Big)\Bigg\}\Bigg)K_h(x_i-x_0).
\end{aligned}
\tag{S1.1}
$$

As a result, the maximum likelihood estimators for the unknown parameter functions at the local point x_0, $\widehat{\boldsymbol{\beta}}_1,\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\boldsymbol{\beta}}_3$ are the maximizer of (S1.1). Accordingly, the local kernel weighted estimator of the conditional quantile function for the survival time T with order τ is then given by

$$
\widehat{Q}_\tau(T\mid X=x_0)=g^{-1}\Big\{g\big(\widehat{\eta}(x_0)\big)\ +\ \widehat{\phi}(x_0)\cdot C_{\widehat{\alpha}(x_0)}(\tau)\Big\},
$$

where $\widehat{\eta}(x_0)=g^{-1}(\widehat{\beta}_{10}),\ \widehat{\phi}(x_0)=\exp(\widehat{\beta}_{20}),\ \widehat{\alpha}(x_0)=\dfrac{\exp\big(\widehat{\beta}_{30}\big)}{1+\exp\big(\widehat{\beta}_{30}\big)}$ and

$$
\begin{aligned}
C_{\widehat{\alpha}(x_0)}(\tau)\ &=\ \frac{1}{1-\widehat{\alpha}(x_0)}\ln\left(\frac{\tau}{\widehat{\alpha}(x_0)}\right)\mathbb{I}(\tau<\widehat{\alpha}(x_0))\\
&\quad-\frac{1}{\widehat{\alpha}(x_0)}\ln\left(\frac{1-\tau}{1-\widehat{\alpha}(x_0)}\right)\mathbb{I}(\tau\geq\widehat{\alpha}(x_0)).
\end{aligned}
$$

Consequently, by employing (6) the local estimator of conditional survival function for the lifetime T at a local point x_0 is given by

$$
\begin{aligned}
&\widehat{S}_{T|X}(t;\ \widehat{\boldsymbol{\theta}}(x_0))\\
&=\Big(\mathbb{I}\{t\geq\widehat{\eta}(x_0)\}-\widehat{\alpha}(x_0)\Big)\exp\Bigg(\Big(\mathbb{I}\{t<\widehat{\eta}(x_0)\}-\widehat{\alpha}(x_0)\Big)\Big\{\frac{g(t)-g\big(\widehat{\eta}(x_0)\big)}{\widehat{\phi}(x_0)}\Big\}\Bigg)\\
&\quad+\mathbb{I}\{t<\widehat{\eta}(x_0)\}.
\end{aligned}
$$

S1.2 Partially linear modelling

We now formulate the Laplace-based log-likelihood functions in the estimation algorithm for the partially linear model, see Section 3.3 of the main paper:

$$
\mathcal{L}_n^{\{1\}}(\boldsymbol{\beta}_2,\boldsymbol{\beta}_3;\ \widehat{\boldsymbol{\beta}}_1^{\{0\}},x_0,h)
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \delta_i \Big\{ \ln\big(\alpha(x_{2i})(1-\alpha(x_{2i}))\big) \; + \; \ln\{g'(y_i)\} \; - \; \mathbf{x}_{2i,1}^T \boldsymbol{\beta}_2 \\
&\qquad\qquad - \rho_{\alpha(x_{2i})}\Big(\frac{g(y_i) - g^{-1}(\mathbf{x}_{1i}^T \widehat{\boldsymbol{\beta}}_1^{\{0\}})}{\exp\big(\mathbf{x}_{2i,1}^T \boldsymbol{\beta}_2\big)} \Big) \Big\} K_h(x_{2i} - x_0) \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \Big\{ \ln \Big[1 - \alpha(x_{2i}) \exp\Big\{ - \rho_{\alpha(x_{2i})}\Big(\frac{g(y_i) - g^{-1}(\mathbf{x}_{1i}^T \widehat{\boldsymbol{\beta}}_1^{\{0\}})}{\exp\big(\mathbf{x}_{2i,1}^T \boldsymbol{\beta}_2\big)} \Big) \Big\} \Big] \\
&\qquad\qquad \times \mathbb{I}\big(y_i < g^{-1}(\mathbf{x}_{1i}^T \widehat{\boldsymbol{\beta}}_1^{\{0\}})\big) \\
&\qquad\qquad + \Big\{ \ln\big(1-\alpha(x_{2i})\big) - \rho_{\alpha(x_{2i})}\Big(\frac{g(y_i) - g^{-1}(\mathbf{x}_{1i}^T \widehat{\boldsymbol{\beta}}_1^{\{0\}})}{\exp\big(\mathbf{x}_{2i,1}^T \boldsymbol{\beta}_2\big)} \Big) \Big\} \\
&\qquad\qquad \mathbb{I}\big(y_i > g^{-1}(\mathbf{x}_{1i}^T \widehat{\boldsymbol{\beta}}_1^{\{0\}})\big) \Big\} K_h(x_{2i} - x_0),
\end{aligned}
$$

where $\alpha(x_{2i})$ is approximated by $\mathbf{x}_{2i,1}^T \boldsymbol{\beta}_3$ as defined in (8) with $\mathbf{x}_{2i,1} = (1, (x_{2i} - x_0))^T$ and $\mathbf{x}_{1i} = (x_{11i}, \ldots, x_{1d_1 i})^T$; $i = 1, 2, \ldots, n$. Similarly,

$$
\begin{aligned}
&\mathcal{L}_n^{\{2\}}\big(\boldsymbol{\beta}_1;\ \widehat{\phi}^{\{1\}}(\cdot), \widehat{\alpha}^{\{1\}}(\cdot)\big) \\
&= \sum_{i=1}^{n} \delta_i \Big\{ \ln\big(\widehat{\alpha}^{\{1\}}(x_{2i})(1-\widehat{\alpha}^{\{1\}}(x_{2i}))\big) \; + \; \ln\{g'(y_i)\} - \; \ln\{\widehat{\phi}^{\{1\}}(x_{2i})\} \\
&\qquad\qquad - \rho_{\widehat{\alpha}^{\{1\}}(x_{2i})}\big(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{1i}^T \boldsymbol{\beta}_1\big)}{\widehat{\phi}^{\{1\}}(x_{2i})} \big) \Big\} \\
&\quad + \sum_{i=1}^{n} (1-\delta_i) \Big\{ \ln \Big[1 - \widehat{\alpha}^{\{1\}}(x_{2i}) \exp\big\{ - \rho_{\widehat{\alpha}^{\{1\}}(x_{2i})}\big(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{1i}^T \boldsymbol{\beta}_1\big)}{\widehat{\phi}^{\{1\}}(x_{2i})} \big) \big\} \Big] \\
&\qquad\qquad \times \mathbb{I}\big(y_i < g^{-1}(\mathbf{x}_{1i}^T \boldsymbol{\beta}_1)\big) \\
&\qquad\qquad + \Big(\ln\big(1-\widehat{\alpha}^{\{1\}}(x_{2i})\big) - \rho_{\widehat{\alpha}^{\{1\}}(x_{2i})}\big(\frac{g(y_i) - g^{-1}\big(\mathbf{x}_{1i}^T \boldsymbol{\beta}_1\big)}{\widehat{\phi}^{\{1\}}(x_{2i})} \big) \Big) \\
&\qquad\qquad \times \mathbb{I}\big(y_i > g^{-1}(\mathbf{x}_{1i}^T \boldsymbol{\beta}_1)\big) \Big\}.
\end{aligned}
$$

S2 Asymptotic properties

S2.1 Notations and conditions

Recall that the parameter vector $\boldsymbol{\theta}(x) = \big(\eta(x), \phi(x), \alpha(x)\big)^T$ belongs to the parameter space $\boldsymbol{\Theta} = (0, +\infty) \times (0, +\infty) \times (0, 1)$. Further, the vector $\tilde{\boldsymbol{\theta}}(x) = \big(\theta_1(x), \theta_2(x), \theta_3(x)\big)^T$ belongs to the parameter space $\mathbb{R}^3$. It is worth mentioning that the form of $f_{\alpha(x)}(t;\ \boldsymbol{\theta}(x))$ is supposed to be known, however, the parameters $\theta_r(x),\ r = 1, 2, 3$ are unspecified real-valued functions of the covariate $X = x$.

Although the asymptotic properties of the local MLE with multi-parameter likelihood models for complete data are provided in the literature, see for example, in Aerts & Claeskens [1] (for standard likelihood) and Gijbels et al. [2] (for non-standard likelihood), these results should be modified to the censoring and some additional assumptions are required that deal with the censoring.

Denote conditional log-likelihood function given the data $Y = y, \Delta = \delta$, and $X = x$ by

$$\ell(u_1, u_2, u_3;\ y, \delta, x) = \delta \ln f_{\alpha(x)}(y;\ u_1, u_2, u_3) + (1 - \delta) \ln S_{\alpha(x)}(y;\ u_1, u_2, u_3),$$

where $\delta = 1$ for uncensored and 0 for censored data. We further denote the partial derivative of $\ell(u_1, u_2, u_3;\ y, \delta, x)$, for $r, s, w \in \{1, 2, 3\}$ by

$$\begin{aligned}
&\psi_r(v_1(x), v_2(x), v_3(x);\ y, \delta) \\
&\qquad = \frac{\partial}{\partial u_r} \ell(u_1, u_2, u_3;\ y, \delta, x) \mid_{(u_1,u_2,u_3)=(v_1(x),v_2(x),v_3(x))} \\
&\psi_{rs}(v_1(x), v_2(x), v_3(x);\ y, \delta) \\
&\qquad = \frac{\partial^2}{\partial u_r \partial u_s} \ell(u_1, u_2, u_3;\ y, \delta, x) \mid_{(u_1,u_2,u_3)=(v_1(x),v_2(x),v_3(x))} \\
&\psi_{rsw}(v_1(x), v_2(x), v_3(x);\ y, \delta) \\
&\qquad = \frac{\partial^3}{\partial u_r \partial u_s \partial u_w} \ell(u_1, u_2, u_3;\ y, \delta, x) \mid_{(u_1,u_2,u_3)=(v_1(x),v_2(x),v_3(x))},
\end{aligned}$$

where $v_1 = g^{-1}(u_1),\ v_2 = \exp(u_2),\ v_3 = \dfrac{\exp(u_2)}{1 + \exp(u_2)}$. Then applying the chain rule to differentiate the log-likelihood function with respect to u, for $r, s \in \{1, 2, 3\}$, results in

$$\begin{aligned}
\frac{\partial}{\partial u_r} \ell(u_1, u_2, u_3;\ y, \delta, x) &= \frac{\partial}{\partial v_r} \ell(v_1, v_2, v_3;\ y, \delta, x) \frac{\partial v_r}{\partial u_r} \\
\frac{\partial^2}{\partial u_r u_s} \ell(u_1, u_2, u_3;\ y, \delta, x) &= \frac{\partial^2}{\partial v_r v_s} \ell(v_1, v_2, v_3;\ y, \delta, x) \frac{\partial v_r}{\partial u_r} \frac{\partial v_s}{\partial u_s},
\end{aligned}$$

where

$$\frac{\partial v_1}{\partial u_1} = (g^{-1})'(u_1), \quad \frac{\partial v_2}{\partial u_2} = v_2, \quad \text{and} \quad \frac{\partial v_3}{\partial u_3} = \frac{v_3}{1 + \exp(u_3)}.$$

Although the conditional TPA density given in (3) and its corresponding survival function in (6) are continuous everywhere for all $X = x$, it is not differentiable at points $y = g^{-1}(\theta_1(x))$. This brings us to establish the asymptotic behaviour of the local kernel likelihood estimators under *non-standard* conditions. One of the standard regularity conditions in (local-) maximum likelihood estimation is that the expected value of the score function with respect to each parameter needs to exist and equate to zero at the true parameter value. We also need this condition to build up the asymptotic properties in the non-standard case. More specifically, consider for $r, s \in \{1, 2, 3\}$,

$$\lambda_r(v_1(x), v_2(x), v_3(x)) = \mathbb{E}_{Y,\Delta|X}\{\psi_r(v_1(X), v_2(X), v_3(X);\ Y, \Delta \mid X = x)\}$$
$$\lambda_{rs}(v_1(x), v_2(x), v_3(x)) = \frac{\partial}{\partial u_s}\lambda_r(u_1, u_2, u_3) \mid_{(u_1,u_2,u_3)=(v_1(x),v_2(x),v_3(x))} .$$

Throughout, the notation $\mathbb{E}_{Y,\Delta|X}$ denotes the joint expectation for Y and Δ conditional on $X = x$.

We now turn to state the assumptions/conditions that are essentially needed to establish the asymptotic properties of the local kernel weighted estimators in case of non-standard local likelihood models with right censored data. These mild conditions mainly depend on the conditional distributions of T and C given $X = x$, the link function $g(\cdot)$, the smoothness of $\theta_r(\cdot)$, the design density f_X, the kernel function $K(\cdot)$ with its bandwidth parameter h.

Conditions on the conditional distributions of T and C given $X = x$:

(R1) The conditional TPA density $f_{\alpha(x)}(t;\ \boldsymbol{\theta}(x))$ has a common support for all x on an open subset $\mathring{\boldsymbol{\Theta}}$ of the parameter space $\boldsymbol{\Theta}$ containing the true parameters $\tilde{\boldsymbol{\theta}}(x) = \big(\theta_1(x), \theta_2(x), \theta_3(x)\big)^T$ for all x.

(R2) The reference symmetric density $f_0(z)$ satisfies $\lim\limits_{z\to+\infty} z f_0(z) = 0$ or $\int_0^\infty z f_0'(z)dz = -\frac{1}{2}$.

(R3) $g(\cdot)\colon \mathbb{R}_0^+ \to \mathbb{R}$ is monotone increasing and differentiable function, such that $\lim\limits_{t\to 0} g(t) = 0$ and $\lim\limits_{t\to+\infty} g(t) = +\infty$.

(R4) The distributions of T and C are conditionally independent given $X = x$.

(R5) $\lim\limits_{z\to+\infty} z^{r-1} f_0(z) G[g_{l\boldsymbol{\theta}}(z \mid X = x)] = 0$, and $\int_0^\infty z^{r-1} f_0(z) dG[g_{l\boldsymbol{\theta}}(z)] < +\infty$ for $r, l = 1, 2$, where $g_{1\boldsymbol{\theta}}(z \mid X = x) = g^{-1}\{g(\eta(x)) - \frac{\phi(x)}{1-\alpha(x)} z\}$ and $g_{2\boldsymbol{\theta}}(z \mid X = x) = g^{-1}\{g(\eta(x)) + \frac{\phi(x)}{\alpha(x)} z\}$.

(R6) $\mathcal{I}_{rs}(\theta_1(x), \theta_2(x), \theta_3(x)) = \mathbb{E}_{Y,\Delta|X}\{\psi_r(Y, \Delta;\ \theta_1(x), \theta_2(x), \theta_3(x)) \cdot \psi_s(Y, \Delta;\ \theta_1(x), \theta_2(x), \theta_3(x))\};\ r, s = 1, 2, 3$ is Lipschitz continuous and differentiable at $\big(\theta_1(x), \theta_2(x), \theta_3(x)\big)$ for all x; and the local Fisher information matrix

$$\mathcal{I}(\theta_1(x), \theta_2(x), \theta_3(x)) = \begin{bmatrix} \mathcal{I}_{11}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{12}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{13}(\theta_1(x), \theta_2(x), \theta_3(x)) \\ \mathcal{I}_{21}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{22}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{23}(\theta_1(x), \theta_2(x), \theta_3(x)) \\ \mathcal{I}_{31}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{32}(\theta_1(x), \theta_2(x), \theta_3(x)) & \mathcal{I}_{33}(\theta_1(x), \theta_2(x), \theta_3(x)) \end{bmatrix}$$

is positive definite, which depends on the censoring distribution G and the link function $g(\cdot)$.

Conditions on the design and the smoothness of $\theta_r(x)$ for all x in the support of X; $r = 1, 2, 3$:

(S1) The smooth function $\theta_r(\cdot)$ has a continuous derivative of order $(p_r + 1)$th for p_r odd and $(p_r + 2)$nd for p_r even, $r = 1, 2, 3$.

(S2) The design density f_X is differentiable and continuous on the interior of $\text{supp}(f_X)$.

(S3) For each $x \in \text{supp}(f_X)$, $g_r(\cdot)$ is continuous and permits third order derivatives and $g_r(\theta_r(x)), r = 1, 2, 3$, is nonzero.

Conditions on the kernel function and the bandwidth parameter:

(K1) The kernel function K is a symmetric bounded probability density compactly supported on $[-1, 1]$.

(K2) The bandwidth sequence $h = h_n$ satisfies: $h_n \to 0$ and $nh_n^3 \to +\infty$, as $n \to +\infty$.

The first condition (R1) is standard on the conditional probability density of the survival time. The conditions (R2)–(R5) are claimed to ensure that the expected value of the score function at the true value of the parameter of interest is zero. Condition (R6) is stated to provide a finite variance-covariance matrix of the local kernel weighted maximum likelihood estimators. Conditions (S1)–(S2) are postulated for the design density f_X and smoothness of $\theta_r(\cdot)$ on the support of a continuous covariate X, which are needed for the consistency and asymptotic normality of the estimators. The condition (S3) is required to translate the asymptotic results to the parameters of interest.The last two conditions (K1) and (K2) are targeted on the nature of the kernel function and the bandwidth parameter.

To prove the asymptotic normality of our estimator, we need some extra notation. These notations are adopted from Aerts & Claeskens [1], Gijbels et al. [2] and references their in. The main ingredients in this asymptotic normality result are: the existence of a positive semidefinite information matrix at the local point x, the kernel density K with the bandwidth parameter h, and the design density f_X. For the easy of illustration with condition (S2), let the local point x be lying in the interior of $\text{supp}(f_X)$. A point x will be called an interior point of $\text{supp}(f_X)$ if and only if $\{u : h^{-1}(u - x) \in \mathcal{A}\} \subseteq \text{supp}(f_X)$ with $\mathcal{A}$ denoting the support of $K(\cdot)$. Denote $\nu_j(\mathcal{A}) = \int_{\mathcal{A}} u^j K(u)du$. In addition, let $\boldsymbol{N}_{p_r p_s}(x)$, $\boldsymbol{T}_{p_r p_s}(x)$ and $\boldsymbol{Q}_{p_r p_s}(x)$ be the $(p_r + 1) \times (p_s + 1)$ dimension of matrices having $(k+1, l+1)$th entry equals, respectively, $\nu_{k+l}(\mathcal{A})$, $\int_{\mathcal{A}} u^{k+l} K^2(u)du$ and $\nu_{k+l+1}(\mathcal{A})$ $(k = 0, 1, \ldots, p_r;\ l = 0, 1, \ldots, p_s)$. Further, let $\boldsymbol{M}_{vp_s}(u)$ be the matrix obtained from $\boldsymbol{N}_{p_r p_s}(x)$ by replacing its $(v + 1)$th column by $(1,\ u,\ \ldots,\ u^{p_s})^T$, and for $\mid \boldsymbol{N}_{p_r p_s}(x) \mid$, define $K_{vp_s}(u) = v!\{ \mid \boldsymbol{M}_{vp_s}(u) \mid / \mid \boldsymbol{N}_{p_r p_s}(x) \mid \}K(u)$, where $K_{vp_s}(u)$ is called an equivalent kernel. Finally, we define, for $r \in \{1, 2, 3\}$

$$\boldsymbol{H}_{p_r} = \text{diag}(1,\ h, \ldots, h^{p_r})$$

$$\begin{aligned}\boldsymbol{\Sigma}_x &= f_X(x)\mathcal{I}(\theta_1(x),\theta_2(x),\theta_3(x))\otimes \boldsymbol{N}(x)\\ \boldsymbol{\Gamma}_x &= f_X(x)\mathcal{I}(\theta_1(x),\theta_2(x),\theta_3(x))\otimes \boldsymbol{T}(x)\\ \boldsymbol{\Lambda}_x &= \boldsymbol{D}(x)\otimes \boldsymbol{Q}(x),\end{aligned}$$

where

$$\boldsymbol{N}(x) = \begin{bmatrix} \boldsymbol{N}_{p_1p_1}(x) & \boldsymbol{N}_{p_1p_2}(x) & \boldsymbol{N}_{p_1p_3}(x)\\ \boldsymbol{N}_{p_2p_1}(x) & \boldsymbol{N}_{p_2p_2}(x) & \boldsymbol{N}_{p_2p_3}(x)\\ \boldsymbol{N}_{p_3p_1}(x) & \boldsymbol{N}_{p_3p_2}(x) & \boldsymbol{N}_{p_3p_3}(x)\end{bmatrix},$$

$$\boldsymbol{T}(x) = \begin{bmatrix} \boldsymbol{T}_{p_1p_1}(x) & \boldsymbol{T}_{p_1p_2}(x) & \boldsymbol{T}_{p_1p_3}(x)\\ \boldsymbol{T}_{p_2p_1}(x) & \boldsymbol{T}_{p_2p_2}(x) & \boldsymbol{T}_{p_2p_3}(x)\\ \boldsymbol{T}_{p_3p_1}(x) & \boldsymbol{T}_{p_3p_2}(x) & \boldsymbol{T}_{p_3p_3}(x)\end{bmatrix},$$

$$\boldsymbol{Q}(x) = \begin{bmatrix} \boldsymbol{Q}_{p_1p_1}(x) & \boldsymbol{Q}_{p_1p_2}(x) & \boldsymbol{Q}_{p_1p_3}(x)\\ \boldsymbol{Q}_{p_2p_1}(x) & \boldsymbol{Q}_{p_2p_2}(x) & \boldsymbol{Q}_{p_2p_3}(x)\\ \boldsymbol{Q}_{p_3p_1}(x) & \boldsymbol{Q}_{p_3p_2}(x) & \boldsymbol{Q}_{p_3p_3}(x)\end{bmatrix},$$

$$\boldsymbol{D}(x) = \frac{d}{dx}\big\{f_X(x)\mathcal{I}(\theta_1(x),\theta_2(x),\theta_3(x))\big\},$$

and $\otimes$ denotes a generalized Kronecker product. To be more precise, $\boldsymbol{D}(x)$ denotes element-by-element derivative of the matrix $f_X(x)\mathcal{I}(\theta_1(x),\theta_2(x),\theta_3(x))$, with respect to the local point x.

Further $\boldsymbol{W}^n(x) = \big(\boldsymbol{W}_1^n(x)^T, \boldsymbol{W}_2^n(x)^T, \boldsymbol{W}_3^n(x)^T\big)^T$, where $\boldsymbol{W}_r^n(x), r \in \{1,2,3\}$, is a column vector of dimension p_r+1, with $(k+1)$st component the partial derivative of the local kernel weighted log-likelihood function given by

$$\begin{aligned} W_{rk}^n(x) = \frac{1}{\sqrt{nh^{2k+1}}}\sum_{i=1}^n \psi_r\big(\bar\theta_1(X_i,x),\bar\theta_2(X_i,x),\bar\theta_3(X_i,x);\ Y_i,\Delta_i\big)\\ \times K\big\{(X_i-x)/h\big\}(X_i-x)^k,\end{aligned}$$

with $\bar\theta_r(X_i,x) = \sum_{j=0}^{p_r}\theta_{rj}(x)\big(X_i-x\big)^j;\ \ r\in\{1,2,3\}$.

S3 Additional simulation results

S3.1 Additional results to Section 5.1 in the main paper

Additional tables and figures for simulations under Scenario II

Figure S1 presents the boxplots of the AISE values for the two models for sample size $n = 300$. Figure S2 shows the boxplots of the AQRL values in quantile estimation from the simulations for Model II under **Scenario II**, for sample size $n = 300$. Finally, Table S1 lists the median AQRL values under this scenario for Model II for sample sizes $n = 100$ and $n = 300$.

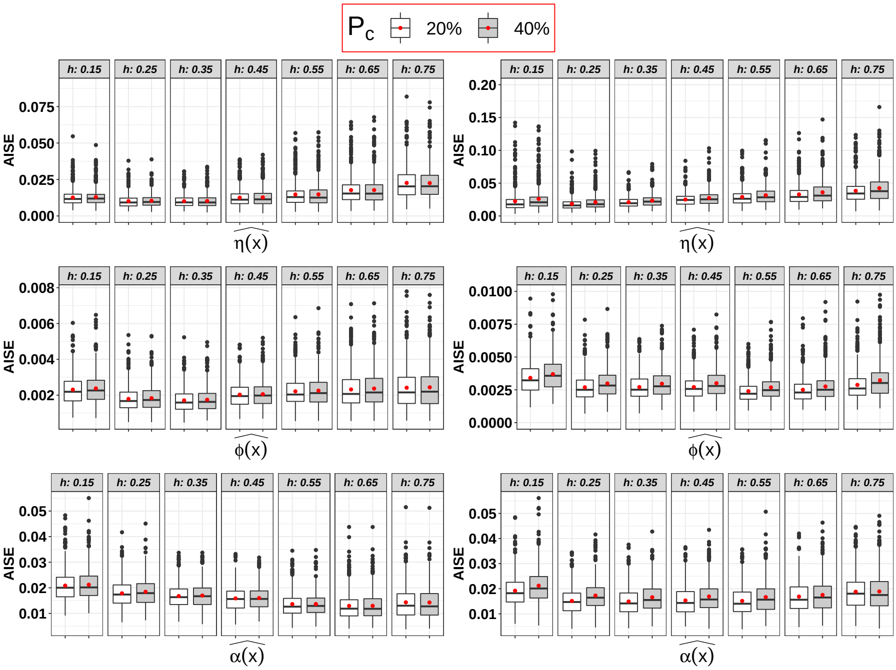


Fig. S1: **Scenario II.** AISE value across seven bandwidth values for the three estimates over 500 simulated samples of size $n = 300$ with 20% and 40% censoring proportions; Model I (left column) and Model II (right column).

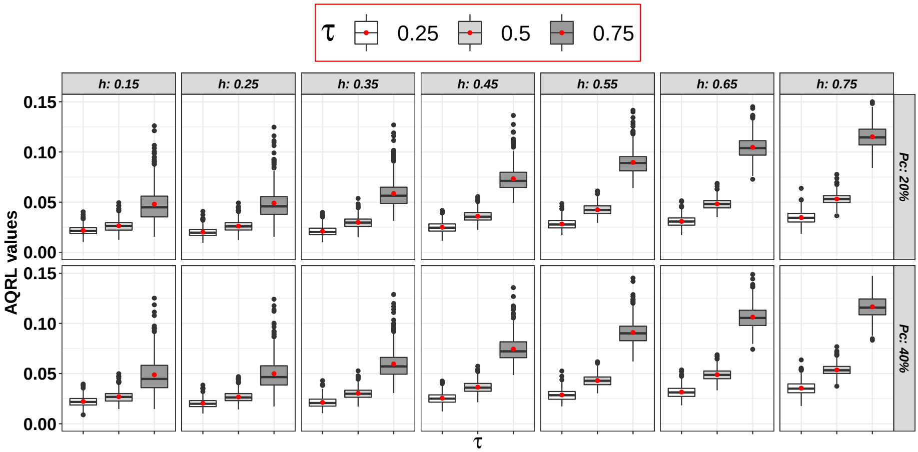
Simulation results under Scenario I

All the simulation results under **Scenario I** are reported here, using similar graphical presentations and similar presentations for tables as for Scenario II.

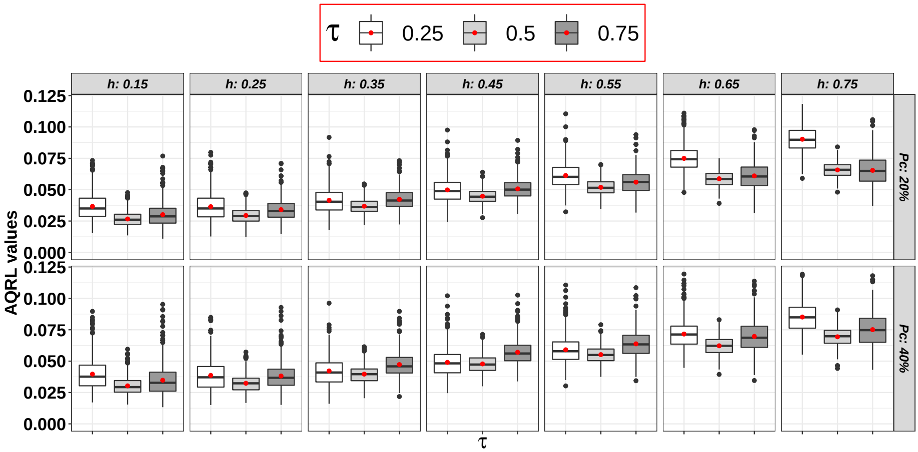
The AISE values for different fixed bandwidths are presented in boxplots in Figures S3 and S4, for respectively sample sizes $n = 100$ and $n = 300$. Median AISE values for Models I and II are listed in, respectively, Tables S2 and S3, for sample sizes $n = 100$ and $n = 300$.

Focusing on quantile estimation, Figures S5 and S6 provide the boxplots for the AQRL values, for respectively sample sizes $n = 100$ and $n = 300$. Tables S4 and S5 list the median AQRL values obtained for respectively Models I and II.

Finally, Figures S7 and S8 depict the true curves, together with three representatives curves of the estimated $\widehat{\boldsymbol{\theta}}(\cdot)$ and $\widehat{Q}_\tau(T \mid X = x)$, respectively.



(a) Model I



(b) Model II

Fig. S2: **Scenario II.** AQRL value across seven bandwidth values for the three quantiles over 500 simulated samples of size $n = 300$ with 20% and 40% censoring proportions; Model I (a) and Model II (b).

Table S1: **Scenario II.** Median AQRL values for the quantile estimator $\widehat{Q}_\tau(T \mid x)$, over a grid of x-values. Results for Model II, for three values of τ, two censoring proportions P_c, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

			bandwidth value h						
sample size	τ value	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	0.25	20%	0.052	0.044	0.044	0.050	0.060	0.073	0.087
		40%	0.056	0.048	0.047	0.051	0.058	0.068	0.082
	0.5	20%	0.042	0.039	0.042	0.049	0.056	0.063	0.069
		40%	0.047	0.043	0.046	0.053	0.061	0.066	0.073
	0.75	20%	0.042	0.042	0.048	0.056	0.062	0.068	0.072
		40%	0.048	0.047	0.054	0.063	0.072	0.078	0.081
$n = 300$	0.25	20%	0.035	0.035	0.040	0.049	0.060	0.074	0.090
		40%	0.038	0.037	0.041	0.048	0.058	0.071	0.085
	0.5	20%	0.026	0.029	0.036	0.044	0.052	0.059	0.066
		40%	0.029	0.032	0.040	0.047	0.055	0.062	0.070
	0.75	20%	0.029	0.033	0.041	0.050	0.056	0.061	0.065
		40%	0.033	0.037	0.046	0.056	0.063	0.069	0.075

Table S2: **Scenario I.** Median AISE values for the estimators $\widehat{\eta}(x)$, $\widehat{\phi}(x)$ and $\widehat{\alpha}(x)$ for Model I, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

	estimator		bandwidth value h						
sample size		P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	$\widehat{\eta}$	20%	0.582	0.465	0.380	0.384	0.430	0.500	0.657
		40%	1.543	0.710	0.523	0.504	0.557	0.613	0.770
	$\widehat{\phi}$	20%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
		40%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
	$\widehat{\alpha}$	20%	0.036	0.030	0.026	0.022	0.020	0.018	0.017
		40%	0.046	0.034	0.029	0.026	0.024	0.021	0.021
$n = 300$	$\widehat{\eta}$	20%	0.241	0.198	0.195	0.225	0.239	0.306	0.432
		40%	0.307	0.255	0.243	0.269	0.277	0.336	0.451
	$\widehat{\phi}$	20%	0.002	0.002	0.002	0.002	0.002	0.002	0.002
		40%	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	$\widehat{\alpha}$	20%	0.020	0.018	0.017	0.015	0.013	0.011	0.013
		40%	0.024	0.021	0.018	0.016	0.015	0.013	0.015

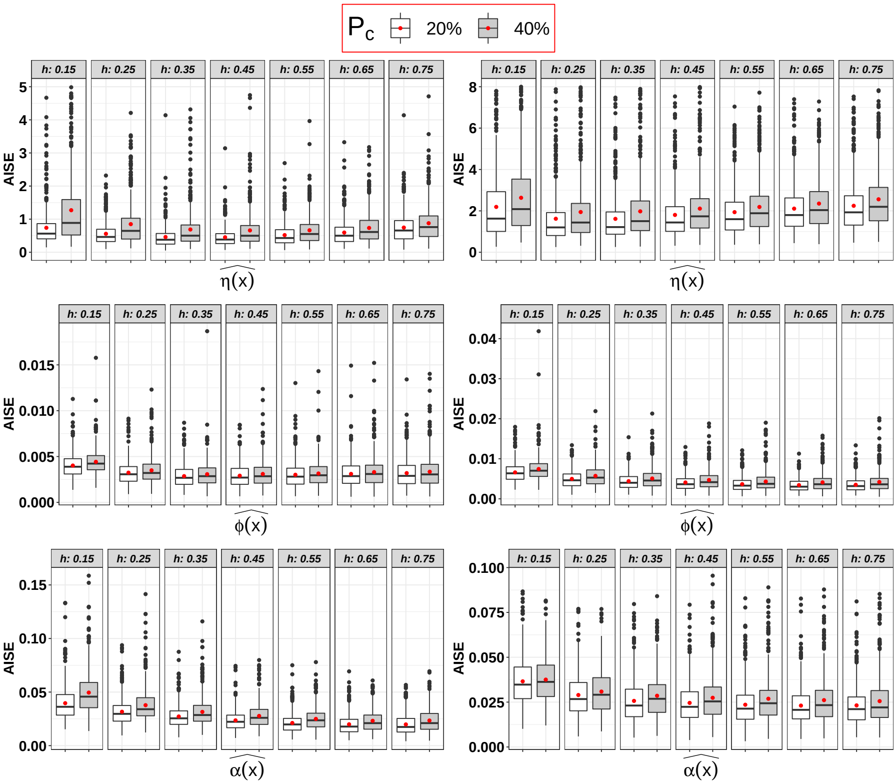


Fig. S3: **Scenario I.** AISE values across seven bandwidth values for the three estimates over 500 simulated samples of size $n = 100$ with 20% and 40% censoring proportions; Model I (left column) and Model II (right column).

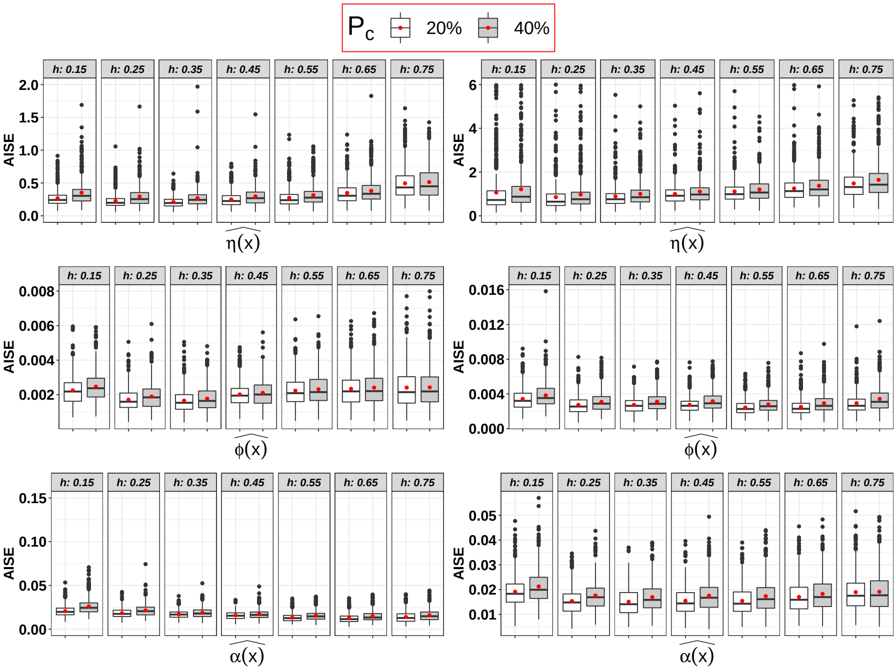
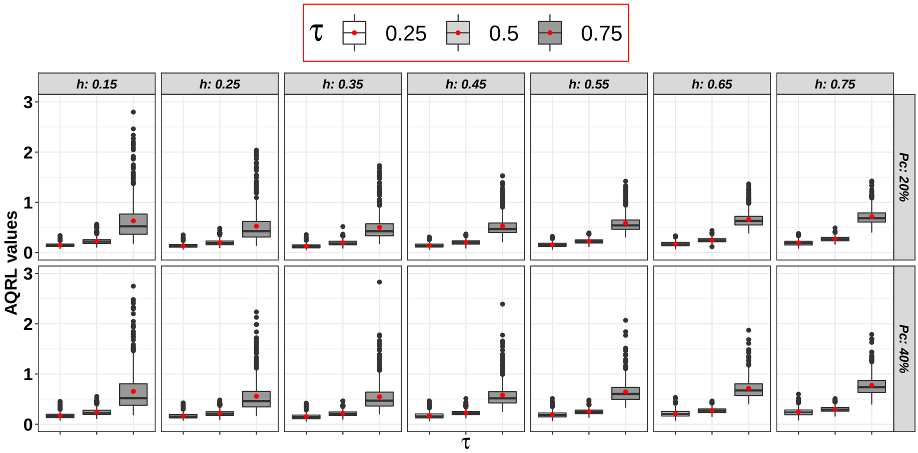


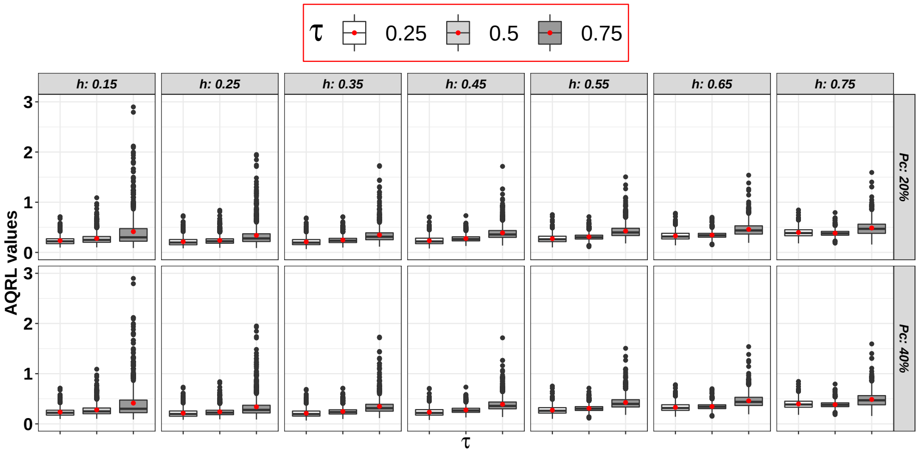
Fig. S4: **Scenario I.** AISE values across seven bandwidth values for the three estimates over 500 simulated samples of size $n = 300$ with 20% and 40% censoring proportions; Model I (left column) and Model II (right column).

Table S3: **Scenario I.** Median AISE values for the estimators $\widehat{\eta}(x)$, $\widehat{\phi}(x)$ and $\widehat{\alpha}(x)$ for Model II, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

sample size	estimator	P_c	bandwidth value h 0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	$\widehat{\eta}$	20%	2.100	1.310	1.260	1.451	1.631	1.806	1.950
		40%	2.919	1.606	1.586	1.777	1.910	2.054	2.219
	$\widehat{\phi}$	20%	0.006	0.005	0.004	0.004	0.003	0.003	0.003
		40%	0.007	0.005	0.005	0.004	0.004	0.004	0.004
	$\widehat{\alpha}$	20%	0.035	0.027	0.023	0.022	0.021	0.021	0.021
		40%	0.036	0.029	0.027	0.025	0.024	0.023	0.022
$n = 300$	$\widehat{\eta}$	20%	0.748	0.654	0.758	0.917	0.989	1.130	1.314
		40%	0.925	0.762	0.849	0.973	1.068	1.206	1.424
	$\widehat{\phi}$	20%	0.003	0.003	0.003	0.003	0.002	0.002	0.003
		40%	0.004	0.003	0.003	0.003	0.003	0.003	0.003
	$\widehat{\alpha}$	20%	0.018	0.015	0.014	0.014	0.014	0.016	0.018
		40%	0.020	0.017	0.016	0.017	0.016	0.017	0.018

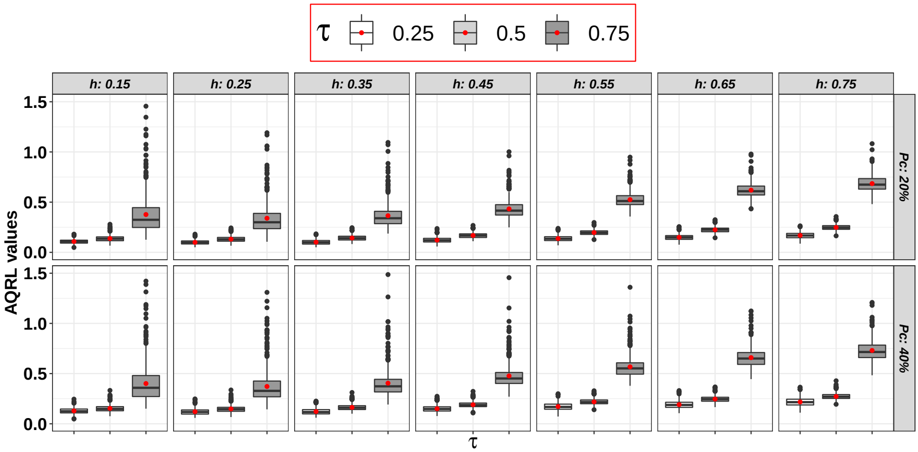


(a) Model I

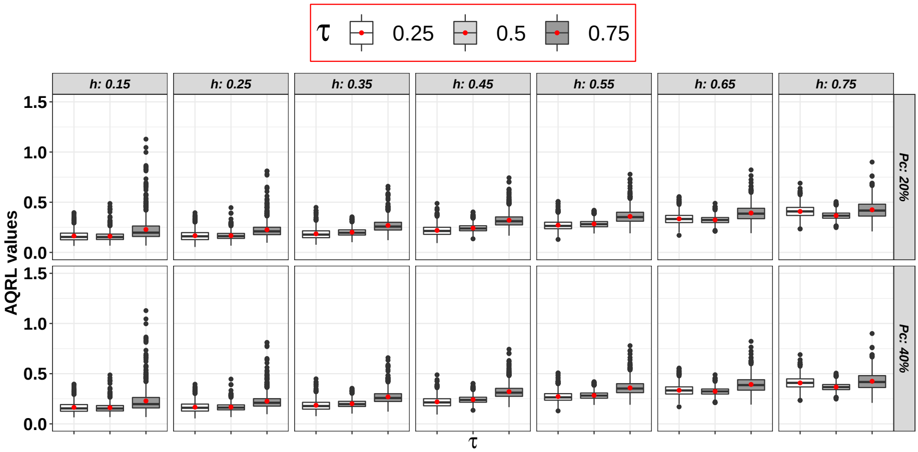


(b) Model II

Fig. S5: **Scenario I.** AQRL values across seven bandwidth values for the three quantiles over 500 simulated samples of size $n = 100$ with 20% and 40% censoring proportions; Model I (a) and Model II (b).



(a) Model I



(b) Model II

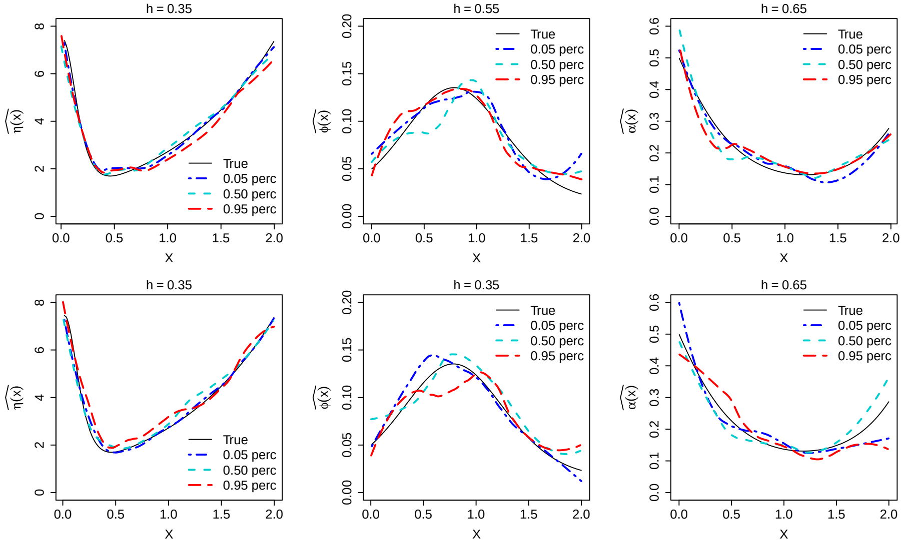
Fig. S6: **Scenario I.** AQRL values across seven bandwidth values for the three quantiles over 500 simulated samples of size $n = 300$ with 20% and 40% censoring proportions; Model I (a) and Model II (b).

Table S4: **Scenario I.** Median AQRL values for the quantile estimator $\widehat{Q}_\tau(T \mid x)$, over a grid of x-values. Results for Model I, for three values of τ, two censoring proportions P_c, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

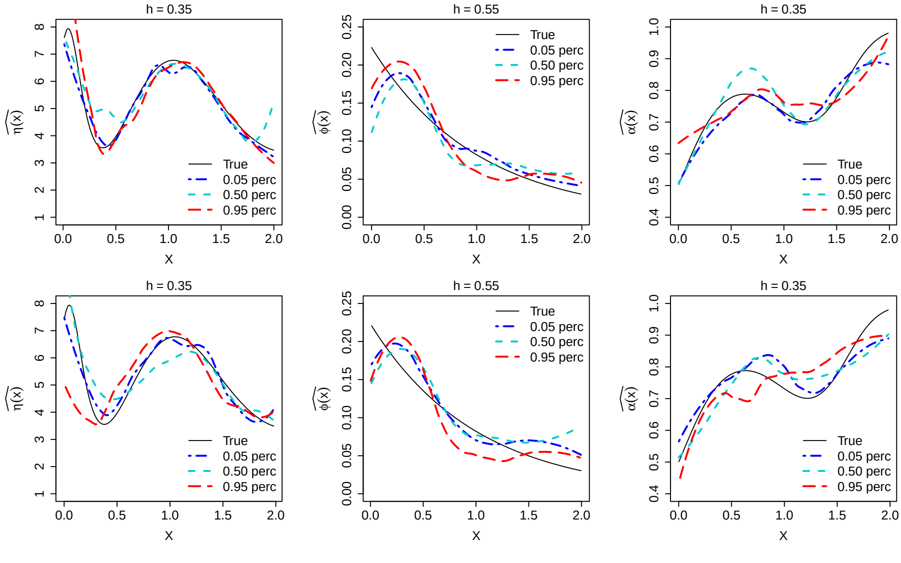
			bandwidth value h						
sample size	τ value	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	0.25	20%	0.145	0.133	0.120	0.134	0.151	0.167	0.189
		40%	0.166	0.155	0.141	0.157	0.182	0.208	0.236
	0.5	20%	0.218	0.189	0.184	0.202	0.221	0.244	0.266
		40%	0.229	0.209	0.204	0.221	0.246	0.268	0.291
	0.75	20%	0.525	0.427	0.425	0.468	0.542	0.630	0.685
		40%	0.521	0.461	0.469	0.516	0.602	0.673	0.738
$n = 300$	0.25	20%	0.107	0.098	0.099	0.120	0.136	0.148	0.167
		40%	0.125	0.118	0.117	0.148	0.168	0.188	0.215
	0.5	20%	0.135	0.129	0.142	0.168	0.196	0.225	0.247
		40%	0.149	0.145	0.160	0.187	0.216	0.245	0.270
	0.75	20%	0.324	0.300	0.340	0.415	0.511	0.609	0.675
		40%	0.359	0.328	0.373	0.451	0.552	0.650	0.715

Table S5: **Scenario I.** Median AQRL values for the quantile estimator $\widehat{Q}_\tau(T \mid x)$, over a grid of x-values. Results for Model II, for three values of τ, two censoring proportions P_c, for the different bandwidth values. Sample sizes $n = 100$ and $n = 300$.

			bandwidth value h						
sample size	τ value	P_c	0.15	0.25	0.35	0.45	0.55	0.65	0.75
$n = 100$	0.25	20%	0.213	0.190	0.188	0.219	0.272	0.328	0.403
		40%	0.228	0.207	0.201	0.216	0.250	0.304	0.367
	0.5	20%	0.230	0.207	0.222	0.252	0.288	0.331	0.371
		40%	0.268	0.236	0.250	0.285	0.317	0.352	0.391
	0.75	20%	0.276	0.255	0.284	0.332	0.369	0.409	0.433
		40%	0.331	0.307	0.336	0.390	0.433	0.481	0.503
$n = 300$	0.25	20%	0.149	0.155	0.181	0.218	0.276	0.343	0.420
		40%	0.162	0.166	0.175	0.210	0.258	0.316	0.388
	0.5	20%	0.142	0.152	0.186	0.230	0.272	0.312	0.359
		40%	0.168	0.175	0.207	0.247	0.290	0.333	0.378
	0.75	20%	0.180	0.193	0.237	0.291	0.329	0.358	0.387
		40%	0.212	0.226	0.277	0.334	0.379	0.419	0.448

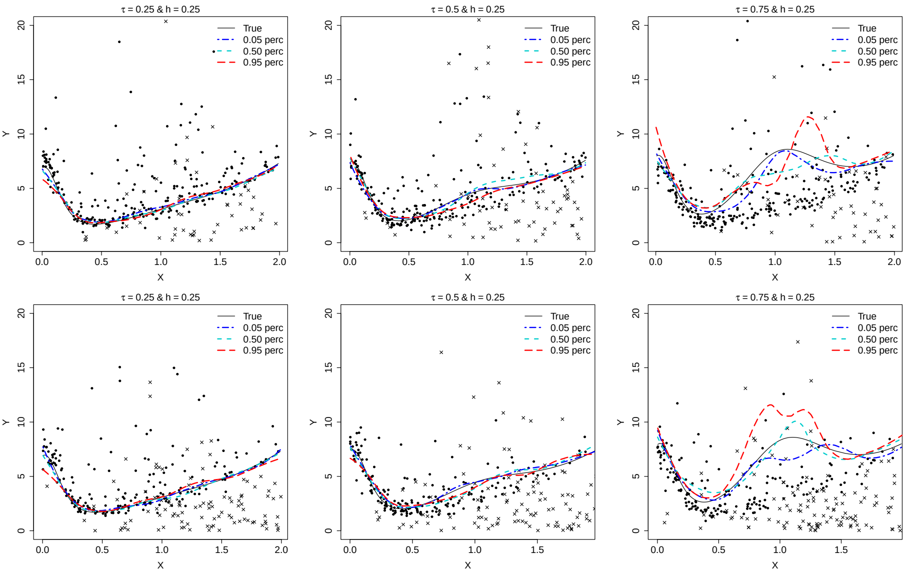


(a) Model I

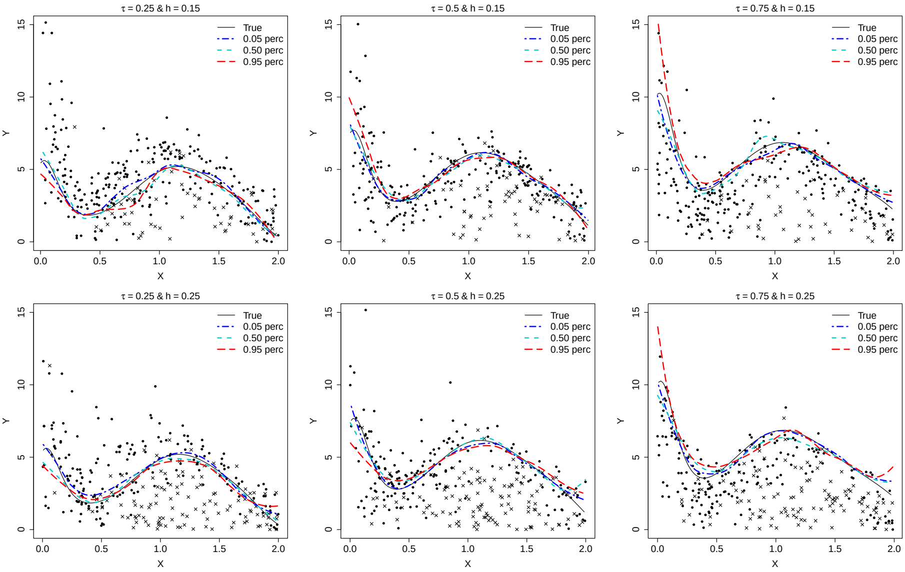


(b) Model II

Fig. S7: **Scenario I.** True and estimated curves of $\widehat{\eta}(x), \widehat{\phi}(x)$ and $\widehat{\phi}(x)$ for $n = 300$ and 20% censoring (first row) and 40% censoring (second row). Model I **(a)** and Model II **(b)**.



(a) Model I



(b) Model II

Fig. S8: **Scenario I.** True and estimated curves of $\widehat{Q}_\tau(x)(\tau = 0.25, 0.50, 0.75)$ for Model I **(a)** and Model II **(b)**. The sample size is $n = 300$ with 20% censoring (first row) and 40% censoring (second row). The scatter plot represents the sample data selected according to the 0.50th AQRL percentile. Dots ($\bullet$) represent the uncensored cases and crossed ($\times$) for censored cases.

S3.2 Weighted versus unweighted AQRL

In the main paper, both unweighted and weighted AQRL criteria have been used to measure the performance of the quantile estimates. The reason for using the weighted version of AQRL in the second simulation part has been explained in the main paper. For further illustrative purposes, we applied both criteria under **Scenario I** for Model I in the first part of simulation study. Figures S9 and S10, respectively, depict the average AQRL and WAQRL for $n = 100$ and $n = 300$ over the 500 simulated samples across the seven different bandwidth values. Note that the conclusive remarks are the same when using the weighted or unweighted performance measure.

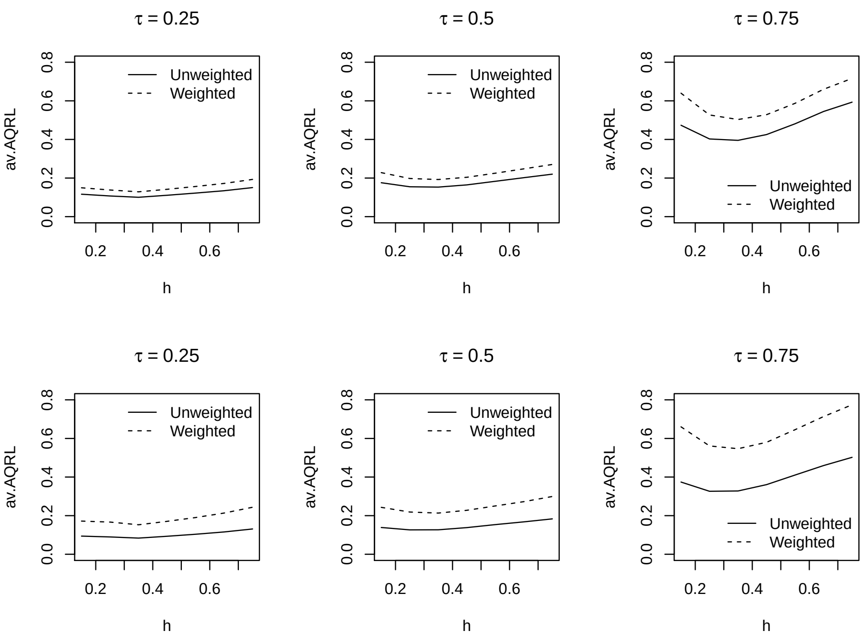


Fig. S9: **Scenario I:** average AQRL and WAQRL values for Model I and $n = 100$ with 20% censoring proportion (first row) and 40% censoring proportion (second row).

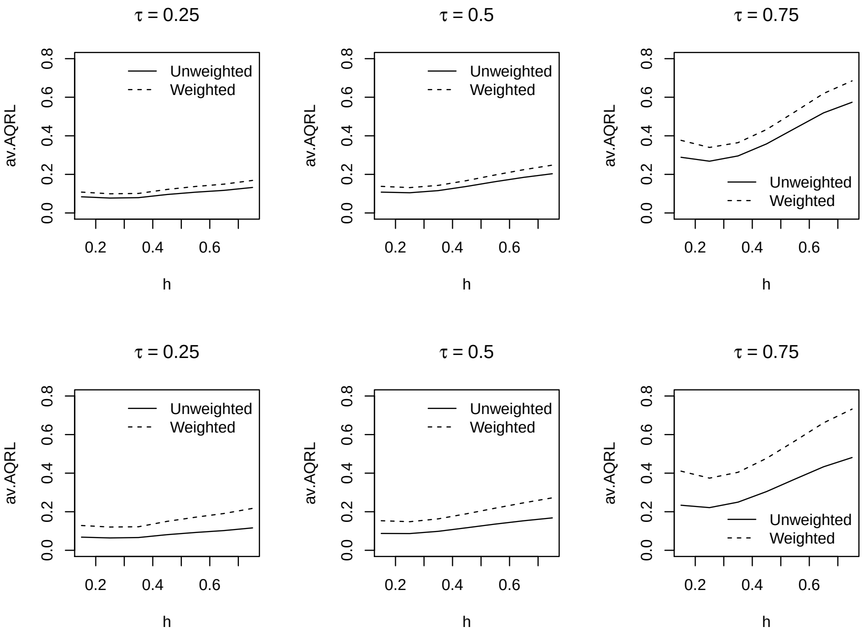


Fig. S10: **Scenario I:** average AQRL and WAQRL values for Model I and $n = 300$ with 20% censoring proportion (first row) and 40% censoring proportion (second row).

S4 SCLC data

Figures S11 and S12 provide some extra illustrations for the SCLC data supporting the statements in the main paper.

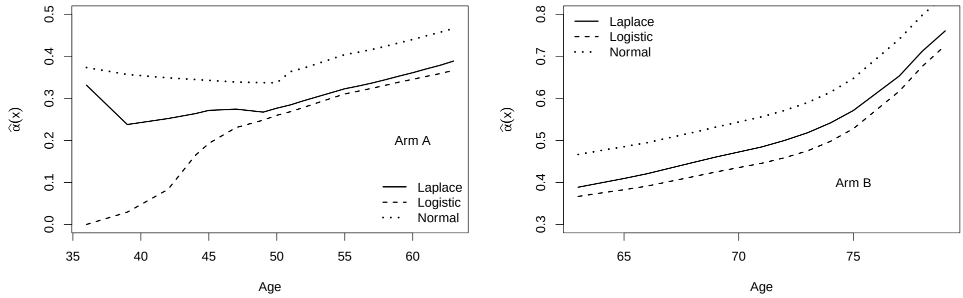


Fig. S11: Estimated $\widehat{\alpha}(x)$ curves for the SCLC data with conditional $\alpha(\cdot)$; solid line for the TPA Laplace, dashed line for the TPA logistic and dotted line for the TPA normal model.

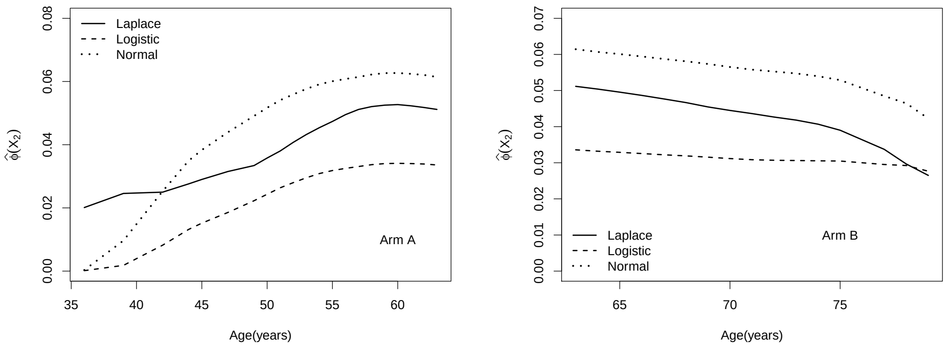


Fig. S12: Estimated $\widehat{\phi}(x)$ curves for the SCLC data with conditional $\alpha(\cdot)$; solid line for the TPA Laplace, dashed line for the TPA logistic and dotted line for the TPA normal model.

References

[1] Aerts, M., & Claeskens, G. (1997). Local polynomial estimation in multiparameter likelihood models. *Journal of the American Statistical Association*, **92**(440), 1536–1545.

[2] Gijbels, I., Karim, R. & Verhasselt, A. (2021). Semiparametric quantile regression using quantile-based asymmetric family of densities. *Computational Statistics and Data Analysis*, **157**, Article #107129.